Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 22 (2020), Issue 9 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Tutorial

Original Papers

Letters to the Editor

Corrigenda and Addenda

Short Paper

Review

Digital Interventions for Screening and Treating Common Mental Disorders or Symptoms of Common Mental Illness in Adults: Systematic Review and Meta-analysis

Jacqueline Sin[1,2], RMN, PhD; Gian Galeazzi[3], PhD, MD; Elicia McGregor[4], MSc; Jennifer Collom[4], MSc; Anna Taylor[4], MSc; Barbara Barrett[4], PhD; Vanessa Lawrence[4], PhD; Claire Henderson[4], PhD, FRCPsych

[1]School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom

[2]Population Health Research Institute, St George's, University of London, London, United Kingdom

[3]Department of Biomedical, Metabolic, and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

[4]Health Services & Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Jacqueline Sin, RMN, PhD
School of Psychology and Clinical Language Sciences
University of Reading
Earley Gate, Whiteknights
Reading, RG6 6AL
United Kingdom
Phone: 44 7817027035
Email: jacqueline.sin@reading.ac.uk

Abstract

Background: Digital interventions targeting common mental disorders (CMDs) or symptoms of CMDs are growing rapidly and gaining popularity, probably in response to the increased prevalence of CMDs and better awareness of early help-seeking and self-care. However, no previous systematic reviews that focus on these novel interventions were found.

Objective: This systematic review aims to scope entirely web-based interventions that provided screening and signposting for treatment, including self-management strategies, for people with CMDs or subthreshold symptoms. In addition, a meta-analysis was conducted to evaluate the effectiveness of these interventions for mental well-being and mental health outcomes.

Methods: Ten electronic databases including MEDLINE, PsycINFO, and EMBASE were searched from January 1, 1999, to early April 2020. We included randomized controlled trials (RCTs) that evaluated a digital intervention (1) targeting adults with symptoms of CMDs, (2) providing both screening and signposting to other resources including self-care, and (3) delivered entirely through the internet. Intervention characteristics including target population, platform used, key design features, and outcome measure results were extracted and compared. Trial outcome results were included in a meta-analysis on the effectiveness of users' well-being and mental health outcomes. We also rated the meta-analysis results with the Grading of Recommendations, Assessment, Development, and Evaluations approach to establish the quality of the evidence.

Results: The electronic searches yielded 21 papers describing 16 discrete digital interventions. These interventions were investigated in 19 unique trials including 1 (5%) health economic study. Most studies were conducted in Australia and North America. The targeted populations varied from the general population to allied health professionals. All interventions offered algorithm-driven screening with measures to assess symptom levels and to assign treatment options including automatic web-based psychoeducation, self-care strategies, and signposting to existing services. A meta-analysis of usable trial data showed that digital interventions improved well-being (3 randomized controlled trials [RCTs]; n=1307; standardized mean difference [SMD] 0.40; 95% CI 0.29 to 0.51; I^2=28%; fixed effect), symptoms of mental illness (6 RCTs; n=992; SMD −0.29; 95% CI −0.49 to −0.09; I^2=51%; random effects), and work and social functioning (3 RCTs; n=795; SMD −0.16; 95% CI −0.30 to −0.02; I^2=0%; fixed effect) compared with waitlist or attention control. However, some follow-up data failed to show any sustained effects beyond the post intervention time point. Data on mechanisms of change and cost-effectiveness were also lacking, precluding further analysis.

Conclusions: Digital mental health interventions to assess and signpost people experiencing symptoms of CMDs appear to be acceptable to a sufficient number of people and appear to have enough evidence for effectiveness to warrant further study. We recommend that future studies incorporate economic analysis and process evaluation to assess the mechanisms of action and cost-effectiveness to aid scaling of the implementation.

(J Med Internet Res 2020;22(9):e20581) doi:10.2196/20581

KEYWORDS

eHealth; mHealth; psychiatric illness; mental disorders; common mental illness; depression; anxiety; self-care

Introduction

Background

There are several reasons to study stand-alone digital technology interventions as the first step in the assessment and management of symptoms of common mental disorders (CMDs). CMDs include different types of depression and anxiety and can cause marked emotional distress and interfere with daily functioning [1,2]. First, access to digital technologies is high in many countries and is increasing in many others [1,2]. Second, mild disorders frequently remit without professional treatment, and, instead, self-management strategies can be learned to ameliorate symptoms and prevent future episodes [3]. Third, there are many digital interventions available for CMDs and their related problems, such as poor sleep [4], and for the promotion of mental well-being such as mindfulness [5]. Some have been subjected to rigorous evaluation [6], whereas others have not been subjected per se but are digital applications of evidence-based therapies such as cognitive behavioral therapy (CBT). Fourth, there is evidence that CMDs are increasing in prevalence in groups such as young women and people aged 55 to 64 years [7], and it is not possible to meet these needs in primary care or specialist mental health service based on current resources and workforce supply [8,9]. Fifth, it should not be assumed that digital interventions are a cost-effective way to meet needs that cannot presently be met by the health workforce. They carry development and maintenance costs, and the work entailed must ensure usability and acceptability. Furthermore, for costs to be offset, the intervention must be accessed by a sufficient number of people who experience benefits above and beyond any other service they may be accessing; ensuring this widespread awareness among people likely to benefit also carries costs [10]. Finally, many people prefer to manage their symptoms without recourse to professional services, often because of a desire for self-reliance but also for reasons such as fear of stigmatization and discrimination and barriers to accessing specialist mental health treatment, for example, because of working long hours, the need for a general practitioner or medical referral, or living in a rural area [11,12].

Our starting point for this review is the development and launch in 2017 of one such digital intervention, *Good Thinking*, for people living and working in London, United Kingdom. Good Thinking provides an initial assessment and signposting to web-based self-guided interventions, including self-care and community-based resources, virtual or otherwise, entirely on the web. This comprises 4 modules: sleep problems, stress, low mood, and anxiety, and includes a self-assessment and signposting to mental health self-management apps, digital therapies (eg, Sleepio for sleep problems [13] or FearFighter—a web-based CBT for social phobia or panic disorder [14]), and conventional services. The apps were approved by *NHS Digital*, the organization in charge of digital services within the UK National Health Service (NHS) using a pre-existing quality control process that included considering the evidence base applied in the digital treatment [15]. The user can choose 1 of these 4 modules and be signposted based on responses to questions on the web-based platform, which can be answered regarding the self or someone they know. Alternatively, the user can use a self-assessment tailored for signposting based on algorithms used for the national telephone helpline, NHS 111.

Good Thinking thus differs from digital therapy delivery, which has been the subject of previous reviews [16-21]. Although these reviews focused on CMDs (such as depression and anxiety disorders [6,18], posttraumatic stress disorder [17,22], and insomnia [16]), they investigated the effectiveness of digital psychotherapies, mostly on CBT provided by health care professionals, although with varying degrees of synchronized or asynchronized guidance delivered on the web. Such interventions tend to follow an assessment conducted by a health professional to validate the diagnosis and include further therapist-delivered psychological interventions using various media. In contrast, Good Thinking exemplifies a new breed of digital mental health interventions that allow users to be in complete control of the process (from access to assessment), intervention (emphasizing self-management), and outcome assessments. These users may have symptoms of CMDs, not necessarily meeting diagnostic or mental health service thresholds or not needing specialist services or conventional therapist-led interventions. Many are primarily interested in seeking digital applications that promote self-care for well-being and signposting to alternative services such as a helpline and peer support forums. As such, this broad range of interventions is likely to be sought by a wide population at a time when many countries are promoting awareness and self-care for mental health, such as Every Mind Matters in England and BeyondBlue in Australia).

To the best of our knowledge, no previous reviews have focused on potentially heterogeneous populations using interventions such as Good Thinking that include a self-assessment to help a web-based user choose their next step in terms of self-management or help-seeking. This review of the interventions and their evaluation will contribute to the development and implementation of more successful applications and hence more effective and sustainable web-based interventions.

Objectives

This study aims to conduct a comprehensive systematic review of studies of digital mental health services that provide web-based self-assessment and treatment that emphasize on self-care for people with common mental health disorders or subthreshold symptoms. We examined randomized controlled trials (RCTs), the fairest and most robust study design in evaluating the effectiveness of entirely web-based interventions aimed at optimizing mental health–related and intermediate outcomes, including self-care, informal support, and treatment services. We planned to conduct meta-analyses on the (cost-) effectiveness of the interventions on mental well-being and CMD symptom outcomes. Using the research evidence, we also aimed to examine the evidence for the mechanisms of action of such interventions through intermediate or health behavioral change outcomes to mental health outcomes.

Methods

Data Sources and Search Strategy

Searches for papers written in English, from January 1, 1999 (when electronic and digital health interventions were first documented) to September 20, 2018, were conducted using MEDLINE (Medical Literature Analysis and Retrieval System Online) and MEDLINE in-process, PsycINFO (Psychological Information), CINAHL (Cumulative Index of Nursing and Allied Health Literature), EMBASE (Excerpta Medica dataBASE), CENTRAL (Cochrane Central Register of Controlled Trials), WoS (Web of Science), ASSIA (Applied Social Sciences Index and Abstracts), DARE (Database of Abstracts of Reviews of Effect), HTA (Health Technology Assessment) published and in-process, and NHS EED (NHS Economic Evaluation Database). Once an initial set of papers from the databases were identified, we performed backward and forward searches in their reference lists and citations of the identified papers for any additional studies. We also contacted the authors of the included papers to retrieve relevant information about their study if this was unclear from the published article. To identify articles not included in our original search, we tracked published protocols of trials identified in 2018 and conducted an update search on MEDLINE, PsycINFO, EMBASE, ASSIA, and WoS for any new publications up to April 9, 2020.

We devised search terms using the population, intervention, comparison, and outcome of interest approach [23]. As the search aimed to be highly sensitive, we employed an initial search strategy that combined search terms for populations (eg, common mental health disorders, adults, depression, and anxiety) and interventions (eg, digital/ ehealth* /mhealth* /web /online /internet adj3 intervention/program*/initiative*/group*). We refined and adapted the search terms used to suit different database search systems. We have published a review protocol in PROSPERO (Prospective Register of Systematic Reviews, CRD42017079085) [24]. The review process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25].

Study Eligibility and Selection

We included studies that targeted adults aged 18 years, with no upper age limit. According to the UK Adult Psychiatric Morbidity Survey (APMS [7]), CMDs include different types of depression and anxiety and can cause marked emotional distress and interfere with daily functioning, but do not usually affect insight or cognition. Symptoms of CMD include somatic symptoms, fatigue, sleep problems, irritability, worry about physical health, concentration and forgetfulness, depression, generalized worry, anxiety, phobias, panic, compulsions, and obsessions [7]. We also consulted experts in the field to establish whether certain types or symptoms of illnesses, not covered by the APMS definitions, fit the criteria of CMD. Examples include perinatal depression.

We included studies of any digital mental health interventions that aimed to support individuals directly and were fully delivered using web-based information and communication technology (ICT). Facilitation by nondigital resources, such as professionals or lay persons, did not affect study inclusion as far as the intervention was fully delivered using web-based ICT. We specified that intervention contents must include screening or diagnostic assessment and self-care for mental health promotion or symptom management as part of the treatment that can also include information giving, signposting or recommendations, informal support, and pre-existing treatment options. We excluded interventions designed to solely provide assessment or treatment, but not both. To examine the (cost-) effectiveness of the identified interventions, we included only empirical studies using a web-based RCT design for optimal external and internal validity [26] and with intervention recipients' outcomes reported using validated quantitative measures.

One author (AT, JC, EM, or JS) screened all retrieved items through their titles, abstracts, and then full text. Another author (JS or GG) conducted an independent check on a random 20% sample of all the items at each step and a third author (CH) reviewed a proportion of searches, screening, and study selection. Disagreements were resolved through (1) seeking additional data or clarification from study authors when possible and (2) reviewing during team discussions. All study selection processes were conducted using EndNote software version 8.0 (Clarivate Analytics).

Outcomes and Measures

For this comprehensive review, we set a range of primary outcomes focusing on participants' symptoms of CMDs and their related domains. These included symptoms of mental illness, well-being, quality of life, perceived social support, work and social functioning, self-efficacy or coping, and adverse events. Process and/or intermediate outcomes were specified as health behavior change or proxy measures that are conduits to primary outcomes. These included the uptake of recommendations on self-care strategies and increased behavioral activation (such as, goal setting, self-monitoring, and general communication skills) [26]. In addition, we examined data on satisfaction or perceived acceptability of the intervention.

Data Extraction and Analysis

Relevant extracted data from the included studies were entered into a summary table devised by the review team. We extracted study design and data variables from each included study for further analysis, including sample size, setting, participant characteristics (such as age, gender, diagnosis or symptoms or complaints, and ethnicity), outcome measures, time points, and control condition or comparator. Data on the intervention extracted were as follows: aims, theoretical framework if used and described, content and features, and duration of intervention both in terms of usage hours if specified and the period during which the intervention was undertaken.

Regarding the theoretical framework, we scoped the theoretical basis used by the studies (eg, social cognitive theory, health belief model), the use of theory (eg, theory or predictors used to select recipients for the intervention) in informing intervention design [27], and any behavior change techniques employed by the identified intervention (eg, stress management, goal setting) [28]. We devised a coding system for these factors as they have been established to be particularly effective in promoting intervention uptake and effectiveness [28-30].

Data extracted on the content and features included the following:

1. The modes of delivery, access, and overall approach of the interventions.
2. Web based (ie, eHealth), mobile health (mHealth), or both eHealth and mHealth.
3. With social networking function, no social network, or combined therapy and social networking.
4. Free versus paid versus depending on contract.
5. Treatment options including self-care or management, informal support such as using peer support or community support resources, or signposting to formal or statutory services.

Data analysis started with an overview of study and intervention characteristics, followed by the tabulation of extracted data. All data deemed relevant for each review objective were grouped together and synthesized using a narrative approach. When sufficient homogeneous data were available, we conducted meta-analyses to investigate the effectiveness of treatment using Review Manager (version 5.3, the Cochrane Collaboration). A meta-regression to investigate the significance of identified moderators on treatment effectiveness was considered in the event that 10 studies were included in a meta-analysis [31]. We used a fixed-effects model when <5 studies were included in the meta-analysis and a random-effects model when ≥5 studies were included in the meta-analysis [31]. In addition to conducting overall analyses comparing digital interventions with all comparators pooled together, we also conducted separate comparisons of digital interventions against all inactive controls (eg, waitlist or usual care) and digital interventions against active controls (eg, interventions augmented with a nondigital element such as therapist support via face-to-face or phone contact or attention controls). As the outcomes were measured with different validated scales, we calculated standardized mean difference (SMD) and 95% CI for continuous outcomes and risk ratio and its 95% CI for dichotomous data [32]. Statistical heterogeneity was quantified using the I^2 statistics in addition to the visual inspection of the forest plots, with I^2 values >50% interpreted as evidence of substantial levels of heterogeneity [31]. Although some consider SMDs of 0.2, 0.5, and 0.8 as small, medium, and large effects, respectively, the magnitude of these effects alone has been criticized as not having any relationship with their clinical importance [31]. Instead, SMDs should be interpreted within the context of overall quantity and quality of the data included in the meta-analysis (see following sections).

Assessment of Study and Evidence Quality

We used the integrated criteria for a review of multiple study designs (ICROMS [33]) to assess the quality of the included studies. All studies were assessed for 7 dimensions: clear aims and justification; managing bias in sampling or between groups, in follow-ups, and in other study aspects; analytical rigor; and managing bias in reporting or ethical considerations. Each criterion was evaluated on a 3-point scale (2=criterion met, 1=unclear, 0=criterion not met). The ICROMS minimum score requirement for RCTs, including cluster (ie, 22), was used to rate the trial quality rather than to exclude studies on grounds of quality to retain usable data [33]. In addition, we also used the CONSORT (Consolidated Standards of Reporting Trials) eHealth Checklist (v.1.6.1) [34] to assess trial reporting quality. For health economic studies, we used the Consolidated Health Economic Evaluation Reporting Standards (CHEERS Checklist [35]) to assess specialty study quality. Quality assessment was independently conducted by 2 authors (EM, GG, or JS), and health economic studies were assessed by an expert in the field (BB). In the event of discrepant assessment results, we resolved them for consensus through (1) seeking additional data or clarification from study authors when possible and (2) reviewing during team discussions.

For collective data pooled into meta-analyses, we assessed the quality of the evidence for each analysis using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach [31,36]. One of the 4 levels—high, moderate, low, or very low—were assigned to the overall quality of evidence for each outcome according to factors including a within-study risk of bias (methodological quality), directness of evidence, heterogeneity, precision of effect estimates, and risk of publication bias.

Results

The search initially retrieved 25,586 records. A stepwise process of screening titles, abstracts, and full-text papers against our eligibility criteria was used to identify 417 full-text articles for the final screening stage. Of these, 21 papers including 19 discrete study data sets were included [5,37-54]. One RCT paper [55] included partial data from a previous paper that reported on the same tailored eHealth intervention investigated with the same sample in the Netherlands [41]; hence, we only used data extracted from the latter, which also reported trial registration details. Similarly, we included the main paper out of the 2 that reported on the same trial of a digital public mental health program in Hong Kong [43,56]. Results from the search process

are shown in Figure 1, and a summary of the included studies is presented in Table 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

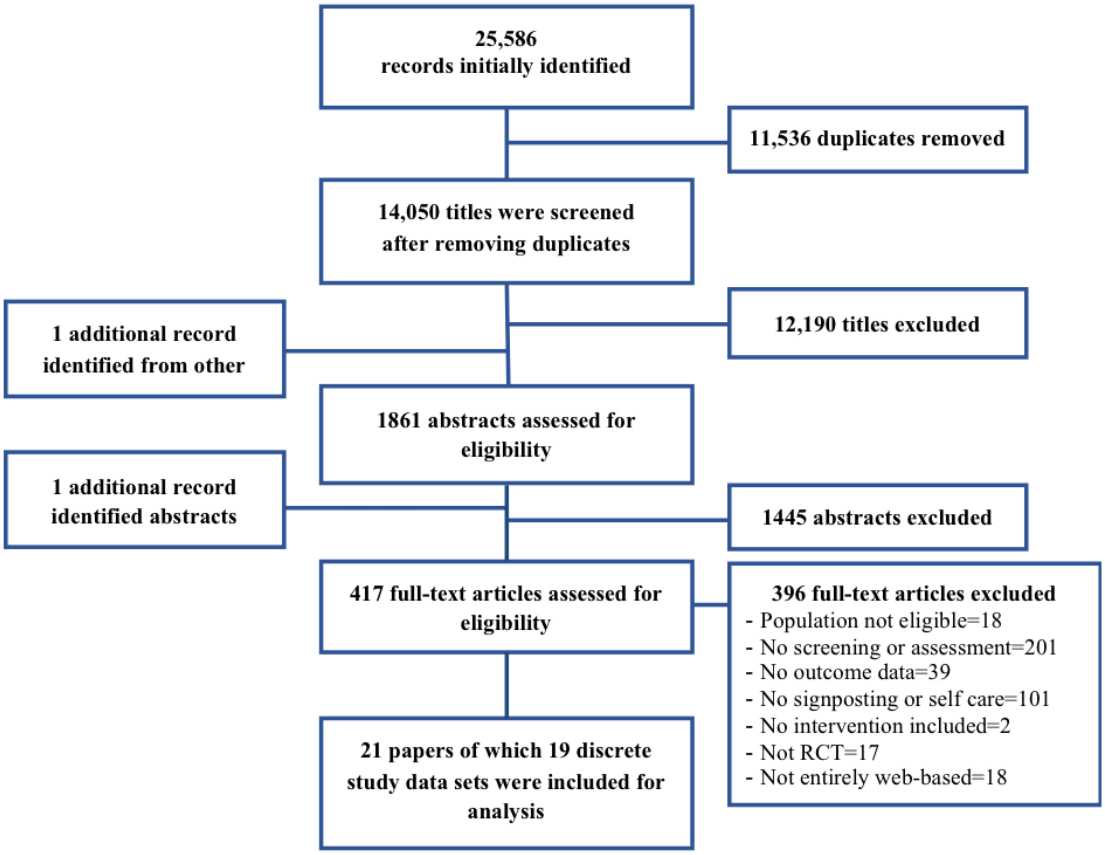


Table 1. Summary of the included studies.

Reference; country	Targeted CMD[a]	Intervention approach (n)[b], gender distribution[c] (%F/M/other), and age	Comparisons (n), gender distribution[c] (%F/M/other), and age	Outcomes with validated measures
Batterham [39]; Australia	Depression and anxiety	Web-based assessment with tailored feedback and health information on depression or anxiety, respectively (n=1342, US[d])	No tailored feedback, just generic advice (n=1431, US)	AHSQ[e], PHQ-9[f], GHSQ[g], AQoL-4D[h]
Batterham [38]; Australia	Depression, anxiety, substance use, and suicidal ideation	FitMindKit, a tailored feedback with 10 core and 8 elective behavior therapy modules based on symptom profile (n=66, 86% F, 14% M, US)	Static FitMindKit—with no tailored feedback (n=62); attention control, a web-based HealthWatch program (n=62, 86% F, 14% M, US)	PHQ-9, GAD-7[i], PADIS[j], SOPHS[k], AUDIT[l], DUDIT[m], SIDAS[n]
Billings [40]; United States	Stress, depression, anxiety, and substance abuse	Stress and Mood Management, a web-based multimedia health promotion CBT[o] program (n=154, 71% F, 29% M, US)	Waitlist control (n=155, 71% F, 29% M, US)	SDS[p], PNAS[q], CES-D[r], BAI[s], ATSPPPH-SF[t], SRSQ[u], WLQ[v]
Chiauzzi [42]; Unites States	Stress, anxiety, and health behaviors	MyStudentBody, a stress-tailored motivational feedback upon completion of 5 web-based questionnaires (n=80, 48% M, 52% F, US)	Control website with no tailoring (n=80); no treatment control (n=80, 48% M, 52% F, US)	PSS-10[w], HPLP-II[x], CAS[y]
Eimontas [51]; Lithuania	Adjustment disorder	BADI[z], a web-based unguided self-help psychological intervention for ICD-11[aa] adjustment disorder (n=516, 82% F, 18% M, mean age 35 years)	BADI-T[ab] group—BADI intervention augmented with web-based therapist support (n=561, 82% F, 18% M, mean age 35 years)	ADNM-8[ac], WHO-5[ad]
Eimontas [50]; Lithuania	Adjustment disorder	BADI, a web-based unguided self-help psychological intervention for ICD-11 adjustment disorder (n=156, 82% F, 18% M, mean age 35 years)	Waitlist control (n=128, 82% F, 18% M, mean age 35 years)	ADNM-8, WHO-5
Farrer [49]; Australia	Depression and anxiety	UVC[ae], a multicomponent, transdiagnostic web-based mental health program designed for university students (n=102, 78% F, 17% M, 5% other, mean age 22 years)	Waitlist control (n=98, 78% F, 17% M, 5% other, mean age 22 years)	PHQ-9, GAD-7, SOPHS, K10[af], EURO-HIS 8[ag], GSE-10[ah], CSEI[ai], AT-SPPH-SF
Fulmer [48]; Unites States	Depression, anxiety	Tess[aj], 2 versions of an integrative psychological artificial intelligence chatbox fully automated intervention for 2 weeks with daily check-ins (n=24) or 4 weeks with biweekly check-ins (n=26, 70% F, 29% M, 1% other, mean age 23 years)	Attention control—link to an electronic book on depression (n=24, 70% F, 29% M, 1% other, mean age 23 years)	PHQ-9, GAD-7, PANAS
Haga [53]; Norway	Perinatal depressive symptoms	Mamma Mia, fully automated preventive intervention for perinatal depressive symptoms and usual care (n=678, 100% F, mean age 31 years)	Treatment as usual (up to 14 consultations at well-baby clinic, n=664, 100% F, mean age 31 years)	EPDS[ak]
Ketalaar[al] [41]; the Netherlands	Stress, functioning, and fatigue	Screening and personalized feedback followed by tailored offer of self-help e-mental health intervention based on symptoms (n=178, 83% F, 17% M, mean age 37 years)	Waitlist control (n=188, 77% F, 23% M, mean age 42 years)	NWFQ[am], 4DSQ[an], QEEW[ao], WAI[ap], IES[aq] (Dutch)
Ludtke [47]; Germany	Depression	Be Good to Yourself CBT-based mobile self-help app (n=44, 82% F, 18% M, mean age 41 years)	Waitlist control (n=44, 75% F, 25% M, mean age 45 years)	PHQ-9, Rosenberg Self-Esteem Scale, WHOQOL-BREF[ar], URICA[as], CSQ-8[at]
Mak[al] [43]; Hong Kong	Psychological distress	Living With Heart App providing a mindfulness-based program (n=703) or a self-compassion program (n=705, 73% F, 27% M, mean age 34 years)	Web-based cognitive behavioral psychoeducation program (n=753, 73%, 27% M, mean age 34 years)	WHO-5, K6[au], MAAS[av], Self-Compassion Scale

Reference; country	Targeted CMD[a]	Intervention approach (n)[b], gender distribution[c] (%F/M/other), and age	Comparisons (n), gender distribution[c] (%F/M/other), and age	Outcomes with validated measures
Moberg [5]; United States	Stress, anxiety, and depression	Pacifica, fully automated app for the self-management of stress, anxiety, and depression app (n=253, 74% F, 23% M, 3% other, mean age 30 years)	Waitlist (n=247, 75% F, 23% M, 2% other, mean age 30 years)	DASS-21[aw], PHQ-8[ax], GAD-7, GSE-10
Proudfoot [37]; Australia	Depression, anxiety, and stress	myCompass—a fully automated, non–therapist-supported psychological treatment tailored to the user (n=472, 70% F, 30% M, mean age 39 years)	Waitlist (n=230, 70% F, 30% M, mean age 38 years); attention control (n=248, 70% F, 30% M, mean age 40 years)	DASS-21, WSAS[ay]
Querstret [54]; United Kingdom	Stress, depression, and anxiety	Be Mindful Online —a web-based mindfulness-based cognitive therapy course (n=60, 81% F, 19% M, mean age 40 years)	Waitlist (n=58, 81% F, 19% M, mean age 42 years)	Symptom severity
Solomon [52]; Australia	Depression, anxiety, and stress	MyCompass —same as Proudfoot et al [37]. Sample size not applicable because of modeling and simulation used (US)	Antidepressant medication or CBT (US)	Quality-adjusted life years
Stallman [46]; Australia	Psychological distress	My Coping Plan app, offering automated support to building an individualized coping plan (n=28, 91% F, 9% M, mean age 29 years)	Waitlist (n=28, 91% F, 9% M, mean age 29 years)	K10, CI[az], WHO-5
Viskovich [44]; Australia	Psychological distress	YOLO[ba] program, a web-based multimedia acceptance and commitment therapy with 4 modules, offered in 3 derivatives: (1) complete 1 module per week but fully flexible (n=40, 75% F, 25% M, mean age 27 years),	(2) to complete the YOLO program in 4 weeks (n=43, 75% F, 25% M, mean age 27 years) and (3) to access a YOLO module 3 days after completion of the previous module (n=47, 75% F, 25% M, mean age 27 years)	DASS-21, MHC-SF[bb], SCS-SF[bc], SWLS[bd], DDQR[be], AAQ-II[bf], CFQ[bg], PVQII[bh] education values subscale, ELS[bi], MAAS, SUS[bj]

Reference; country	Targeted CMD[a]	Intervention approach (n)[b], gender distribution[c] (%F/M/other), and age	Comparisons (n), gender distribution[c] (%F/M/other), and age	Outcomes with validated measures
Viskovich [45]; Australia	Depression, anxiety, and stress	YOLO program—a multimedia acceptance and commitment therapy with 4 modules, as above (n=596, 68% F, 32% M, mean age 27 years)	Waitlist (n=566, 68% F, 32% M, mean age 27 years)	DASS-21, MHC-SF, SCS-SF, SWLS, AAQ-II, CFQ, PVQII education values subscale, ELS, MAAS, SUS

[a]CMD: common mental disorder.
[b](n): sample size.
[c]Gender distribution: percentage of female, male, or other/unspecified participants.
[d]US: unspecified.
[e]AHSQ: Actual Help Seeking Questionnaire.
[f]PHQ-9: Patient Health Questionnaire-9 items.
[g]GHSQ: General Help Seeking Questionnaire.
[h]AQoL: Assessment of Quality of Life.
[i]GAD-7: Generalized Anxiety Disorder-7.
[j]PADIS: Panic Disorder Screener.
[k]SOPHS: Social Phobia Screener.
[l]AUDIT: Alcohol Use Disorders Identification Test.
[m]DUDIT: Drug Use Disorders Identification Test.
[n]SIDAS: Suicidal Ideation Attribution Scale.
[o]CBT: cognitive behavioral therapy.
[p]SDS: Symptoms of Distress scale.
[q]PANAS: positive and negative affect schedule.
[r]CES-D: Centre for Epidemiologic Studies Depression Scale.
[s]BAI: Beck Anxiety Inventory.
[t]ATSPPPH-SF: Attitudes Towards Seeking Professional Psychological Help Scale-Short Form.
[u]SRSQ: Stress Relief Strategies Questionnaire.
[v]WLQ: Work Limitations Questionnaire.
[w]PSS: Perceived Stress Scale.
[x]HPLP-II: Health-Promoting Lifestyle Profile II.
[y]CAS: College Adjustment Scales.
[z]BADI: Brief Adjustment Disorder Intervention.
[aa]ICD-11: International Classification of Diseases 11th Revision.
[ab]BADI-T: Brief Adjustment Disorder Intervention – Therapist support.
[ac]ADNM-8: Brief Adjustment Disorder New Model Scale.
[ad]WHO-5: World Health Organization well-being index.
[ae]UVC: Uni Virtual Clinic.
[af]K10: Kessler 10 items Psychological Distress Scale.
[ag]EURO-HIS 8: shortened version of the World Health Organization Quality of Life Instrument-Abbreviated Version.
[ah]GSE-10: General Self-Efficacy Scale.
[ai]CSEI: College Self-Efficacy Inventory.
[aj]Tess: name of the intervention.
[ak]EPDS: Edinburgh Postnatal Depression Scale.
[al]Denotes the major publication for the same study sample and data.
[am]NWFQ: Nurses Workforce Functioning Questionnaire.
[an]4DSQ: Four Dimensional Symptoms Questionnaire.
[ao]QEEW: questionnaire on the experience and evaluation of work.
[ap]WAI: Work Ability Index.
[aq]IES: Impact of Event Scale.
[ar]WHOQOL-BREF: World Health Organization Quality of Life Instrument-abbreviated version.
[as]URICA: University of Rhode Island Change Assessment.
[at]CSQ-8: client satisfaction questionnaire.
[au]K6: Kessler 6-Item Psychological Distress Scale.

[av]MAAS: Mindful Attention and Awareness Scale.
[aw]DASS-21: Depression Anxiety and Stress Scales-21.
[ax]PHQ-8: Patient Health Questionnaire-8 items.
[ay]WSAS: Work and Social Adjustment Scale.
[az]CI: coping index.
[ba]YOLO: You Only Live Once.
[bb]MHC-SF: Mental Health Continuum-Short Form.
[bc]SCS-SF: Self-Compassion Scale-Short Form.
[bd]SWLS: Satisfaction with Life Scale.
[be]DDQR: Daily Drinking Questionnaire Revised.
[bf]AAQ-II: Acceptance and Action Questionnaire II.
[bg]CFQ: Cognitive Fusion Questionnaire.
[bh]PVQII: Personal Value Questionnaire II.
[bi]ELS: Engaged Living Scale.
[bj]SUS: System Usability Scale.

Overview of the Included Studies

Overall, the included studies covered 6223 participants in intervention conditions and 5797 participants in comparison conditions. Nearly half of the studies (8/19, 42%) including a cost-effectiveness study[37-39,44-46,49,52] were conducted in Australia. Four (4/19, 21%) studies were conducted in the United States [5,40,42,48]. The remaining studies took place in Europe, including Lithuania [50,51], the United Kingdom [54], the Netherlands [41], Germany [47], and Norway [53]. Finally, 1/19 (5%) study originated from Hong Kong, China [43].

Studies recruited adults with subclinical or mild symptoms of CMDs among the general population in the community through social media (Facebook and Twitter) advertisements [5,37,38]. Nearly half of the studies aimed at promoting positive well-being and targeted users with some indication of clinical symptoms, including university students [42,44-46,48,49] and the general public who were interested in self-care to promote well-being [5,38,39,43,47,54]. The remaining studies targeted populations with an increased risk of mental health morbidities either because of work-related stress or health conditions. These included nurses and allied health professionals [41], technology company employees [40], and pregnant or postpartum women and their partners to prevent or manage postpartum depression [53]. Very few studies targeted populations with symptoms of CMDs that were above the clinical threshold. The exceptions included studies trialing an electronic mental health treatment for those with mild-to-moderate depression [37] or marked adjustment disorder symptoms [50,51].

Across the included studies, female participants comprised, on average, three-fourth of the overall sample (from 66% to 90%). Participants were largely in their early adulthood (aged 20-30 years). Few studies provided details on other sociodemographic characteristics, beyond age and gender, of the participants, an exception being ethnicity for studies from the United States and Australia. One trial from the United States on university students reported that half of the participants were Asians (50%), outweighing those who were Whites (43%), with only 3% of African Americans or Black participants [48]. The other studies from the United States showed instead a majority of Whites over Asian and Black or African American participants: the percentages were 59%, 18%, and 13%, respectively, in another study on students [42]; 82%, 4%, and 10%, respectively, in a further US app trial [5]; and 65%, 23%, and 7%, respectively, in a web-based stress management program [40]. In Australia, a trial reported that about half (53%) of the participants were Whites, 15% were Asians, 3% were Africans, 0.8% were Aboriginal or Torres Strait Islanders, and a further 17% preferred not to provide ethnicity details [44]. Finally, in an Australian study of a university student virtual clinic, 65% of the participants were Whites, 28% were Asians, 1% were Africans, and 1% were Aboriginal, Torres Strait, and Pacific Islanders [49].

Intervention Design and Features

Sixteen digital interventions were reported in the 19 included studies: 1 brief adjustment disorder intervention was trialed in 2 RCTs in Lithuania [50,51], a web-based acceptance and commitment therapy intervention was tested in 2 studies in Australia [44,45], and a web-based intervention targeting mild-to-moderate depression was reported in both an effectiveness trial [37] and a health economic study, [52] also in Australia.

In terms of intervention approaches, most offered web-based screening using various validated CMD measures followed by automatically generated (individualized) feedback, including classifying the users' CMD symptom levels from no risk to high risk. All interventions included offered signposting to relevant services or resources, including self-management strategies such as mood or progress monitoring; relaxation strategies including meditation, mindfulness, and self-compassion; goal setting; journaling; and activating exercises. Some interventions further used the screening results to assign individuals to a relevant web-based mental health treatment pathway using artificial intelligence (AI) algorithms [39,48].

The mode of delivery and design features of the interventions are summarized in Table 2. Most were delivered through a web-based portal allowing users to access it through any device with a web browser [37-39,44,49,50,53,54]. Some were specifically developed and trialed as mobile apps [5,43,46,47]. There was 1 fully AI chat box [48]. All included trials tested digital self-care interventions, often incorporating

psychoeducation [39,40,53] and various other psychological intervention modalities. The most commonly employed intervention strategies included mindfulness [5,43,47,50,54], compassion, CBT [5,47,50], acceptance and commitment therapy [44,45], motivational interviewing [48], and positive psychology mobilizing the individual's strengths [46,48]. Five interventions included an interactive forum where users can exchange discussions with one another [5,38,39,43,46].

Table 2. Mode of delivery used by the included interventions.

References	Delivery platform				Social network	Treatment recommendations				Cost		
	App	Computers	Both	Other		Self-care	Informal support	Formal service	Other	Free	Paid	Not stated
Batterham et al [39]	N/A[a]	N/A	X[b]	X	X	X	X	N/A	X	N/A	N/A	X
Batterham et al [38]	N/A	X	N/A	N/A	X	X	X	N/A	X	X	N/A	N/A
Billings et al [40]	N/A	X	N/A	N/A	N/A	X	N/A	X	X	X	N/A	N/A
Chiauzzi et al [42]	N/A	X	N/A	N/A	N/A	X	N/A	N/A	X	N/A	N/A	X
Eimontas et al[c] [50]	N/A	N/A	X	N/A	N/A	X	N/A	N/A	N/A	X	N/A	N/A
Farrer et al [49]	N/A	N/A	X	N/A	N/A	X	N/A	N/A	X	N/A	N/A	X
Fulmer et al [48]	N/A	N/A	N/A	X	N/A	X	N/A	N/A	X	X	N/A	N/A
Haga et al [53]	N/A	N/A	X	N/A	N/A	X	N/A	N/A	N/A	X	N/A	N/A
Ketelaar et al [41]	N/A	X	N/A	N/A	N/A	X	N/A	N/A	X	N/A	N/A	X
Mak et al [43]	N/A	N/A	X	N/A	X	X	N/A	N/A	X	X	N/A	N/A
Ludtke et al [47]	X	N/A	N/A	N/A	N/A	X	N/A	N/A	X	N/A	N/A	X
Moberg et al [5]	X	N/A	N/A	N/A	X	X	X	N/A	N/A	N/A	N/A	X
Proudfoot et al[c] [37]	N/A	N/A	X	N/A	N/A	X	N/A	N/A	X	X	N/A	N/A
Querstret et al [54]	N/A	X	N/A	N/A	N/A	X	N/A	N/A	N/A	X	N/A	N/A
Stallman et al [46]	X	N/A	N/A	N/A	X	X	X	X	N/A	X	N/A	N/A
Viskovich et al[c] [45]	N/A	X	N/A	N/A	N/A	X	N/A	N/A	N/A	X	N/A	N/A

[a]N/A: not applicable.

[b]X: indicated feature.

[c]Data from Eimontas 2018 [50], Proudfoot 2013 [37], and Viskovich 2019 [45] used for intervention description here.

Limited details of the digital intervention designs and ICT features were reported. An explicit theoretical basis underpinning the design and delivery integrating algorithm and web-based behavioral change techniques was generally lacking. Across studies, only a few web-based behavioral change techniques were explicitly adopted by the interventions, including provision of feedback on performance [39,51], goal setting [46], and prompts for self-monitoring of behavior and progress [37,43,49]. Intervention duration and intensity varied widely across studies, with most interventions lasting 4 weeks [43,44,46], a few lasting 3 months [38,39,41], and the longest lasting 11 months [53]. Most interventions did not stipulate the minimum usage requirement and recommended that the users use the intervention as preferred [38]. Some interventions had a set number of modules to be undertaken over a set time frame. However, these did not necessarily translate into minimum usage requirement, intervention duration, or intensity [37,38,53].

Study Design and Outcome Measures

All but one of the included studies used an individual-level RCT design. Only 1 study used a cluster RCT design at a ward level where nurses and allied health professionals were allocated according to their work base within a hospital in the Netherlands [41]. All studies examined digital intervention effectiveness,

with 1 including a health economic modeling study comparing cost-effectiveness of the digital intervention with antidepressant medication (as treatment as usual) or CBT for mild-to-moderate depression in Australia [37,52]. The comparison conditions used in the included RCTs were grouped into (1) inactive controls and (2) active controls. The former includes usual care delivered using a conventional medium [53] or waitlist controls [5,37,40,41,45-47,49,50,54]. The latter comprises attention controls (eg, static websites with information or an electronic book [37-39,48]). One trial included 3 arms, comparing the digital intervention with both attention and waitlist controls [37]; we used such data in separate analyses. Two 3-arm trials compared 3 different formats of the same digital intervention head-to-head with no other comparison groups comprising nondigital elements [43,44]. No usable comparison data could be extracted for analyses. Data from 1 trial that compared an entirely web-based self-care intervention for university students with a version of the intervention augmented with therapist input also delivered through its web-based platform was not usable in the analysis [51].

All trials that investigated the effectiveness of digital interventions used outcome measures of symptoms of mental illness, including stress, anxiety, depression, and general distress. Three studies measured well-being [45,46,50] and only 1 study measured quality of life at post intervention and 3-month follow-up, respectively [39,47]. Help-seeking attitude [40] and service use [39] were each measured by 1 study at each time point. Work or general functioning was assessed in 3 studies [37,40,41]. Two studies reported coping as an outcome, with each focused on overall coping [46] or negative coping [40]. Satisfaction with intervention, if assessed, focused only on the intervention group participants and the measures or tools used were often unvalidated or devised by the study teams on an ad hoc basis [44,45,50].

In terms of intermediate outcomes, 1 study measured knowledge of symptoms of CMD, prevention, and treatment [40]. Use of health-promoting behaviors was covered in only 1 study [42], although many reported therapy-specific measures to assess engagement with therapy approaches (eg, compassion, cognitive flexibility, willingness to change). Although behavior change techniques, most often goal setting, prompts for self-monitoring, or action planning, have been reported to form part of the intervention design [37,43,46], no data on uptake of recommendations or behavioral activation outcomes were available if measured.

Overall Study Quality

Our evaluation of the study quality and the comparison of the global ICROMS score of each study against the ICROMS minimal score requirement is presented in Table 3. The ICROMS global quality scores ranged from 14 to 29; 6 (33%) trials were rated below the minimum score of 22. Although the RCTs were published relatively recently, some did not fully adhere to the CONSORT or CONSORT-eHealth checklist [40]. Many of the RCTs did not publish their protocols or prospectively register the study on trial databases to provide details on the intervention design and required minimum intervention exposure (ie, per-protocol use) or state a priori primary outcomes [42,45,46]. Although randomization and allocation using a computerized or web-based system were often cited, details on the randomization sequence generation and allocation concealment were often minimal if at all reported [5,43,44,47,50,51]. Given that waitlist control or usual care was most commonly used as the comparator, it was not feasible to blind the participants, although there were few exceptions [49,54]. Although outcome data collection using web-based questionnaires with the participants directly reduced bias in assessment, limited considerations were conveyed to establish whether the researchers or trial statisticians who conducted the data analysis were blinded to group allocation [5,43,44,47,50,51]. Nonetheless, the most significant quality issue identified here concerns retention and completion rates in digital health intervention trials. An intention-to-treat analysis was not always used, and there was a lack of available data for noncompleters [40,41,44,45,50,51]; these quality issues might bias the study results and overall evidence. Another area of potential bias lies in reporting or ethical considerations, as not all studies reported their funding sources and conflicts of interest. Furthermore, some trialists reported a digital intervention produced by commercial enterprises in which they had a financial interest [5,42,48].

We rated the quality of the health economic study [52] as satisfactory according to CHEERS [35]. The paper addressed 18 of the 24 (67%) CHEERS quality criteria, including clear reporting of method, analysis, results, and discussions. Four checklist criteria were deemed irrelevant in this study (eg, not a single study-based economic evaluation and hence no such study parameters). Quality criteria that were not addressed were discount rates used for costs and outcomes (if any) and justification of the choice of model used.

Table 3. Quality assessment of the included studies using integrated criteria for review of multiple study designs (ICROMS).

References (first author only)	Study design[a]	Aims and justifications	Sequence generation and allocation concealment	Outcome measures and blinding	Follow-up	Other study aspects	Analytical rigor	Other considerations	Global quality score
Batterham et al [38]	RCT[b]	2	3	3	5	2	2	11	28[c]
Batterham et al [39]	RCT	2	4	4	6	2	2	8	28[c]
Billings et al [40]	RCT	0	2	1	3	2	1	5	14[d]
Chiauzzi et al [42]	RCT	2	4	2	5	2	2	7	24[c]
Eimontas et al [50]	RCT	1	1	2	5	2	1	7	19[d]
Eimontas et al [51]	RCT	2	1	4	5	2	1	8	23[c]
Farrer et al [49]	RCT	2	4	4	5	2	2	9	28[c]
Fulmer et al [48]	RCT	2	2	4	4	2	1	4	19[d]
Haga et al [53]	RCT	2	3	4	6	2	2	9	28[c]
Ketelaar et al [41]	cRCT[e]	2	4	2	6	2	0	6	22[c]
Ludtke et al [47]	RCT	1	2	4	6	2	2	7	24[c]
Mak et al [43]	RCT	2	2	4	6	2	2	8	26[c]
Moberg et al [5]	RCT	2	1	4	4	2	1	5	19[d]
Proudfoot et al [37]	RCT	2	3	2	5	2	2	8	24[c]
Querstret et al [54]	RCT	2	4	6	6	2	2	7	29[c]
Stallman [46]	RCT	2	4	4	5	2	1	7	25[c]
Viskovich and Pakenham [44]	RCT	1	1	4	4	1	1	6	18[d]
Viskovich and Pakenham [45]	RCT	2	2	3	4	1	1	6	19[d]

[a]ICROMS minimal score requirement for (cluster) randomized controlled trial=22.

[b]RCT: randomized controlled trial.

[c]Comparison against minimal score requirement: above requirement.

[d]Comparison against minimal score requirement: below requirement.

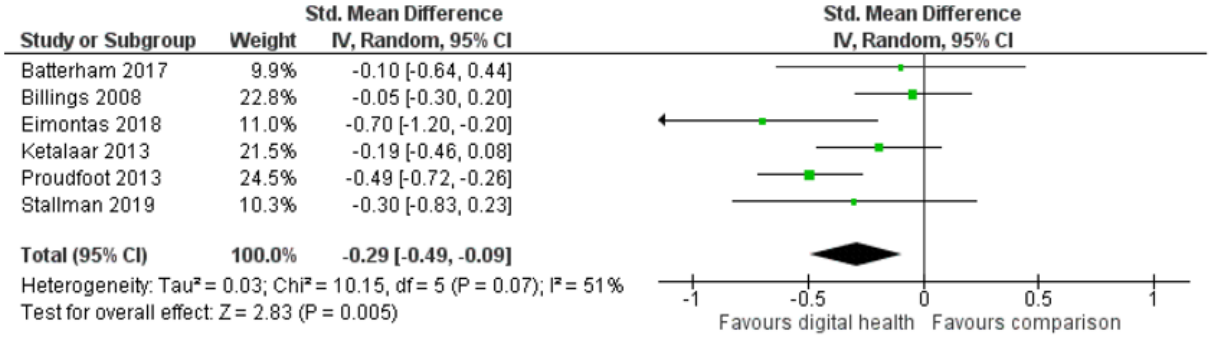
[e]cRCT: cluster randomized controlled trial.

Effectiveness

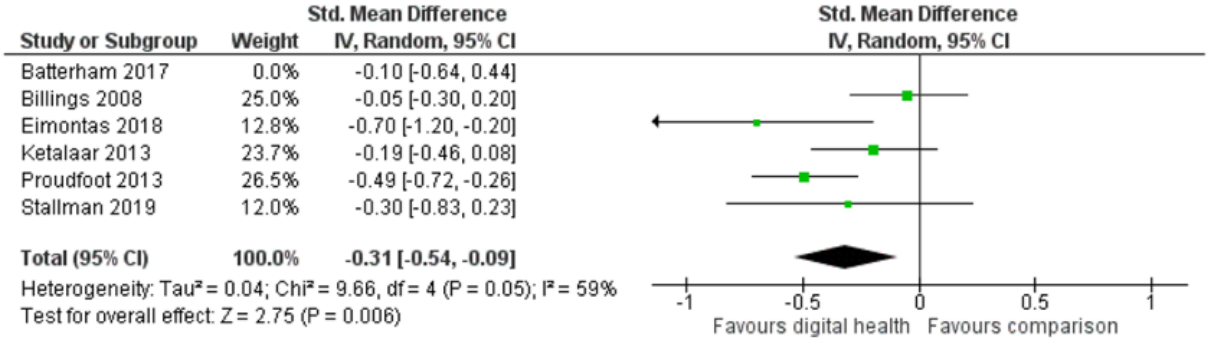
Six RCTs [37,38,40,41,46,50] reported outcomes using measures of mental illness symptoms (as a composite measure encompassing depression, anxiety, and distress or psychological distress) at the end of the intervention use. These studies examined the effectiveness of tailored digital interventions compared with waitlist controls [37,40,41,46,50] or attention controls [38]. The meta-analysis including these 6 studies showed an overall significant small effect of digital interventions compared with controls in reducing the symptoms of mental illness (6 RCTs; n=992; SMD −0.29; 95% CI −0.49 to −0.09; I^2=51%; random effects; GRADE quality of evidence=moderate). Comparing digital interventions with waitlist controls only using data from 5 trials led to a similar

result favoring digital interventions (5 RCTs; n=939; SMD −0.31; 95% CI −0.54 to −0.09; I^2=59%; random effects; GRADE quality of evidence=low). Only 2 trials provided data for comparing digital interventions with attention controls [37,38]. The meta-analysis including these data still yielded a significant result favoring digital intervention (2 RCTs; n=374; SMD −0.31; 95% CI −0.52 to −0.10; I^2=0%; fixed effect; GRADE quality of evidence=very low). Figure 2 shows the meta-analyses on the outcome of symptoms of mental illness.

Figure 2. Meta-analysis on outcome of mental illness symptoms 2(a) Comparison of digital interventions with any comparators using all available data.

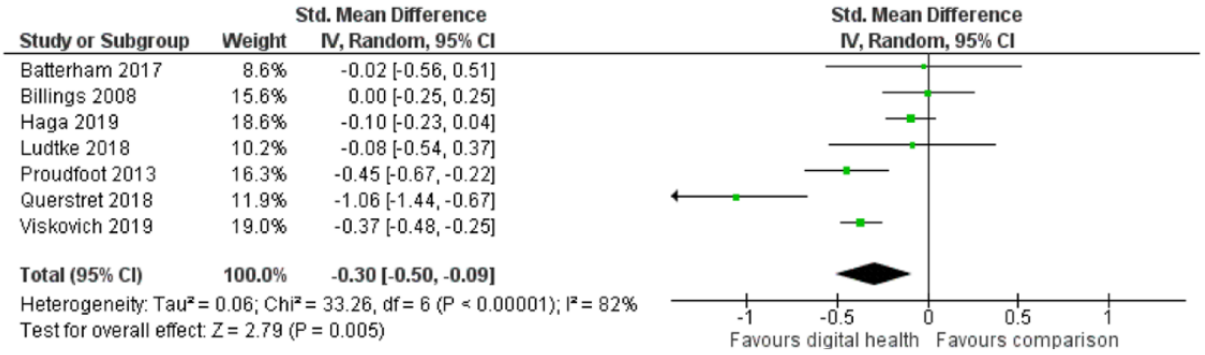


2(b) Comparison of digital interventions with waitlist controls only



Seven studies measured participants' depressive symptoms [37,38,40], comparing digital interventions with inactive controls [37,40,45,47,53,54] or attention controls [38]. Digital interventions showed a small but significant positive effect over any comparison (7 RCTs; n=2824; SMD −0.30; 95% CI −0.50 to −0.09; I^2=82%; random effects; GRADE quality of evidence=low). Heterogeneity of this meta-analysis was high: 3 were European studies, including 1 focusing on postnatal depression in new mothers through a year-long intervention across the perinatal period [53]; 3 were conducted in Australia, comprising nearly half of the total participants in this analysis; and the remainder was conducted in the United States. Figure 3 shows the meta-analysis of depressive symptoms.

Figure 3. Meta-analysis on outcome of depressive symptoms.



A meta-analysis of participants' anxiety symptoms from 5 studies produced similar positive results favoring digital interventions over inactive or attention controls (5 RCTs; n=1893; SMD −0.37; 95% CI −0.65 to −0.08; I^2=84%; random effects; GRADE quality of evidence=low). The high heterogeneity is likely because of diverse intervention, population, and methodological factors [37,38,40,45,54]. Figure 4 shows the meta-analysis of the anxiety symptoms. Three studies reported stress outcomes, but only data from 2 of these were used in the meta-analysis [41,45,54]. The analysis showed a significant positive effect over waitlist controls (2 RCTs; n=1280; SMD −0.43; 95% CI −0.54 to −0.32; I^2=94%; fixed effect; GRADE quality of evidence=very low). Of note, the heterogeneity of these 2 studies was high: 1 trialed a web-based mindfulness CBT for UK workers [54] and the other investigated a web-based acceptance and commitment therapy for university students in Australia [45]. Figure 5 shows the meta-analysis on stress outcomes.

Figure 4. Meta-analysis on outcome of anxiety symptoms.

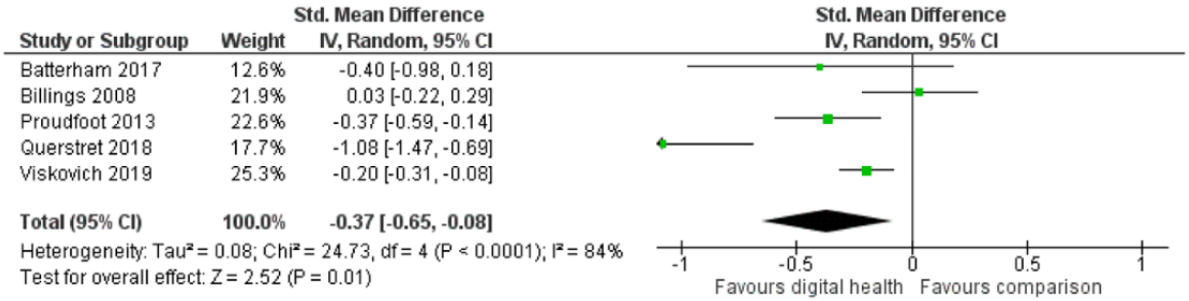
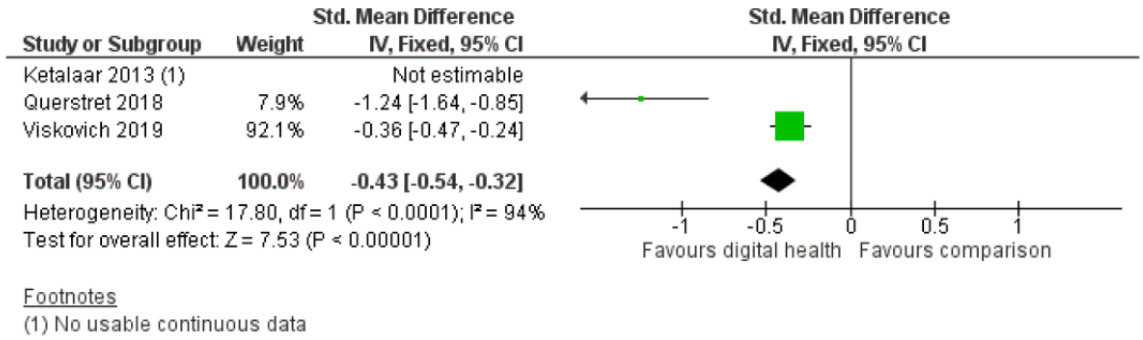
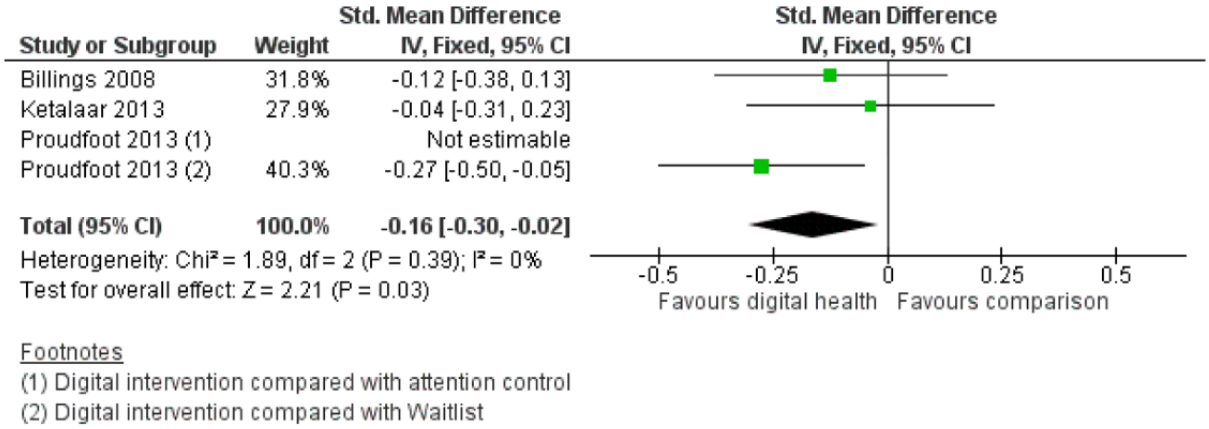


Figure 5. Meta-analysis on outcome of stress symptoms.



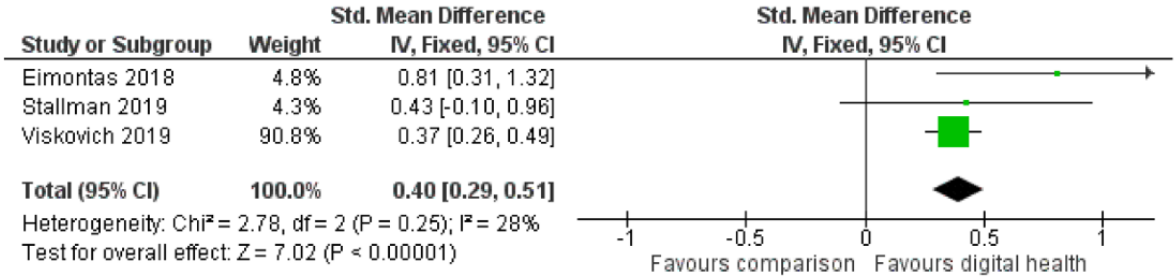
In terms of work and social functioning outcomes, 3 studies compared digital interventions with inactive controls [37,40,41], whereas 1 study included a second control group using attention controls [37]. Results comparing digital interventions with any comparators were equivocal across groups (3 RCTs; n=792; SMD −0.13; 95% CI −0.27 to 0.01; I^2=0%; fixed effect; GRADE quality of evidence=low). However, when comparing digital interventions with inactive controls only, digital interventions showed a significant although small effect over waitlist controls (3 RCTs; n=795; SMD −0.16; 95% CI −0.30 to −0.02; I^2=0%; fixed effect; GRADE quality of evidence=low). Figure 6 provides the meta-analysis on work and social functioning outcomes.

Figure 6. Meta-analysis on outcome of work and social functioning comparing digital interventions with inactive controls.



Three studies examined the effectiveness of digital interventions on well-being [45,46,50]; digital interventions delivered as web-based CBT, acceptance, and commitment therapy or mobile app showed a significant positive effect over waitlist controls (3 RCTs; n=1307; SMD 0.40; 95% CI 0.29 to 0.51; I^2=28%; fixed effect; GRADE quality of evidence=low). It is worth noting that this result was weighted heavily by 1 study conducted in Australia including >1100 university students [45]. Figure 7 provides the meta-analysis on the well-being outcome. Only 1 study measured participants' quality of life as an outcome when comparing digital intervention with waitlist control [47].

Figure 7. Meta-analysis on outcome of wellbeing comparing digital interventions with inactive controls.



Follow-Up Outcome Data

Follow-up data (beyond 3 months) were limited. Four studies [37-39,41] provided data, with 1 study delivering 2 active interventions that focused on depression or anxiety management compared with attention controls [39]. Meta-analyses using the available 3-month follow-up data revealed no significant differences in mental health, work, and social functioning outcomes between digital interventions and controls, active or inactive. Table 4 provides a summary of the meta-analysis results using the fixed effect model.

Table 4. Summary of meta-analyses on the 3-month follow-up outcome measures.

Outcome measures	Studies, n	Sample, N (n/n)[a]	SMD[b]	95% CI	I^{2c} (%)
Symptoms of mental illness	3	521 (194/327)	−0.12	−0.30 to 0.05	1
Depression	3	1209 (509/700)	−0.04	−0.15 to 0.08	0
Anxiety	3	1044 (431/613)	−0.20	−0.87 to 0.47	0
Work and social functioning	2	476 (171/305)	−0.13	−0.32 to 0.06	0

[a]Total number of participants included in the analysis (number of participants in digital interventions or number of participants in comparator groups).

[b]SMD: standardized mean difference.

[c]I^2: Statistical heterogeneity.

Health Economic Outcomes

No RCTs included a cost-effectiveness evaluation. One Australian RCT on a digital intervention, myCompass, designed to treat mild-to-moderate depression in the general population [37], was used as the basis of a decision analytic model [52]. The model employed a cost-utility framework to compare the costs of myCompass with each treatment as usual (antidepressant treatment and face-to-face CBT). The results of the model suggested that the myCompass intervention provided the highest net monetary benefit, and the authors concluded that digital interventions could provide a cost-effective route to treatment as part of a stepped care model [52].

Intermediate or Process Outcomes

There were no usable data available from RCTs on any of our prespecified intermediate or process outcomes (eg, uptake of self-care or informal support, health behavior change), precluding analysis on such outcomes in its own right or meta-regression on any association between intermediate and health outcomes. Some studies reported therapy-specific mediating measures, such as willingness to change measure in a CBT-based mobile app [47], self-compassion, or 5-facet mindfulness questionnaires in third-wave web-based CBTs [45,54]. These fell short of health behavior change outcomes and were therapy specific; therefore, we considered it inappropriate to compare such outcomes across studies.

Perceived Acceptability of Interventions

If reported, study findings on satisfaction were collated via self-devised measures or unvalidated survey post intervention use, lacking corroboration from validated outcome data and comparison with the control groups or any other interventions. No analysis of this outcome was feasible.

Discussion

Principal Findings

This comprehensive review included 18 RCTs and 1 health economic study on 16 interventions to examine the effectiveness of digital interventions that provided both initial assessment and treatment that emphasize self-care, using a web-based medium entirely. Fourteen of the included trials were only

published in the last 5 years, suggesting that despite the popularity of digital mental health interventions, rigorous research undertaken in this field is still emerging.

Our review identified some evidence to support the effectiveness of digital interventions in promoting well-being among university students [45,46,50] and in reducing symptoms of CMDs, depression, anxiety, stress, and promoting social and work functioning. These positive results on the symptoms of CMDs came from studies on nonclinical young adult samples (aged between their early 20s and 30s) among the general population with mild baseline symptoms [37,38,41,45,46,50,54]. It is highly plausible that the recruited study samples included a high proportion of people who had a low intensity of CMD symptoms that might not meet the threshold of clinical caseness or the need for conventional mental health interventions delivered by clinicians (eg, CBT or counseling). Uptake of interventions showed that the majority of the participants had a White background, with Asians being the second most frequent group reported and Blacks being third. Unfortunately, information about ethnicity of the participants was available for only 6 studies [5,40,42,44,48,49], limiting the analysis of plausible cultural determinants of digital health performance. Similar to conventional trials on psychological interventions delivered face-to-face, two-third of the study participants were female [16-18]. Furthermore, some of the included studies were designed primarily as a mental health promotion or preventative intervention, for example, for college students and new mothers [46,49,53]. Despite this aim, there was, in general, a lack of focus on positive psychological outcomes such as well-being or quality of life. Furthermore, there may be a ceiling effect with respect to the population means at baseline or study entry, leaving little room for improvement in the outcomes.

Most of the interventions examined were designed to be accessed and used autonomously by the users [5,37,43,48,50]. Commonly, users were advised to use the intervention flexibly to suit their own preference as much or as often as necessary or desired, although encouraged to make full use of the intervention elements (eg, forum, exercises, and monitoring) and content. A small proportion of interventions, however, guided their users through core content through a specific sequence (eg, to complete 4 modules in a predetermined order [38,47]) or over a specific timeframe (eg, 1 module a week or at a certain time point, such as 3 weeks after giving birth [53]). Although web-based recruitment across studies was largely successful, retention and completion rates reported across trials are concerning. With a couple of trials achieving retention rates ≥80% (eg, [46,48]) as exceptions, attrition rates range from 27% in an Australian trial of a digital depression and anxiety intervention [37] to 78% in a mobile app trial [5,43] and 87% in a web-based intervention [50] post intervention. Attrition at short-term follow-up is equally high (eg, 83% at 3-month follow-up [39]), whereas most of the included studies did not report follow-up beyond the immediate post-intervention time point. Furthermore, the low usage or adherence rate across trials was often cited to account partly for the equivocal results across groups [5,43,50], raising the possibility that no effect was because of low or no minimally sufficient treatment dosage. The low use of digital interventions also prompts doubts over the value of the automatic reminders (as emails or SMS, or mobile app prompts) integral to digital intervention design and delivery in the entirely self-guided treatment. Our review demonstrated that although many interventions sent automatic daily or weekly reminders or prompts to the participants directly, they were not responding accordingly. These issues, although consistent with the inherent challenges of conducting digital intervention trials [57,58], remain critical to be resolved. For any digital interventions to effect meaningful changes in their users, developers need to articulate the essential intervention elements and the required intervention exposure or usage to achieve that as a crucial part of the intervention design [59]. Most importantly, it is essential for digital interventions to optimize their engagement and facilitation strategies to ensure users get the intended benefits of the intervention when enjoying their autonomy in pursuing individualized treatment. The effects of reminders and prompts functions and indeed other communication strategies afforded by digital interventions should be carefully investigated to inform both the intervention and the study designs.

In addition to the paucity of research in the growing field of digital health interventions, we note some limitations in the included studies and the data they reported. Although all interventions examined included a self-assessment component, we found no data pertaining to the effectiveness or efficiency of the assessment function independent of the overall intervention, including their treatment component. Thus, no conclusions could be drawn on the impact of assessment on the users' initial engagement with the intervention, subsequent signposting based on AI, or the users' mental health outcomes. Follow-up data were sparse, limiting analysis on outcomes beyond the 3-month follow-up. The lack of reporting on intermediate outcomes and process evaluation data (if used) precluded any analysis to convey how digital interventions might work to instill health outcome changes [26,27]. Although some behavior change techniques were incorporated as intervention design (eg, prompts for self-monitoring or goal setting), data on the target health behavior change outcomes were generally not collected or not reported. Although it is often argued that digital interventions carry with them the benefit to be expanded and delivered to whole populations at a relatively low cost, no data were available regarding estimated cost-effectiveness and only 1 paper included economic modeling [52]. This, coupled with the unclear intervention design description, limits the generalizability of the results and the scope of replication and wider implementation.

Limitations

This review has several limitations. First, we are mindful that our results are synthesized from studies that reported different interventions targeting a wide range of populations, ranging from those promoting positive mental health to others identifying and treating mild-to-moderate depression. The results therefore fall short of identifying specific intervention designs (eg, with specific ICT features), which may be particularly effective for specific populations or groups with CMDs (symptoms). This approach also, in part, accounted for the high heterogeneity observed in the results of the meta-analysis. Second, given the limited amount of usable data included in the

analysis, especially in the follow-up timepoint, we conducted a meta-analysis using the fixed-effect model on endpoint mean score whenever <5 study data sets were available. Although the fixed-effect model is deemed most suitable for a meta-analysis including ≤5 studies, this approach is inferior in taking baseline measurements into consideration, which is particularly important in small trials [32]. We therefore downgraded the GRADE quality ratings accordingly [31].

Implications for Research and Practice

Although the results from the studies reviewed appear promising, they are limited in terms of generalizability to digital interventions scaled up for use by whole populations. For example, a key implication of the results for both research and practice is the need for economic evaluation of digital mental health services for general population samples [10,60]. Although the usual trial methods for cost-effectiveness evaluation would be informative, economic evaluation of the scaling up of digital interventions to whole populations is also important, as a key consideration for economic evaluation is the potential range of reach of digital services. Whereas widespread awareness and usage of a digital service may increase its cost-effectiveness, creating that awareness also has to be done in a cost-efficient manner. People who use non–digital health or other services can be informed of a digital service at these services, whereas those who only do so rarely or when in crisis but may benefit must be reached by other means. Given that the interventions are web based, the most obvious approach is to use social media advertising in response to mental health–related search terms [9,61]. An economic evaluation of scaling up requires a study of the costs to create awareness of the service and modeling methods using the usage data from the service. Such models must take into account as 1 of their assumptions the additional use of other services, both digital and nondigital, by some users. This is likely to vary as usage increases: as more people take up a digital intervention, the proportions that were previously using something else (and what that was) versus nothing is likely to change; similarly, the intervention's cost-effectiveness is likely to vary by demographic and clinical groups, which again changes with increasing levels of use.

Outside of a research or practice setting, the extent to which a digital service is trusted is important in addition to its usability [26,28,58]. One implication is the need for research into aspects that affect this trust and how this varies within the general population, for example, the need to use personal information to register before using the intervention and the use of health services or government logos [62].

Conclusions

Digital mental health interventions to assess and signpost people experiencing symptoms of CMDs appear to be acceptable to a sufficient number of people and have enough evidence for effectiveness to warrant further studies. We recommend that future studies incorporate economic analysis; much of the work in this area appears to rest on the untested assumption that digital interventions are cost-effective by their nature. We also suggest clarification of the theoretical models for interventions. Many apply therapies, such as CBT and psychoeducation, to a sample with milder problems than those presently receiving them and state their aims as including both reduction in symptoms and promotion of mental health. However, positive mental health outcomes such as mental well-being, self-esteem, self-efficacy, coping skills, or resilience are rarely used. This may obscure their effectiveness in the target population. Finally, process evaluation to assess implementation and mechanisms of action are needed to understand the outcomes reported, if needed, in a separate publication.

Acknowledgments

This systematic review forms part of a bigger public health program titled "London Digital Mental Wellbeing evaluation project," which is funded by the North East London Commissioning Support Unit and NHS Tower Hamlets Clinical Commissioning Group, United Kingdom.

Conflicts of Interest

None declared.

References

1. Mason EC, Andrews G. The use of automated assessments in internet-based CBT: the computer will be with you shortly. Internet Interv 2014 Oct;1(4):216-224. [doi: 10.1016/j.invent.2014.10.003]
2. Andrews G, Titov N. Is internet treatment for depressive and anxiety disorders ready for prime time? Med J Aust 2010 Jun 7;192(S11):S45-S47. [Medline: 20528709]
3. Titov N, Andrews G, Kemp A, Robinson E. Characteristics of adults with anxiety or depression treated at an internet clinic: comparison with a national survey and an outpatient clinic. PLoS One 2010 May 28;5(5):e10885 [FREE Full text] [doi: 10.1371/journal.pone.0010885] [Medline: 20526371]
4. Elison S, Ward J, Williams C, Espie C, Davies G, Dugdale S, et al. Feasibility of a uk community-based, etherapy mental health service in greater Manchester: repeated-measures and between-groups study of 'living life to the full interactive', 'sleepio' and 'breaking free online' at 'self help services'. BMJ Open 2017 Jul 20;7(7):e016392 [FREE Full text] [doi: 10.1136/bmjopen-2017-016392] [Medline: 28729322]

5. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 8;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
6. Andersson G, Hesser H, Hummerdal D, Bergman-Nordgren L, Carlbring P. A 3.5-year follow-up of Internet-delivered cognitive behavior therapy for major depression. J Ment Health 2013 Apr;22(2):155-164. [doi: 10.3109/09638237.2011.608747] [Medline: 21957933]
7. McManus S, Bebbington P, Jenkins R, Brugha T. Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014. NHS Digital. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/556596/apms-2014-full-rpt.pdf [accessed 2020-08-18]
8. Espie CA, Carl JR, Stott R, Henry AL, Miller CB. Digital medicine needs to work. Lancet 2018;392(10165):2694. [doi: 10.1016/S0140-6736(18)32519-4] [Medline: 30587365]
9. Greaves F, Joshi I, Campbell M, Roberts S, Patel N, Powell J. What is an appropriate level of evidence for a digital health intervention? Lancet 2019 Dec 22;392(10165):2665-2667. [doi: 10.1016/S0140-6736(18)33129-5] [Medline: 30545779]
10. Evidence Standards Framework for Digital Health Technologies. The National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2020-08-18]
11. Clement S, Brohan E, Jeffery D, Henderson C, Hatch SL, Thornicroft G. Development and psychometric properties the barriers to access to care evaluation scale (BACE) related to people with mental ill health. BMC Psychiatry 2012 Jun 20;12(1):36. [doi: 10.1186/1471-244x-12-36]
12. Powell J, Newhouse N, Boylan A, Williams V. Digital health citizens and the future of the NHS. Digit Health 2016;2:2055207616672033 [FREE Full text] [doi: 10.1177/2055207616672033] [Medline: 29942569]
13. Freeman D, Sheaves B, Goodwin GM, Yu L, Nickless A, Harrison PJ, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. Lancet Psychiatry 2017 Oct;4(10):749-758 [FREE Full text] [doi: 10.1016/S2215-0366(17)30328-0] [Medline: 28888927]
14. Mathiasen K, Riper H, Ehlers LH, Valentin JB, Rosenberg NK. Internet-based CBT for social phobia and panic disorder in a specialised anxiety clinic in routine care: results of a pilot randomised controlled trial. Internet Interv 2016 May;4:92-98 [FREE Full text] [doi: 10.1016/j.invent.2016.03.001] [Medline: 30135794]
15. Evidence Standards Framework for Digital Health Technologies. The National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2020-08-18]
16. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
17. Sijbrandij M, Kunovski I, Cuijpers P. Effectiveness of internet-delivered cognitive behavioral therapy for posttraumatic stress disorder: a systematic review and meta-analysis. Depress Anxiety 2016 Sep;33(9):783-791. [doi: 10.1002/da.22533] [Medline: 27322710]
18. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565 [FREE Full text] [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
19. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-delivered cognitive behavior therapy as a prequel to face-to-face therapy for depression and anxiety: a naturalistic observation. Front Psychiatry 2019;10:902 [FREE Full text] [doi: 10.3389/fpsyt.2019.00902] [Medline: 31998149]
20. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
21. Andersson G, Paxling B, Roch-Norlund P, Östman G, Norgren A, Almlöv J, et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. Psychother Psychosom 2012;81(6):344-355 [FREE Full text] [doi: 10.1159/000339371] [Medline: 22964540]
22. Lewis C, Roberts N, Bethell A, Robertson L, Bisson J. Internet-based cognitive and behavioural therapies for post-traumatic stress disorder (PTSD) in adults. Cochrane Database Syst Rev 2018 Dec 14;12:CD011710 [FREE Full text] [doi: 10.1002/14651858.CD011710.pub2] [Medline: 30550643]
23. Riva J, Malik K, Burnie S, Endicott A, Busse J. What is your research question? An introduction to the PICOT format for clinicians. J Can Chiropr Assoc 2012 Sep;56(3):167-171 [FREE Full text] [Medline: 22997465]
24. Sin J, Taylor A, Henderson C. A Systematic Review of Digital Mental Health Services That Promote Uptake of Self-care and Informal Support for Common Mental Disorders (CMDs) or People With Symptoms of CMDs. Centre for Reviews and Dissemination - University of York. 2017. URL: http://www.crd.yor.ac.uk/PROSPERO/display_record.php?ID=CRD42017079085 [accessed 2020-04-14]
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]

26. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
27. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
28. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
29. Webb M, Burns J, Collin P. Providing online support for young people with mental health difficulties: challenges and opportunities explored. Early Interv Psychiatry 2008 May;2(2):108-113. [doi: 10.1111/j.1751-7893.2008.00066.x] [Medline: 21352141]
30. Høifødt RS, Mittner M, Lillevoll K, Katla SK, Kolstrup N, Eisemann M, et al. Predictors of response to web-based cognitive behavioral therapy with high-intensity face-to-face therapist guidance for depression: a bayesian analysis. J Med Internet Res 2015 Sep 2;17(9):e197 [FREE Full text] [doi: 10.2196/jmir.4351] [Medline: 26333818]
31. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Oxford, UK: Wiley-Blackwell; 2011.
32. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
33. Zingg W, Castro-Sanchez E, Secci F, Edwards R, Drumright L, Sevdalis N, et al. Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). Public Health 2016 Apr;133:19-37. [doi: 10.1016/j.puhe.2015.10.012] [Medline: 26704633]
34. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
35. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, CHEERS Task Force. Consolidated health economic evaluation reporting standards (CHEERS) statement. Br Med J 2013 Mar 25;346:f1049 [FREE Full text] [doi: 10.1136/bmj.f1049] [Medline: 23529982]
36. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches the GRADE working group. BMC Health Serv Res 2004 Dec 22;4(1):38 [FREE Full text] [doi: 10.1186/1472-6963-4-38] [Medline: 15615589]
37. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
38. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VW. Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interv 2018 Jun;12:91-99 [FREE Full text] [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]
39. Batterham PJ, Calear AL, Sunderland M, Carragher N, Brewer JL. Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. BJPsych Open 2016 Jan;2(1):67-73 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001552] [Medline: 27703756]
40. Billings DW, Cook RF, Hendrickson A, Dove DC. A web-based approach to managing stress and mood disorders in the workforce. J Occup Environ Med 2008 Aug;50(8):960-968 [FREE Full text] [doi: 10.1097/JOM.0b013e31816c435b] [Medline: 18695455]
41. Ketelaar SM, Nieuwenhuijsen K, Gärtner FR, Bolier L, Smeets O, Sluiter JK. Effect of an e-mental health approach to workers' health surveillance versus control group on work functioning of hospital employees: a cluster-RCT. PLoS One 2013;8(9):e72546 [FREE Full text] [doi: 10.1371/journal.pone.0072546] [Medline: 24069148]
42. Chiauzzi E, Brevard J, Thum C, Thurn C, Decembrele S, Lord S. MyStudentBody-stress: an online stress management intervention for college students. J Health Commun 2008 Sep;13(6):555-572. [doi: 10.1080/10810730802281668] [Medline: 18726812]
43. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
44. Viskovich S, Pakenham KI. Pilot evaluation of a web-based acceptance and commitment therapy program to promote mental health skills in university students. J Clin Psychol 2018 Dec;74(12):2047-2069. [doi: 10.1002/jclp.22656] [Medline: 29962090]
45. Viskovich S, Pakenham KI. Randomized controlled trial of a web-based acceptance and commitment therapy (ACT) program to promote mental health in university students. J Clin Psychol 2020 Jun;76(6):929-951. [doi: 10.1002/jclp.22848] [Medline: 31468528]

46. Stallman HM. Efficacy of the my coping plan mobile application in reducing distress: a randomised controlled trial. Clin Psychol 2019 Jun 26;23(3):206-212. [doi: 10.1111/cp.12185]
47. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
48. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
49. Farrer LM, Gulliver A, Katruss N, Fassnacht DB, Kyrios M, Batterham PJ. A novel multi-component online intervention to improve the mental health of university students: randomised controlled trial of the uni virtual clinic. Internet Interv 2019 Dec;18:100276 [FREE Full text] [doi: 10.1016/j.invent.2019.100276] [Medline: 31890625]
50. Eimontas J, Rimsaite Z, Gegieckaite G, Zelviene P, Kazlauskas E. Internet-based self-help intervention for ICD-11 adjustment disorder: preliminary findings. Psychiatr Q 2018 Jun;89(2):451-460. [doi: 10.1007/s11126-017-9547-2] [Medline: 29124500]
51. Eimontas J, Gegieckaite G, Dovydaitiene M, Mazulyte E, Rimsaite Z, Skruibis P, et al. The role of therapist support on effectiveness of an internet-based modular self-help intervention for adjustment disorder: a randomized controlled trial. Anxiety Stress Coping 2018 Mar;31(2):146-158. [doi: 10.1080/10615806.2017.1385065] [Medline: 28984157]
52. Solomon D, Proudfoot J, Clarke J, Christensen H. E-CBT (MyCompass), antidepressant medication, and face-to-face psychological treatment for depression in Australia: a cost-effectiveness comparison. J Med Internet Res 2015 Nov 11;17(11):e255 [FREE Full text] [doi: 10.2196/jmir.4207] [Medline: 26561555]
53. Haga SM, Drozd F, Lisøy C, Wentzel-Larsen T, Slinning K. Mamma Mia - a randomized controlled trial of an internet-based intervention for perinatal depression. Psychol Med 2019 Aug;49(11):1850-1858 [FREE Full text] [doi: 10.1017/S0033291718002544] [Medline: 30191779]
54. Querstret D, Cropley M, Fife-Schaw C. The effects of an online mindfulness intervention on perceived stress, depression and anxiety in a non-clinical sample: a randomised waitlist control trial. Mindfulness (N Y) 2018;9(6):1825-1836 [FREE Full text] [doi: 10.1007/s12671-018-0925-0] [Medline: 30524514]
55. Ketelaar SM, Nieuwenhuijsen K, Gärtner FR, Bolier L, Smeets O, Sluiter JK. Mental vitality @ work: the effectiveness of a mental module for workers' health surveillance for nurses and allied health professionals, comparing two approaches in a cluster-randomised controlled trial. Int Arch Occup Environ Health 2014 Jul;87(5):527-538. [doi: 10.1007/s00420-013-0893-6] [Medline: 23893252]
56. Mak W, Wong C, Chan A, Lau J. Mobile self-compassion programme for promotion of public mental health: a randomised controlled trial. Hong Kong Med J 2019 Feb;25(Suppl 2):15-17 [FREE Full text] [Medline: 30674702]
57. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
58. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
59. Sin J, Henderson C, Spain D, Cornelius V, Chen T, Gillard S. eHealth interventions for family carers of people with long term illness: a promising approach? Clin Psychol Rev 2018 Mar;60:109-125 [FREE Full text] [doi: 10.1016/j.cpr.2018.01.008] [Medline: 29429856]
60. Adenuga OA, Kekwaletswe RM, Coleman A. eHealth integration and interoperability issues: towards a solution through enterprise architecture. Health Inf Sci Syst 2015;3:1 [FREE Full text] [doi: 10.1186/s13755-015-0009-7] [Medline: 26167279]
61. Ridout B, Campbell A. Using Facebook to deliver a social norm intervention to reduce problem drinking at university. Drug Alcohol Rev 2014 Nov;33(6):667-673. [doi: 10.1111/dar.12141] [Medline: 24689339]
62. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]

Abbreviations

AI: artificial intelligence
APMS: Adult Psychiatric Morbidity Survey
ASSIA: Applied Social Sciences Index and Abstracts
CBT: cognitive behavioral therapy
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CMD: common mental disorder
CONSORT: Consolidated Standards of Reporting Trials
EMBASE: Excerpta Medica dataBASE
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
ICROMS: integrated criteria for review of multiple study designs

ICT: information and communication technology
MEDLINE: Medical Literature Analysis and Retrieval System Online
NHS: National Health Service
PsycINFO: Psychological Information
RCT: randomized controlled trial
SMD: standardized mean difference
WoS: Web of Science

Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by A Li, S Zhao; comments to author 24.06.20; revised version received 30.06.20; accepted 14.07.20; published 02.09.20.

Please cite as:
Sin J, Galeazzi G, McGregor E, Collom J, Taylor A, Barrett B, Lawrence V, Henderson C
Digital Interventions for Screening and Treating Common Mental Disorders or Symptoms of Common Mental Illness in Adults: Systematic Review and Meta-analysis
J Med Internet Res 2020;22(9):e20581
URL: https://www.jmir.org/2020/9/e20581
doi:10.2196/20581
PMID:32876577

Review

Internet-Based Supportive Interventions for Family Caregivers of People With Dementia: Systematic Review and Meta-Analysis

Minmin Leng[1,2*], MD; Yajie Zhao[1,2*], MA; Hongmei Xiao[1,2], MA; Ce Li[1,2], MA; Zhiwen Wang[1,2], PhD

[1]School of Nursing, Peking University, Beijing, China

[2]Peking University Health Science Centre for Evidence-Based Nursing: A Joanna Briggs Institute Affiliated Group, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhiwen Wang, PhD
School of Nursing
Peking University
38 Xueyuan Road, Haidian District
Beijing, 100191
China
Phone: 86 15901566817
Email: wzwjing@sina.com

Abstract

Background: Caring for people with dementia is perceived as one of the most stressful and difficult forms of caring. Family caregivers always experience high levels of psychological burden and physical strain, so effective and practical support is essential. Internet-based supportive interventions can provide convenient and efficient support and education to potentially reduce the physical and psychological burden associated with providing care.

Objective: This review aimed to (1) assess the efficacy of internet-based supportive interventions in ameliorating health outcomes for family caregivers of people with dementia, and (2) evaluate the potential effects of internet-based supportive intervention access by caregivers on their care recipients.

Methods: An electronic literature search of the PubMed, EMBASE, Web of Science, CINAHL, Cochrane Library, and PsycINFO databases was conducted up to January 2020. Two reviewers (ML and YZ) worked independently to identify randomized controlled trials (RCTs) that met the inclusion criteria and independently extracted data. The quality of the included RCTs was evaluated using the approach recommended by the Cochrane Handbook for Systematic Reviews of Interventions. Standardized mean differences (SMDs) with 95% CIs were applied to calculate the pooled effect sizes.

Results: In total, 17 RCTs met the eligibility criteria and were included in this systematic review. The meta-analysis showed that internet-based supportive interventions significantly ameliorated depressive symptoms (SMD=–0.21; 95% CI –0.31 to –0.10; *P*<.001), perceived stress (SMD=–0.40; 95% CI –0.55 to –0.24; *P*<.001), anxiety (SMD=–0.33; 95% CI –0.51 to –0.16; *P*<.001), and self-efficacy (SMD=0.19; 95% CI 0.05-0.33; *P*=.007) in dementia caregivers. No significant improvements were found in caregiver burden, coping competence, caregiver reactions to behavioral symptoms, or quality of life. Six studies assessed the unintended effects of internet-based supportive intervention access by caregivers on their care recipients. The results showed that internet-based supportive interventions had potential benefits on the quality of life and neuropsychiatric symptoms in care recipients.

Conclusions: Internet-based supportive interventions are generally effective at ameliorating depressive symptoms, perceived stress, anxiety, and self-efficacy in dementia caregivers and have potential benefits on care recipients. Future studies are encouraged to adopt personalized internet-based supportive interventions to improve the health of family caregivers and their care recipients.

Trial Registration: PROSPERO CRD42020162434; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=162434

(J Med Internet Res 2020;22(9):e19468) doi:10.2196/19468

KEYWORDS

family caregivers; dementia; supportive interventions; internet; systematic review

Introduction

According to World Alzheimer Report 2018 [1], dementia affects almost 50 million people globally, with a new case of dementia occurring around the world every 3 seconds; this number is expected to increase to an alarming 152 million by 2050. Most people with dementia live at home and are cared for primarily by their spouse or adult children, the informal caregivers [2]. Caring for people with dementia is perceived as one of the most stressful experiences, as family caregivers may face long-term problems of managing activities of daily living [3,4], behavioral and psychological symptoms [5,6], and organizing care and providing emotional support [7,8]. However, due to a lack of supportive resources and knowledge of dementia, family caregivers have low confidence in managing caregiving [9] and do not know what to do when their relatives have dementia-related behavioral problems, need emotional support, and require the coordination of dementia care [10]. In addition, improper care behavior can also induce care recipients to develop various behavioral and psychological problems. As a result, dementia family caregivers always experience higher levels of psychological burden, physical strain, and ineffective coping than caregivers of older adults with physical impairments [11,12].

With the given physical and psychological impacts on family caregiver well-being, effective and practical support for family caregivers is essential. The Alzheimer's Disease International points out that help and support for caregivers should be a fundamental lynchpin of any national dementia plan [13]. Most currently available interventions to support family caregivers of people with dementia are "face-to-face" interventions [14-16], but the uptake of such interventions is poor. It is estimated that only a small percentage of family caregivers access caregiver support services, with the difficulty of leaving the care recipient and stigma being the main obstacles to uptake [17,18]. In addition, the continued increase in the number of people with dementia has led to concerns about whether the current labor force can cope with such an increased future caring demand [19].

Internet-based supportive interventions could be an efficient alternative to close the support gap to provide education and support for family caregivers, especially for those finding it difficult to leave their care recipient or requiring flexibility due to caring responsibilities. The benefits of interventions provided via the internet are that they are relatively low cost and more convenient because they enable family caregivers to learn at anytime and anywhere [20,21]. The growing number of randomized controlled trials (RCTs) in the field of internet-based support reflects the increasing demand for strategies that can complement existing services and better support family caregivers providing care to people with dementia.

Recent systematic reviews have concluded that internet-based supportive interventions can improve health outcomes in family caregivers of people with chronic disease [22-24]. However, the findings regarding the effect of internet-based supportive interventions on the improvement of health outcomes in family caregivers of people with dementia have been inconsistent. For example, some RCTs have indicated that the level of caregiver burden [25,26], depressive symptoms [25,27], or coping competence [25,28] was significantly ameliorated in the internet-based group compared with the control group. Conversely, some trials have shown that there were no significant differences detected in caregiver burden [29,30], depressive symptoms [29,31], or coping competence [26,30] between the internet-based and control groups.

Currently, there are several systematic reviews of internet-based supportive interventions for family caregivers of people with dementia. In a previous systematic review [32], the author included 14 empirical studies; the results indicated that computer-mediated interventions were potentially useful as a supportive intervention. Nonetheless, because the study design of the included studies was diverse and the data were insufficient, the authors only described the inconsistent results and did not conduct a meta-analysis. A recent meta-analysis [33] investigated the effects of internet-based interventions on mental health outcomes for home caregivers of people with dementia. However, the number of articles included in this meta-analysis was limited. In another recent systematic review [11] of internet-based interventions for family caregivers of people with dementia, the authors mainly aimed to identify the key components of existing internet-based interventions designed to support family caregivers of people with dementia and explore which components are most valued by family caregivers.

In summary, the effects of internet-based supportive interventions on family caregivers of people with dementia require further exploration. Furthermore, to the best of our knowledge, there has been no systematic review that has evaluated the potential effects of internet-based supportive intervention access by family caregivers on their care recipients. Thus, we performed this systematic review and meta-analysis to further clarify the effects of internet-based supportive interventions on dyads (caregivers + people with dementia). The primary objectives of this study were to assess the efficacy of internet-based supportive interventions in ameliorating health outcomes for family caregivers of people with dementia and examine whether specific types of internet-based supportive interventions had a beneficial impact on family caregivers' health outcomes. The secondary objective was to evaluate the potential effects of internet-based supportive intervention access by caregivers on their care recipients.

Methods

We conducted this systematic review and meta-analysis of RCTs by following the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines [34]. The research protocol was registered in PROSPERO (registration number CRD42020162434).

Literature Search Strategy

An electronic literature search of the PubMed, EMBASE, Web of Science, CINAHL, Cochrane Library, and PsycINFO databases was conducted up to January 2020. Additional relevant studies were identified through the reference lists of the included

studies and previous related systematic reviews. For each database, the search strategy was customised. The key search terms were a combination of medical subject heading terms (MeSH) and entry terms. The detailed information about the search strategies and search results of each database is available in Multimedia Appendix 1.

Inclusion Criteria

Studies were included in this review if they met the following criteria: (1) participants were family caregivers who were currently providing caregiving support to people with dementia, defined as a family member such as their spouse or adult children providing unpaid care; (2) the intervention was a digital one delivered via any internet-based modality, which could include either single-component interventions or multiple-component interventions to family caregivers; (3) comparison was usual care or minimal support control by using paper materials, telephone, or email, etc; (4) primary outcomes included outcome variables related to family caregivers of people with dementia (depressive symptoms, caregiver burden, coping competence, perceived stress, caregiver reaction to behavioral symptoms, anxiety, quality of life, and self-efficacy), and secondary outcomes included outcome variables related to people with dementia (care recipient' quality of life and neuropsychiatric symptoms); and (5) to achieve high levels of evidence, we included only RCTs.

Study Selection and Data Extraction

All of the searched records were imported into EndNote X9 to eliminate duplicate studies. Two reviewers (ML and YZ) worked independently to identify studies that met the inclusion criteria. To further evaluate the eligibility of potential studies, we obtained full-text articles and discussed any disagreements with the third reviewer (ZW). Data were extracted from the included studies by 2 independent reviewers (ML and YZ) using the standardized data extraction tool. From each included study, we extracted data including the author, publication year, country, sample size, participant' mean age, internet-based supportive intervention details (eg, methods, content, and duration), data collection time points, and outcome measurement tools. Any disagreements between the 2 independent reviewers were resolved by the third reviewer (ZW).

Quality Appraisal

The risk of bias in the included studies was assessed using the approach recommended by the Cochrane Handbook for Systematic Reviews of Interventions [35]. The 7 recommended items included random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data, selective reporting, and other bias. All included studies were independently evaluated, and the risk of bias for each item was categorized as "low risk," "unclear," or "high risk." Disagreements between the 2 reviewers were resolved by the third reviewer (ZW).

Data Analysis

Standardized mean differences (SMDs) with 95% CIs were used when studies used different outcome scales, and mean differences (MDs) with 95% CIs were applied when studies used the same outcome scales. The level of heterogeneity was evaluated by the I^2 method, and a value of I^2>50% was regarded as significant heterogeneity [36]. The fixed-effects model was used to calculate the pooled effect size if the data were not significantly heterogeneous. Otherwise, the random-effects model was used. A sensitivity analysis was performed by excluding one study at a time to confirm the consistency of the findings. Publication bias was evaluated by visual inspection of funnel plot. RevMan 5.3 provided by the Cochrane Collaboration was used for all statistical calculations, and a P value <.05 was considered statistically significant.

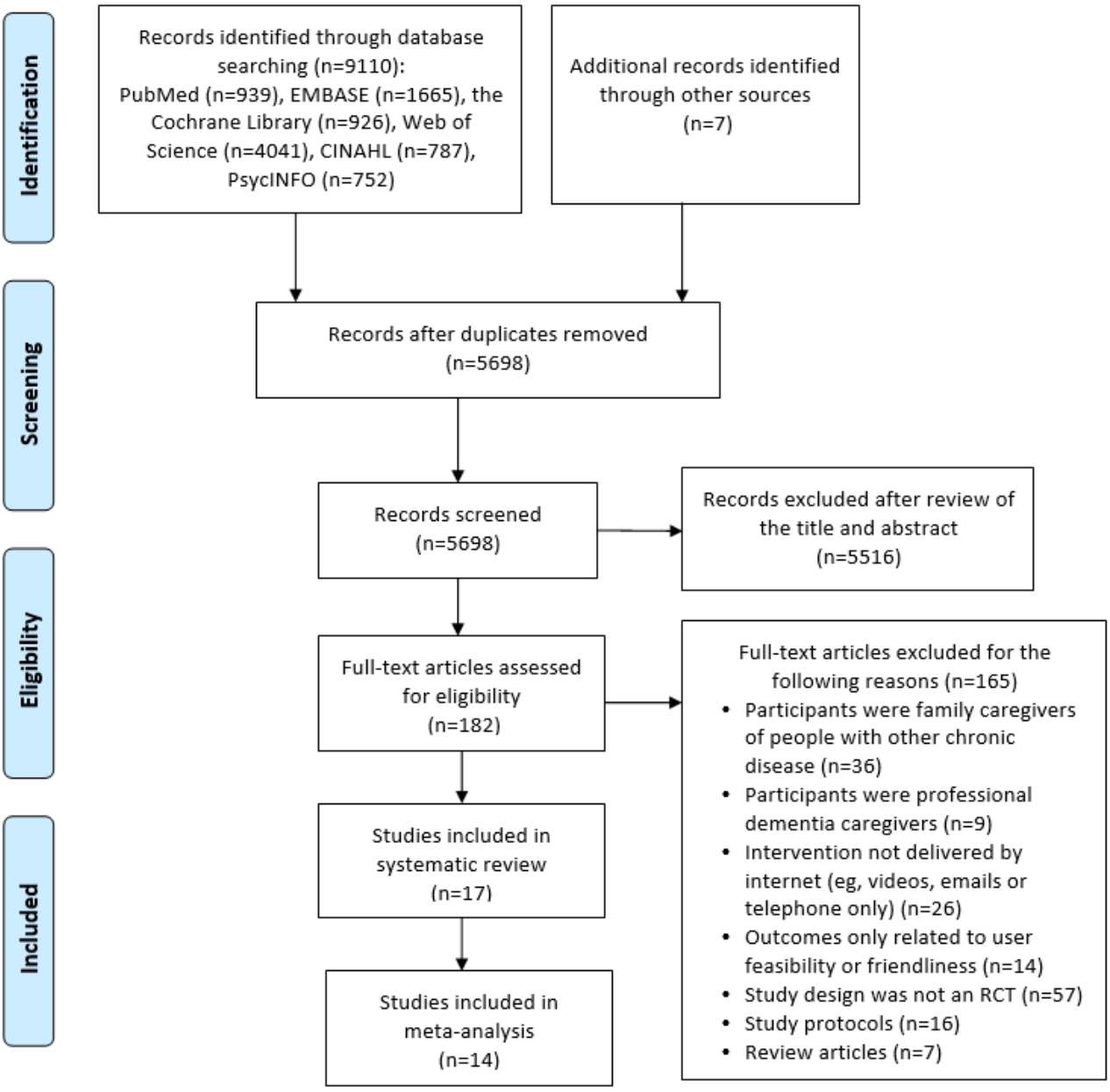
In the systematic review, intervention formats were divided into personalized and nonpersonalized formats. Subgroup analyses were performed to explore which internet-based supportive intervention format was most beneficial for family caregivers of people with dementia.

Results

Study Selection

A total of 9110 records were identified from the electronic databases in the final search, with an additional 7 records identified through other sources. After removal of duplicates and obviously irrelevant records, we retrieved 182 full-text articles to further evaluate their eligibility. In total, 165 articles were excluded because they did not meet the inclusion criteria. Ultimately, a total of 17 studies [25-31,37-46] involving 2202 family caregivers of people with dementia were included in this systematic review. The detailed screening process is illustrated in Figure 1.

Figure 1. Flow diagram for search and selection of the included studies.



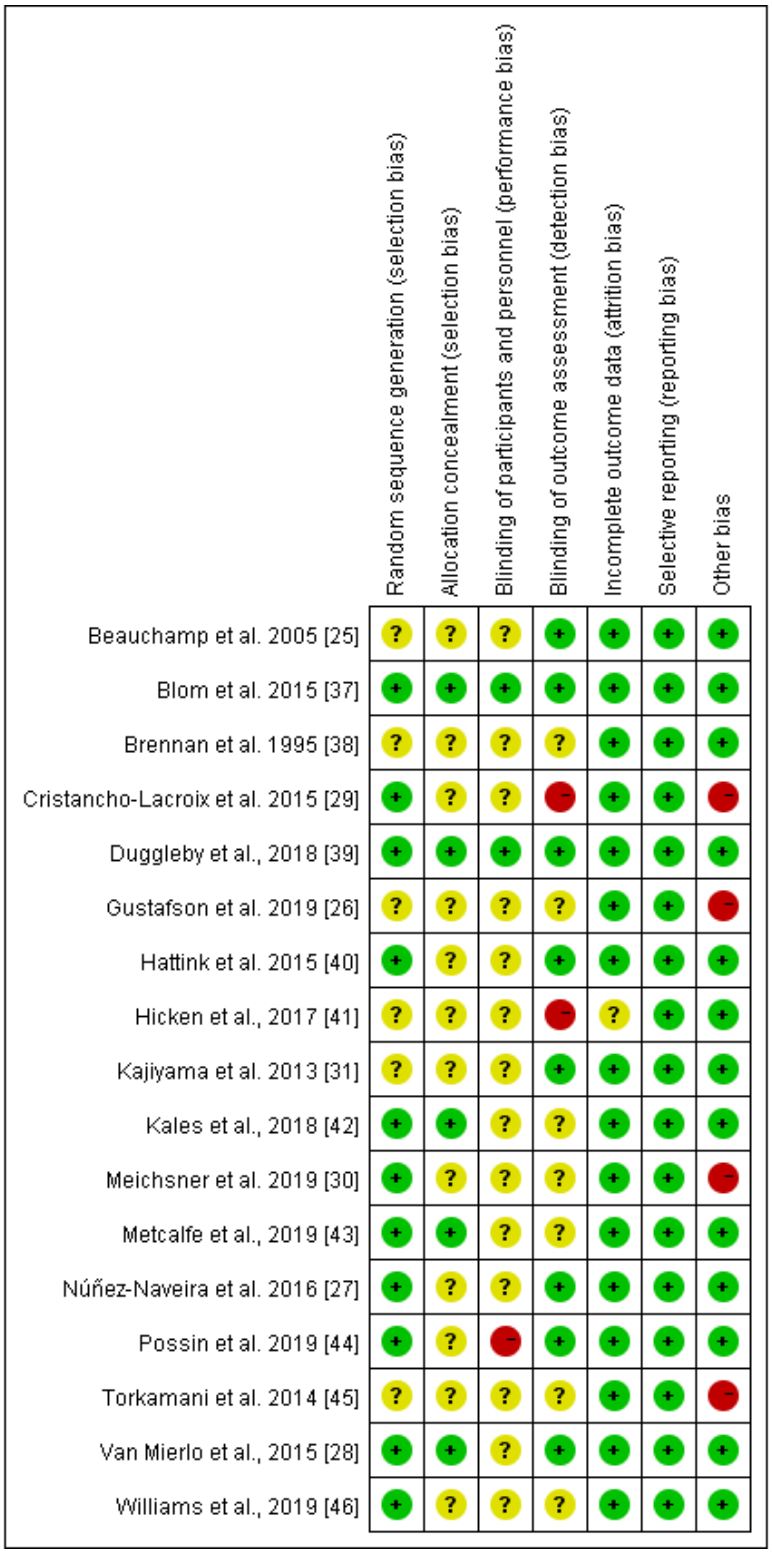
Study Characteristics

The number of participants in each study ranged from 25 to 547. The intervention duration ranged from 4 weeks to 12 months. Eight studies were performed in the United States [25,26,31,38,41,42,44,46], 3 studies were performed in Netherlands [28,37,40], 2 studies were performed in France [29,43], and 1 study each was conducted in Canada [39], Germany [30], Spain [27], and the UK [45]. All of the included studies reported clear inclusion and exclusion criteria for their participants. In 9 studies, the forms of internet-based support interventions were personalized [25,27,28,40-42,44-46] whereas in the other 8 studies [26,29-31,37-39,43], they were nonpersonalized. The nature of the treatment of control group participants differed from one study to another. The control groups in 10 studies [25,27-30,40,42-45] were exposed to the same conditions as the corresponding intervention groups, except for the implementation of internet-based support intervention in the latter. However, in some studies, the control group received e-bulletins [37], booklet [39], or book [26] while the intervention group received internet-based support intervention. Similarly, Brennan et al [38] provided placebo training experience identifying local services and resources for the comparison group. In the studies of Hicken et al [41] and Williams et al [46], caregivers in the control group received telephone-support attention. Kajiyama et al [31] provided a website containing the similar navigational features to caregivers in the control condition, but they did not provide any of the information or skills training content presented in the intervention group. All 17 included studies evaluated the effect of internet-based supportive interventions in ameliorating health outcomes for family caregivers of people with dementia, and 6 studies [28,29,42-45] also reported potential effects of internet-based supportive intervention access by caregivers on their care recipients. The main characteristics of the included studies are presented in Multimedia Appendix 2.

Risk of Bias Assessment

In general, the RCTs included in this systematic review showed an acceptable risk of bias. Most of the included studies reported randomization, but some trials did not describe the allocation concealment details, which could cause potential selection bias. However, only 2 studies [37,39] were judged as having a low risk of performance bias because it was difficult to blind the participants in psychosocial intervention trials. Approximately half of the studies blinded outcome assessors; therefore, their risk of detection bias was categorized as low. The risk of attrition bias was judged as unclear in only 1 study [41], and there was no evidence of selective reporting bias in any of the included studies. The risk of other bias was categorized as high in 4 studies because of the baseline differences [29,45] and small sample sizes [26,30]. The risk of bias assessment of each included study is shown in Figure 2.

Figure 2. Risk of bias summary: review of the authors' judgments about each risk of bias item for each included study.

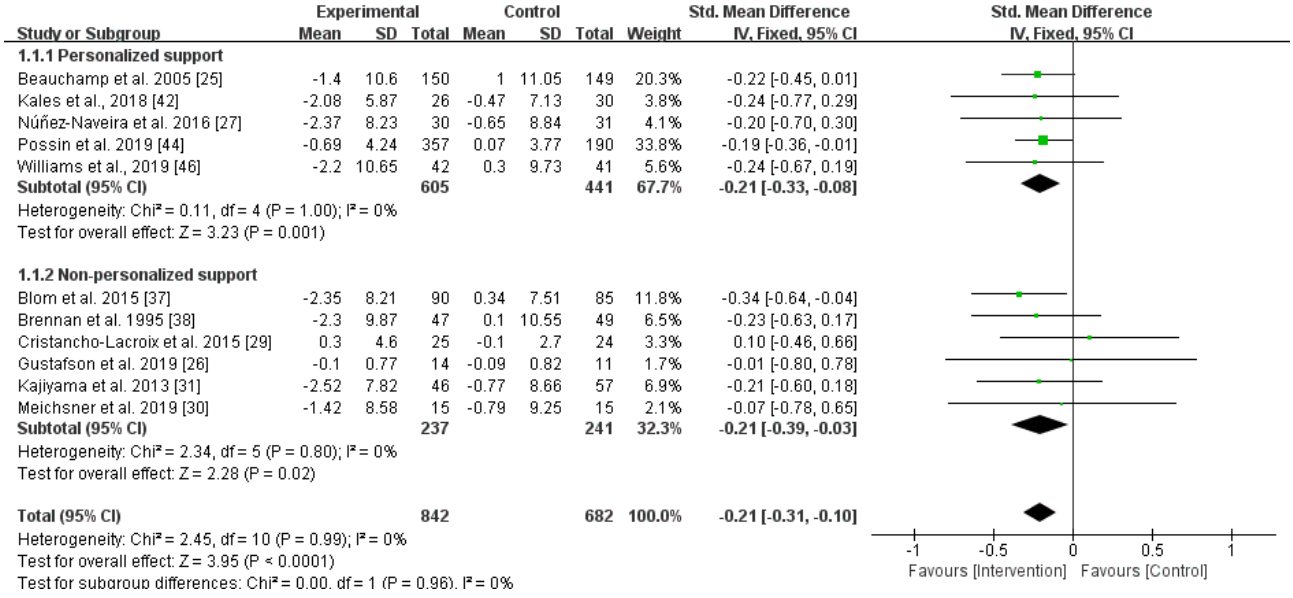


Meta-Analysis Results of Interventions Among Family Caregivers

Depressive Symptoms

A total of 11 studies assessed the effects of internet-based supportive interventions on depressive symptoms using the Centre for Epidemiologic Studies Depression Scale [25,27,30,31,37,38,42,46], Patient Health Questionnaire [26,44], and Beck Depression Inventory-II Scale [29]. Because of the different assessment tools, we used the SMD to represent the pooled effect size. The meta-analysis showed that caregivers in the internet-based supportive intervention group exhibited a significant amelioration of depressive symptoms compared with controls (n=1524; SMD=–0.21; 95% CI –0.31 to –0.10; P<.001; I^2=0%, fixed-effects model; Figure 3).

Figure 3. The effect of internet-based supportive interventions on depressive symptoms.



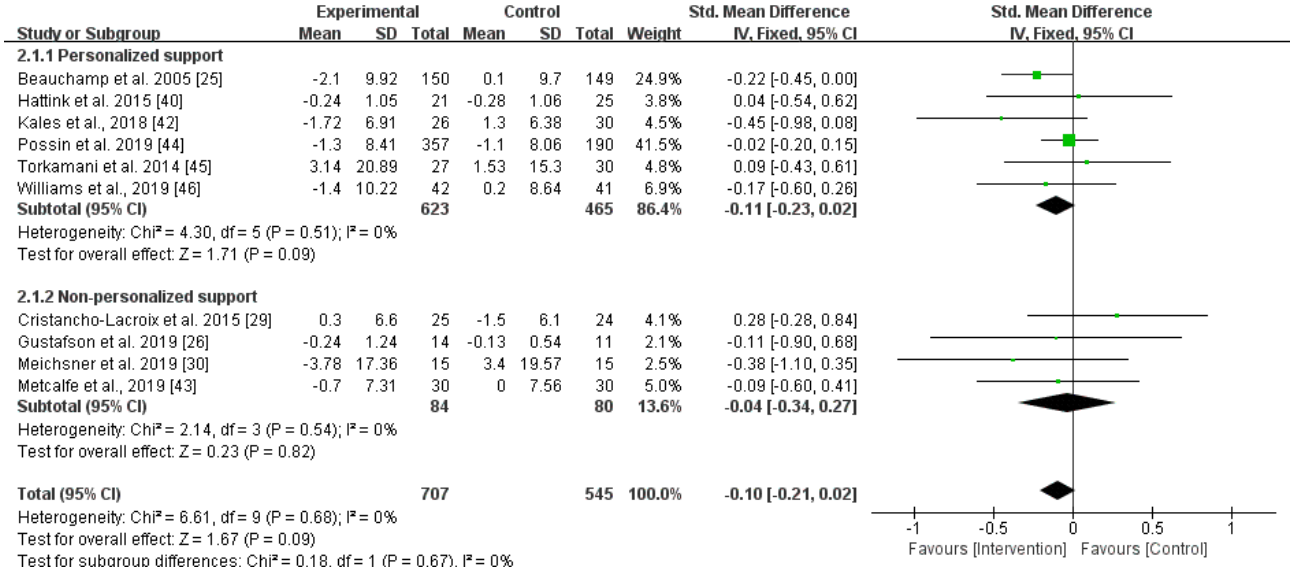
For the subgroup analysis, 5 studies reported detailed data on personalized format interventions [25,27,42,44,46], while 6 studies reported detailed data on nonpersonalized format interventions [26,29-31,37,38]. In the personalized format subgroup, the results showed that the caregivers in the intervention group had a significant reduction in SMD scores for depressive symptoms compared with the caregivers in the control group (n=1046; SMD=–0.21; 95% CI –0.33 to –0.08; P=.001; I^2=0%, fixed-effects model; Figure 3). In the nonpersonalized format subgroup, a significant reduction in SMD scores for depressive symptoms was also observed (n=478; SMD=–0.21; 95% CI –0.39 to –0.03; P=.02; I^2=0%, fixed-effects model; Figure 3).

Caregiver Burden

A total of 10 studies evaluated the effects of internet-based supportive interventions on caregiver burden measured by the Zarit Burden Interview [29,42,44-46], the Caregiver Strain Instrument [25], the Caregiver Load scale [26], the Burden Scale for Family Caregivers [43], the Burden Visual Analog scale [30], and 1 question [40]. Because of the different measuring tools, we applied the SMD to represent the pooled effect size. The meta-analysis showed that the overall combined effect of the internet-based supportive intervention on caregiver burden was not statistically significant (n=1252; SMD=–0.10; 95% CI –0.21 to 0.02; P=.09; I^2=0%; fixed-effects model; Figure 4).

In the personalized format subgroup, the results from 6 studies [25,40,42,44-46] showed that internet-based supportive interventions improved the burden status of caregivers with an SMD score of –0.11, but the difference was not statistically significant (n=1088; SMD=–0.11; 95% CI –0.23 to 0.02; P=.09; I^2=0%; fixed-effects model; Figure 4). In the nonpersonalized format subgroup, the results from 4 studies [26,29,30,43] showed that internet-based supportive interventions improved the burden status of caregivers with an SMD score of –0.04, and this difference was also not statistically significant (n=164; SMD=–0.04; 95% CI –0.34 to 0.27; P=.82; I^2=0%; fixed-effects model; Figure 4).

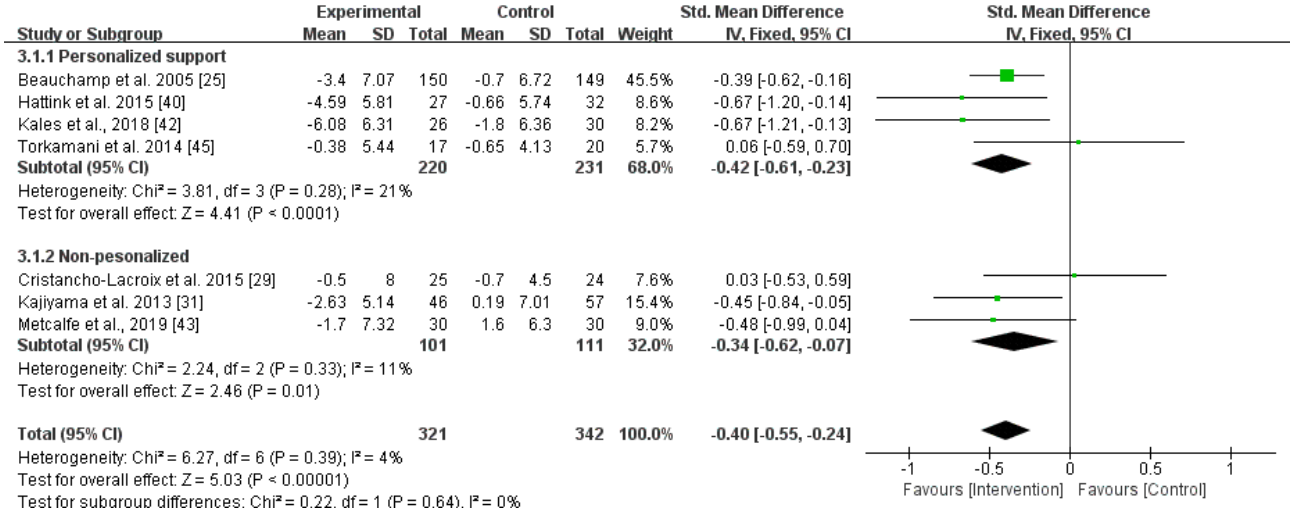
Figure 4. The effect of internet-based supportive interventions on caregiver burden.



Perceived Distress/Stress

A total of 7 studies reported the effects of internet-based supportive interventions on distress/stress measured by the Perceived Stress Scale [29,31,43], Neuropsychiatric Inventory subscale [42,45], Interpersonal Reactivity Index subscale [40], and 2 questions [25]. A significant improvement in distress/stress after internet-based supportive intervention was observed compared with the control condition (n=663; SMD=–0.40; 95% CI –0.55 to –0.24; P<.001; I^2=4%, fixed-effects model; Figure 5).

Figure 5. The effect of internet-based supportive interventions on perceived distress/stress.

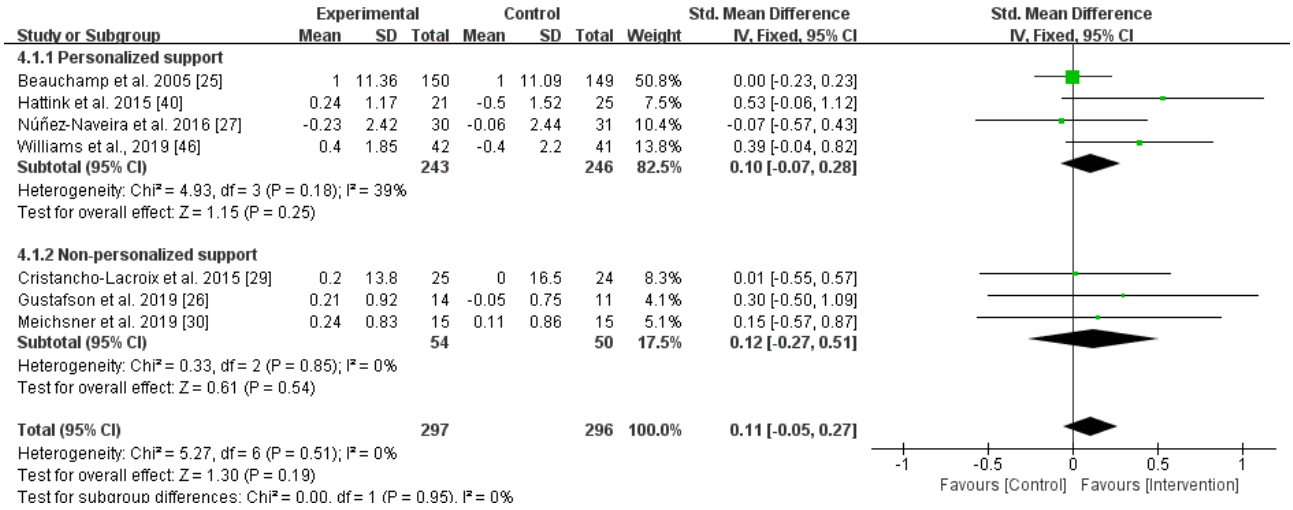


As much as 4 studies reported detailed data on personalized format interventions [25,40,42,45], while 3 studies reported detailed data on nonpersonalized format interventions [29,31,43]. In the personalized format subgroup, the results showed that the internet-based interventions had a significant beneficial effect on distress/stress (n=451; SMD=–0.42; 95% CI –0.61 to –0.23; P<.001; I^2=21%, fixed-effects model; Figure 5). In the nonpersonalized format subgroup, a significant beneficial effect on distress/stress was also observed (n=212; SMD=–0.34; 95% CI –0.62 to –0.07; P=.01; I^2=11%, fixed-effects model; Figure 5).

Coping Competence

A total of 7 studies used coping competence as an outcome variable, using the Short Sense of Competence Questionnaire [40,46], Caregiver Competence Scale [27], Revised Ways of Coping [25], Visual Analog Scale of coping [29], Caregiver Appraisal Scale [26], and Caregiver Grief Scale [30] to assess coping competence. The meta-analysis results showed that internet-based supportive interventions had no significant effect on coping competence (n=593; SMD=0.11; 95% CI –0.05 to 0.27; P=.19; I^2=0%; fixed-effects model; Figure 6).

Figure 6. The effect of internet-based supportive interventions on coping competence.

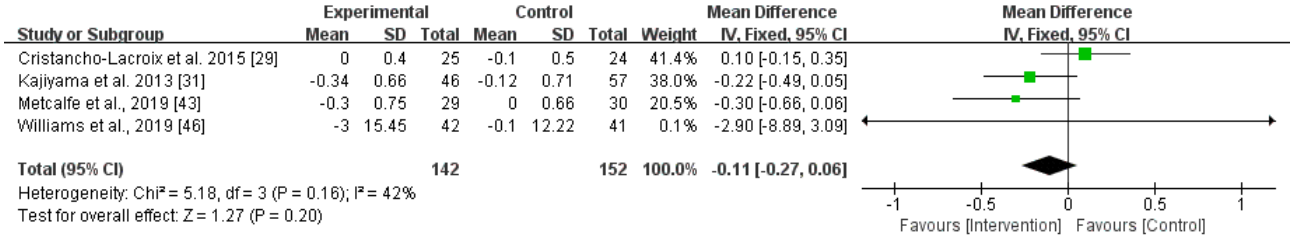


In the personalized format subgroup, the results from 4 studies [25,27,40,46] showed that the effect of the internet-based supportive interventions on coping competence was not statistically significant (n=489; SMD=0.10; 95% CI –0.07 to 0.28; *P*=.25; I^2=39%; fixed-effects model; Figure 6). In the nonpersonalized format subgroup, the results from 3 studies [26,29,30] showed a similar effect (n=108; SMD=0.12; 95% CI –0.27 to 0.51; *P*=.54; I^2=0%; fixed-effects model; Figure 6).

Caregiver Reactions to Behavioral Symptoms

The effects of internet-based supportive interventions on caregiver reactions to behavioral symptoms of people with dementia were evaluated in 4 studies using the Revised Memory and Behavior Problems Checklist [29,31,43,46]. Because the measuring tool was the same, we applied the MD to represent the pooled effect size. The meta-analysis results showed that internet-based supportive interventions had no significant effect on caregiver reaction to behavioral symptoms (n=294; MD=–0.11; 95% CI –0.27 to 0.06; *P*=.20; I^2=42%; fixed-effects model; Figure 7).

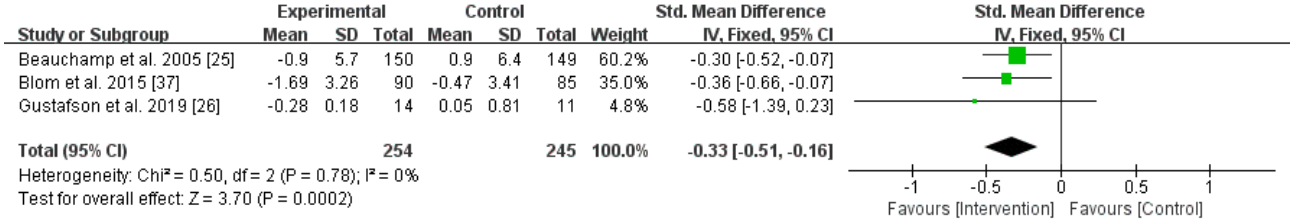
Figure 7. The effect of internet-based supportive interventions on caregiver reaction to behavioral symptoms.



Anxiety

Three studies used anxiety as an outcome variable, using the State-Trait Anxiety Inventory [25], Hospital Anxiety and Depression subscale [37], and Generalized Anxiety Disorder Scale [26] to assess changes in caregiver anxiety status. The results showed that compared with the control caregivers, the caregivers in the intervention group experienced significant amelioration of anxiety (n=499; SMD=–0.33; 95% CI –0.51 to –0.16; *P*<.001; I^2=0%; fixed-effects model; Figure 8).

Figure 8. The effect of internet-based supportive interventions on anxiety.



Quality of Life

Three studies reported the effects of internet-based supportive interventions on quality of life of dementia caregivers measured by the Perceived Quality of Life [31], Quality of Life Scale [45], and 2 distinct questions [40]. The meta-analysis showed that the effect of the internet-based supportive intervention on quality of life was not statistically significant (n=187; SMD=0.15; 95% CI –0.14 to 0.44; *P*=.31; I^2=26%; fixed-effects model; Multimedia Appendix 3).

Self-Efficacy

In 2 studies, self-efficacy was used as the outcome variable, and the Caregiver Self-Efficacy Scale [44] and 6 self-efficacy questions regarding areas of caregiving [25] were used to measure self-efficacy. The results from the 2 studies showed that the internet-based supportive interventions had a significant beneficial effect on self-efficacy (n=846; SMD=0.19; 95% CI 0.05 to 0.33; P=.007; I^2=0%; fixed-effect models; Multimedia Appendix 4).

Description of Studies Not Suitable for Meta-Analysis

Three RCTs [28,39,41] were not included in the meta-analysis due to the limitation of data types. Instead, the results of these studies are described and summarized in the following narrative review. Van Mierlo et al [28] found that an internet-based supportive intervention significantly improved the sense of competence of caregivers after 12 months (P=.03). However, there were no significant differences in emotional stress between the intervention and control groups. Hicken et al [41] evaluated the effects of an internet-based supportive intervention for 4-6 months in family caregivers of veterans with dementia. There were no significant differences in the changes in caregiver burden or depressive symptoms between the groups from baseline to the end of the intervention. The important findings in this study were that the effects partly changed when the analyses were stratified by rurality. For urban caregivers, the burden score remained stable in the control group but decreased in the internet group (P=.014). For rural caregivers, there were no significant differences between the groups. Duggleby et al [39] assessed the effects of 3 months of an internet-based supportive intervention on self-efficacy among caregivers of older adults with Alzheimer's disease and multiple chronic conditions. Although no significant group differences were observed in outcome measures, the caregivers in the intervention group indicated, when asked in interviews, that the internet-based supportive intervention helped them with their transitions.

Potential Effects on Care Recipients

Six studies [28,29,42-45] evaluated the potential effects of internet-based supportive intervention access by caregivers on their care recipients. Van Mierlo et al [28] examined the effects of DEMentia Digital Interactive Social Chart (DEM-DISC) access by caregivers on their care recipients. The results showed no significant differences between the intervention and control groups on the quality of life and neuropsychiatric symptoms of people with dementia as reported by family caregivers at 12 months. Two other studies [29,42] indicated similar results; there were no significant differences in behavioral frequency, severity, or overall neuropsychiatric symptoms between the intervention and control groups. Conversely, a recent RCT [44] showed that the quality of life of care recipients significantly declined more in the usual care group than in the intervention group from baseline to 12 months (P=.04). The rate of emergency department visits over a 12-month period significantly increased more in the usual care group than in the intervention group (P=.04). However, there were no significant differences in rates of ambulance use (P=.12) and hospital use (P=.71) between the intervention and usual care groups. Another study [43] demonstrated that behavior problem frequency of care recipients significantly reduced more in the intervention group than in the control group from baseline to 6 weeks (P=.04). In a previous study conducted by Torkamani et al [45], the comparison revealed a significant difference in neuropsychiatric symptoms, showing worse neuropsychiatric symptoms for care recipients in the intervention group than those in the control group at baseline. The group difference remained significant at 3 and 6 months. Therefore, it is difficult to distinguish whether the intervention is effective for the neuropsychiatric symptoms of care recipients. One positive finding was that the care recipients in the intervention group showed weight gain.

Publication Bias and Sensitivity Analyses

Visual inspection of the funnel plot did not reveal evidence of potential publication bias. The funnel plot is shown in Multimedia Appendix 5. We conducted the sensitivity analysis by removing all the studies included in this meta-analysis one by one, and confirmed that the findings were not significantly influenced by any single study.

Discussion

Summary and Interpretation of Results

A total of 17 studies of internet-based supportive interventions in family caregivers of people with dementia were included in our systematic review, and 14 studies were included in the meta-analyses. Unlike previous reviews in this field, this review not only explored the effects of internet-based interventions on family dementia caregivers but also focused on the effects on care recipients. The meta-analysis showed that internet-based supportive interventions significantly ameliorated depressive symptoms, perceived stress, anxiety, and self-efficacy in dementia caregivers. Both the personalized and nonpersonalized formats of internet-based supportive interventions significantly reduced depressive symptoms and perceived stress. However, current evidence failed to support the efficacy of internet-based supportive interventions on caregiver burden, coping competence, caregiver reactions to behavioral symptoms, or quality of life. The results based on 6 studies [28,29,42-45] showed that internet-based supportive interventions had potential benefits on the quality of life and neuropsychiatric symptoms in care recipients.

This systematic review showed that internet-based supportive interventions had significant beneficial effects on mental health, such as depressive symptoms, perceived distress/stress, and anxiety. Several systematic reviews have also recognized the potential for internet-based interventions in supporting family caregivers of people with dementia to maintain well-being and independence. Our results are comparable to a recently published meta-analysis by Zhao et al [33], who reported that web-based interventions had a positive effect on mental health in home caregivers of people with dementia. Another systematic review [11] identified a broad variety of internet-based interventions that focused on providing information; engaging with social care professionals; and providing peer support, psychological support, and decision support. Although the main focus of that review [11] was not on effectiveness, some internet-based

multiple-component interventions showed promise in reducing depressive symptoms and anxiety in family caregivers. For these positive results, the possible reason is that internet-based supportive interventions can improve the dementia-related knowledge and care skills of family caregivers, thereby enhancing their confidence in managing caregiving and alleviating their negative emotions.

In this systematic review, subgroup analysis showed that both personalized and nonpersonalized formats of internet-based supportive interventions had significant beneficial effects on depressive symptoms and perceived stress. Compared with the effect size of the nonpersonalized format, the combined effect size of the personalized format was larger, which means that personalized format intervention was more beneficial for dementia family caregivers. Because the symptoms of people with dementia are diverse and the coping abilities of home caregivers are uneven, it is important to provide personalized interventions. The American Geriatrics Society points out that providing care that is respectful of and responsive to individual person needs, preferences, and values is one of the pillars of quality health care [47]. Recent guidelines [48-50] that offer best-practice advice on support and care for people with dementia and their families and caregivers stated that dementia care should be personalized to a person's interests, abilities, values, beliefs, personalities, life experiences, likes, and dislikes and should be based on the severity and characteristics of the symptoms. The personalized format of internet-based supportive interventions enables health-care professionals to understand and provide support for the unmet needs of individuals with dementia and their family caregivers. In Núñez-Naveira et al's study [27], at the start of the program, the participants needed to complete an interactive customization questionnaire with questions about the time availability for learning, energy, and preferences of the care provider and about the severity of dementia of the people cared for by the family caregiver. By completing the questionnaire, the information content provided to the family caregivers was personalized and adjusted to their personal situation. In Possin et al's study [44], care team members responded to family caregivers' immediate needs first, then screened for common problems and provided standardized education and tailored support according to the care plan protocols. In Williams et al's study [46], dementia care experts implemented personalized interventions by providing tailored feedback and guidance based on specific care encounters. The personalized interventions offered a user-friendly and promising method of individualizing professional consultation and guidance to maximize the intervention effects.

However, current evidence failed to support the efficacy of internet-based supportive interventions on caregiver burden, coping competence, caregiver reactions to behavioral symptoms, or quality of life. Studies have shown that caregiver outcomes, such as burden, have been shown to increase with time [51,52]. Although the internet-based supportive interventions did not significantly ameliorate those caregiver outcomes, they did lead to nonsignificant improvements in those caregiver outcomes. Internet-based supportive intervention is an emerging field, and the research in this field lacks consistency, such as inconsistent choice of theoretical models, inconsistent content in intervention courses, inconsistent doses of interventions, inconsistent conditions of control groups, and inconsistent measurement tools. These inconsistencies need to be taken seriously, so it is necessary to further explore the effects of these types of interventions.

In recent years, nonpharmacological interventions have been increasingly implemented to prevent and reduce challenging behavioral and psychological symptoms of dementia [53,54]. However, few of these interventions targeted people with dementia and their caregivers as dyads but rather people with dementia or caregivers alone [55]. In our systematic review, the evidence for potential effects on care recipients was limited, as only 5 of 17 RCTs provided effect data about care recipients. One study [44] showed that the quality of life and the rate of emergency department visits of care recipients increased significantly more in the intervention group than in the usual care group, whereas another study [28] indicated no significant differences between the intervention and control groups on the quality of life. Two studies [28,42] showed no significant differences between the intervention and control groups on the neuropsychiatric symptoms of people with dementia, whereas the other study [45] revealed unclear effects because of baseline differences. Another study [43] demonstrated that behavior problem frequency of care recipients was reduced significantly more in the intervention group than in the control group. Although the results were inconsistent across studies, they generally showed potential beneficial effects of internet-based supportive interventions on care recipients. Some outcome changes may not be statistically significant, but they are clinically significant for families who care for loved ones with dementia at home. In our opinion, internet-based supportive interventions can improve the caregiving skills of family caregivers, thereby meeting the complex care needs and providing better care for the people with dementia, and thus indirectly improve the health status of the people with dementia. Because of the limited evidence, further research is needed to explore the effects of internet-based supportive intervention access by caregivers on their care recipients.

Implications for Clinical Practice and Future Research

Our results demonstrate that internet-based supportive interventions have beneficial effects on family dementia caregivers and should be considered as a useful tool in clinical practice. The following issues need to be addressed when applying internet-based supportive interventions. First, the content of internet-based supportive interventions should be tailored according to user preferences, needs, personal situation, and dementia severity of the care recipients. It is better to provide online question-answering functionality on the internet platform so that dementia care experts can provide personalized feedback and guidance in a timely manner based on specific care encounters. Second, the ultimate goal of providing internet-based support services to family dementia caregivers is to improve the well-being of both dementia caregivers and people with dementia. Therefore, future studies should not only explore the effects of internet-based interventions on family dementia caregivers but also focus on their impact on care recipients. Third, privacy and security issues need to be highlighted. The details discussed by many users on the internet

are emotional and personal topics. The overwhelming and vocal concerns about the internet have been data security and privacy [56]. The reasons for choosing and trusting a particular website are more focused on the way to access the website and the content of the website, including password-protected entry methods, personalized content, and unbiased information, rather than complex and busy layouts, irrelevant content, and corporate looks. Fourth, future research into internet-based supportive interventions for caregivers of people with dementia may benefit from a mixed-method approach. Qualitative components gathered from interviews with caregivers who have used internet-based supportive interventions provide a significant supplement to quantitative outcome measures. Qualitative approaches offer deeper insight into caregivers' experiences of using the interventions and taps into factors, such as feeling focused, supported, and less isolated, that are of great importance to caregivers. Finally, there is a need for further research in this field to promote cost-effective care and lower the threshold of seeking support around the world.

Strengths and Limitations

The strength of this systematic review was that we focused on the dyad of dementia, including family dementia caregivers and people with dementia, not only caregivers or only care recipients. In addition, only RCTs with rigorous study designs were included, which implies that the quality of evidence was relatively high. This systematic review also had some limitations. First, the included studies lacked consistency, the content and doses of interventions varied widely, and the tools for measuring outcome variables were diverse. Thus, the optimal intervention design for family dementia caregivers remains unclear. Second, only 6 studies evaluated the potential effects of internet-based supportive intervention access by caregivers on their care recipients. Therefore, the meta-analysis of the effects on care recipients was not performed, but only described and summarized in the narrative review. Third, some outcome variables, such as anxiety and self-efficacy, were not analyzed by subgroup due to the limited number of studies. Therefore, further research is needed to provide sufficient evidence for practice.

Conclusions

Internet-based supportive interventions are generally effective at ameliorating depressive symptoms, perceived stress, anxiety, and self-efficacy in family dementia caregivers and have potential benefits on care recipients, although negative results were found in some RCTs. Future studies are encouraged to adopt personalized internet-based supportive interventions to improve the health of family caregivers and their care recipients. Combining personalized information with the help of dementia care experts and the possibility of communicating with other family dementia caregivers can augment standard dementia care and improve the efficiency of resource utilization.

Acknowledgments

This work was supported by the Beijing Municipal Science & Technology Commission (D171100008217007) and Chinese Nursing Association (ZHKY201915). The authors sincerely thank Drs Katherine Possin and Rachel Kiekhofer for providing the data from their study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search Strategy.
[DOCX File , 26 KB - jmir_v22i9e19468_app1.docx]

Multimedia Appendix 2
Characteristics of the studies included in this systematic review.
[DOCX File , 41 KB - jmir_v22i9e19468_app2.docx]

Multimedia Appendix 3
The effect of internet-based supportive interventions on quality of life.
[PNG File , 9 KB - jmir_v22i9e19468_app3.png]

Multimedia Appendix 4
The effect of internet-based supportive interventions on self-efficacy.
[PNG File , 9 KB - jmir_v22i9e19468_app4.png]

Multimedia Appendix 5
The funnel plot.
[DOCX File , 22 KB - jmir_v22i9e19468_app5.docx]

References

1. Alzheimer's Disease International (ADI). World Alzheimer Report 2018: The State of the Art of Dementia Research: New Frontiers. 2018. URL: https://www.alz.co.uk/research/world-report-2018 [accessed 2020-04-16]
2. Wimo A. Establishing and funding long term care infrastructure. 2018 Jul Presented at: 33rd International Conference of Alzheimer's Disease International; 26-29 July, 2018; Chicago, IL p. 26-29.
3. Drummond N, McCleary L, Garcia L, McGilton K, Molnar F, Dalziel W, et al. Assessing Determinants of Perceived Quality in Transitions for People with Dementia: a Prospective Observational Study. Can Geriatr J 2019 Mar 04;22(1):13-22 [FREE Full text] [doi: 10.5770/cgj.22.332] [Medline: 31501679]
4. Spigelmyer PC, Schreiber JB. A pilot study: Resistive behavior in the context of informal caregiver-assisted activities of daily living. Geriatr Nurs 2019 Jul;40(4):399-404. [doi: 10.1016/j.gerinurse.2019.01.005] [Medline: 30878279]
5. Huis In Het Veld JG, van Asch IFM, Willemse BM, Verkade P, Pot AM, Blom MM, et al. Process Evaluation of Nurse-Led Online Self-Management Support for Family Caregivers to Deal With Behavior Changes of a Relative With Dementia (Part 1): Mixed Methods Study. J Med Internet Res 2019 Oct 11;21(10):e13002 [FREE Full text] [doi: 10.2196/13002] [Medline: 31605517]
6. Baharudin AD, Din NC, Subramaniam P, Razali R. The associations between behavioral-psychological symptoms of dementia (BPSD) and coping strategy, burden of care and personality style among low-income caregivers of patients with dementia. BMC Public Health 2019 Jun 13;19(Suppl 4):447 [FREE Full text] [doi: 10.1186/s12889-019-6868-0] [Medline: 31196141]
7. Zwingmann I, Hoffmann W, Michalowsky B, Dreier-Wolfgramm A, Hertel J, Wucherer D, et al. Supporting family dementia caregivers: testing the efficacy of dementia care management on multifaceted caregivers' burden. Aging Ment Health 2018 Jul 20;22(7):889-896. [doi: 10.1080/13607863.2017.1399341] [Medline: 29156941]
8. Tatangelo G, McCabe M, Macleod A, You E. "I just don't focus on my needs." The unmet health needs of partner and offspring caregivers of people with dementia: A qualitative study. Int J Nurs Stud 2018 Jan;77:8-14. [doi: 10.1016/j.ijnurstu.2017.09.011] [Medline: 28982034]
9. Jennings LA, Reuben DB, Evertson LC, Serrano KS, Ercoli L, Grill J, et al. Unmet needs of caregivers of individuals referred to a dementia care program. J Am Geriatr Soc 2015 Feb 17;63(2):282-289 [FREE Full text] [doi: 10.1111/jgs.13251] [Medline: 25688604]
10. Peeters JM, Van Beek AP, Meerveld JH, Spreeuwenberg PM, Francke AL. Informal caregivers of persons with dementia, their use of and needs for specific professional support: a survey of the National Dementia Programme. BMC Nurs 2010 Jun 07;9(1):9 [FREE Full text] [doi: 10.1186/1472-6955-9-9] [Medline: 20529271]
11. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-Based Interventions Aimed at Supporting Family Caregivers of People With Dementia: Systematic Review. J Med Internet Res 2018 Jun 12;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
12. Joling KJ, O'Dwyer ST, Hertogh CM, van Hout HP. The occurrence and persistence of thoughts of suicide, self-harm and death in family caregivers of people with dementia: a longitudinal data analysis over 2 years. Int J Geriatr Psychiatry 2018 Feb 05;33(2):263-270 [FREE Full text] [doi: 10.1002/gps.4708] [Medline: 28379646]
13. Alzheimer's Disease International (ADI). World Alzheimer Report 2019: Attitudes to Dementia. 2019. URL: https://www.alz.co.uk/research/world-report-2019 [accessed 2020-04-16]
14. Sepe-Monti M, Vanacore N, Bartorelli L, Tognetti A, Giubilei F, Caregiver Study Group Savvy. The Savvy Caregiver Program: A Probe Multicenter Randomized Controlled Pilot Trial in Caregivers of Patients Affected by Alzheimer's Disease. J Alzheimers Dis 2016 Oct 04;54(3):1235-1246. [doi: 10.3233/JAD-160235] [Medline: 27567824]
15. Woods RT, Orrell M, Bruce E, Edwards RT, Hoare Z, Hounsome B, et al. REMCARE: Pragmatic Multi-Centre Randomised Trial of Reminiscence Groups for People with Dementia and their Family Carers: Effectiveness and Economic Analysis. PLoS One 2016 Apr 19;11(4):e0152843 [FREE Full text] [doi: 10.1371/journal.pone.0152843] [Medline: 27093052]
16. Bjørge H, Kvaal K, Ulstein I. The effect of psychosocial support on caregivers' perceived criticism and emotional over-involvement of persons with dementia: an assessor-blinded randomized controlled trial. BMC Health Serv Res 2019 Oct 24;19(1):744 [FREE Full text] [doi: 10.1186/s12913-019-4551-x] [Medline: 31651321]
17. Hamieh N, Sharara E, Salibi N, Mrad P, Chaaya M. Public Knowledge of, Perceptions About and Attitudes Towards Dementia: A Cross-Sectional Survey Among Lebanese Primary Health Care Attenders. Community Ment Health J 2019 Nov 3;55(8):1362-1368. [doi: 10.1007/s10597-019-00436-2] [Medline: 31270647]
18. Brodaty H, Thomson C, Thompson C, Fine M. Why caregivers of people with dementia and memory loss don't use services. Int J Geriatr Psychiatry 2005 Jun;20(6):537-546. [doi: 10.1002/gps.1322] [Medline: 15920707]
19. Novais T, Dauphinot V, Krolak-Salmon P, Mouchoux C. How to explore the needs of informal caregivers of individuals with cognitive impairment in Alzheimer's disease or related diseases? A systematic review of quantitative and qualitative studies. BMC Geriatr 2017 Apr 17;17(1):86 [FREE Full text] [doi: 10.1186/s12877-017-0481-9] [Medline: 28415968]
20. Lewis BA, Williams DM, Neighbors CJ, Jakicic JM, Marcus BH. Cost Analysis of Internet vs. Print Interventions for Physical Activity Promotion. Psychol Sport Exerc 2010 May 01;11(3):246-249 [FREE Full text] [doi: 10.1016/j.psychsport.2009.10.002] [Medline: 20401164]

21. Hayden LJ, Glynn SM, Hahn TJ, Randall F, Randolph E. The use of Internet technology for psychoeducation and support with dementia caregivers. Psychol Serv 2012 May;9(2):215-218. [doi: 10.1037/a0027056] [Medline: 22662739]
22. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-Based Interventions to Improve Mental Health, General Caregiving Outcomes, and General Health for Informal Caregivers of Adults With Chronic Conditions Living in the Community: Rapid Evidence Review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
23. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of Internet-Based Interventions on Caregiver Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2018 Jul 03;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
24. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-Focused, Web-Based Interventions: Systematic Review and Meta-Analysis (Part 2). J Med Internet Res 2018 Oct 26;20(10):e11247 [FREE Full text] [doi: 10.2196/11247] [Medline: 30368439]
25. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec 01;45(6):793-801. [doi: 10.1093/geront/45.6.793] [Medline: 16326661]
26. Gustafson DH, Gustafson DH, Cody OJ, Chih M, Johnston DC, Asthana S. Pilot Test of a Computer-Based System to Help Family Caregivers of Dementia Patients. J Alzheimers Dis 2019;70(2):541-552. [doi: 10.3233/JAD-190052] [Medline: 31256126]
27. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT Platform to Help Informal Caregivers of People with Dementia: A Pilot Randomized Controlled Study. Biomed Res Int 2016;2016:5726465 [FREE Full text] [doi: 10.1155/2016/5726465] [Medline: 28116300]
28. Van Mierlo LD, Meiland FJ, Van de Ven PM, Van Hout HP, Dröes RM. Evaluation of DEM-DISC, customized e-advice on health and social support services for informal carers and case managers of people with dementia; a cluster randomized trial. Int Psychogeriatr 2015 Aug;27(8):1365-1378. [doi: 10.1017/S1041610215000423] [Medline: 25872457]
29. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015 May 12;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
30. Meichsner F, Theurer C, Wilz G. Acceptance and treatment effects of an internet-delivered cognitive-behavioral intervention for family caregivers of people with dementia: A randomized-controlled trial. J Clin Psychol 2019 Apr;75(4):594-613. [doi: 10.1002/jclp.22739] [Medline: 30597537]
31. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
32. McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia-a systematic review. Int Psychogeriatr 2014 Oct;26(10):1619-1637 [FREE Full text] [doi: 10.1017/S1041610214001045] [Medline: 24989249]
33. Zhao Y, Feng H, Hu M, Hu H, Li H, Ning H, et al. Web-Based Interventions to Improve Mental Health in Home Caregivers of People With Dementia: Meta-Analysis. J Med Internet Res 2019 May 06;21(5):e13415 [FREE Full text] [doi: 10.2196/13415] [Medline: 31066680]
34. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
35. Deeks JH, Altman D. Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions. London: Cochrane Collaboration; 2008:211-242.
36. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
37. Blom MM, Zarit SH, Groot Zwaaftink RBM, Cuijpers P, Pot AM. Effectiveness of an Internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015 Feb 13;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
38. Brennan PF, Moore SM, Smyth KA. The effects of a special computer network on caregivers of persons with Alzheimer's disease. Nurs Res 1995;44(3):166-172. [Medline: 7761293]
39. Duggleby W, Ploeg J, McAiney C, Peacock S, Fisher K, Ghosh S, et al. Web-Based Intervention for Family Carers of Persons with Dementia and Multiple Chronic Conditions (My Tools 4 Care): Pragmatic Randomized Controlled Trial. J Med Internet Res 2018 Jun 29;20(6):e10484 [FREE Full text] [doi: 10.2196/10484] [Medline: 29959111]
40. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-Based STAR E-Learning Course Increases Empathy and Understanding in Dementia Caregivers: Results from a Randomized Controlled Trial in the Netherlands and the United Kingdom. J Med Internet Res 2015 Oct 30;17(10):e241 [FREE Full text] [doi: 10.2196/jmir.4025] [Medline: 26519106]
41. Hicken BL, Daniel C, Luptak M, Grant M, Kilian S, Rupper RW. Supporting Caregivers of Rural Veterans Electronically (SCORE). J Rural Health 2017 Jun;33(3):305-313. [doi: 10.1111/jrh.12195] [Medline: 27437642]

42. Kales HC, Gitlin LN, Stanislawski B, Myra Kim H, Marx K, Turnwald M, et al. Effect of the WeCareAdvisor™ on family caregiver outcomes in dementia: a pilot randomized controlled trial. BMC Geriatr 2018 May 10;18(1):113 [FREE Full text] [doi: 10.1186/s12877-018-0801-8] [Medline: 29747583]
43. Metcalfe A, Jones B, Mayer J, Gage H, Oyebode J, Boucault S, et al. Online information and support for carers of people with young-onset dementia: A multi-site randomised controlled pilot study. Int J Geriatr Psychiatry 2019 Oct;34(10):1455-1464. [doi: 10.1002/gps.5154] [Medline: 31111516]
44. Possin KL, Merrilees JJ, Dulaney S, Bonasera SJ, Chiong W, Lee K, et al. Effect of Collaborative Dementia Care via Telephone and Internet on Quality of Life, Caregiver Well-being, and Health Care Use. JAMA Intern Med 2019 Dec 01;179(12):1658 (forthcoming). [doi: 10.1001/jamainternmed.2019.4101]
45. Torkamani M, McDonald L, Saez Aguayo I, Kanios C, Katsanou M, Madeley L, ALADDIN Collaborative Group. A randomized controlled pilot study to evaluate a technology platform for the assisted living of people with dementia and their carers. J Alzheimers Dis 2014;41(2):515-523. [doi: 10.3233/JAD-132156] [Medline: 24643137]
46. Williams K, Perkhounkova Y, Shaw C, Hein M, Vidoni E, Coleman C. Supporting Family Caregivers With Technology for Dementia Home Care: A Randomized Controlled Trial. Innov Aging 2019 Jul;3(3):igz037 [FREE Full text] [doi: 10.1093/geroni/igz037] [Medline: 31660443]
47. American Geriatrics Society Expert Panel on Person-Centered Care T. Person-Centered Care: A Definition and Essential Elements. J Am Geriatr Soc 2016 Jan 02;64(1):15-18. [doi: 10.1111/jgs.13866] [Medline: 26626262]
48. National Institute for Health and Care Excellence (UK). National Institute for Health and Care Excellence: Clinical Guidelines. Dementia: Assessment, Management and Support for People Living With Dementia and Their Carers. London: National Institute for Health and Care Excellence; 2018.
49. Fazio S, Pace D, Maslow K, Zimmerman S, Kallmyer B. Alzheimer's Association Dementia Care Practice Recommendations. Gerontologist 2018 Jan 18;58(suppl_1):S1-S9. [doi: 10.1093/geront/gnx182] [Medline: 29361074]
50. Australian and New Zealand Society for Geriatric Medicine. Position Statement 26 Management of Behavioural and Psychological Symptoms of Dementia (BPSD). Sydney, NSW: Australian and New Zealand Society for Geriatric Medicine; 2016.
51. Dauphinot V, Ravier A, Novais T, Delphin-Combe F, Mouchoux C, Krolak-Salmon P. Risk Factors of Caregiver Burden Evolution, for Patients With Subjective Cognitive Decline or Neurocognitive Disorders: A Longitudinal Analysis. J Am Med Dir Assoc 2016 Nov 01;17(11):1037-1043. [doi: 10.1016/j.jamda.2016.07.003] [Medline: 27575984]
52. Swinkels J, Broese van Groenou MI, de Boer A, Tilburg T. Male and Female Partner-Caregivers' Burden: Does It Get Worse Over Time? Gerontologist 2019 Nov 16;59(6):1103-1111. [doi: 10.1093/geront/gny132] [Medline: 30321338]
53. Morrin H, Fang T, Servant D, Aarsland D, Rajkumar AP. Systematic review of the efficacy of non-pharmacological interventions in people with Lewy body dementia. Int. Psychogeriatr 2017 Oct 09;30(3):395-407. [doi: 10.1017/s1041610217002010]
54. Caspar S, Davis ED, Douziech A, Scott DR. Nonpharmacological Management of Behavioral and Psychological Symptoms of Dementia: What Works, in What Circumstances, and Why? Innov Aging 2018 Jan;2(1):igy001 [FREE Full text] [doi: 10.1093/geroni/igy001] [Medline: 30480128]
55. Gilhooly KJ, Gilhooly MLM, Sullivan MP, McIntyre A, Wilson L, Harding E, et al. A meta-review of stress, coping and interventions in dementia and dementia caregiving. BMC Geriatr 2016 May 18;16:106 [FREE Full text] [doi: 10.1186/s12877-016-0280-8] [Medline: 27193287]
56. Schaller S, Marinova-Schmidt V, Gobin J, Criegee-Rieck M, Griebel L, Engel S, et al. Tailored e-Health services for the dementia care setting: a pilot study of 'eHealthMonitor'. BMC Med Inform Decis Mak 2015 Jul 28;15:58 [FREE Full text] [doi: 10.1186/s12911-015-0182-2] [Medline: 26215731]

Abbreviations

DEM-DISC: DEMentia Digital Interactive Social Chart
MD: mean difference
RCT: randomized controlled trials
SMD: Standardized mean differences

Edited by G Eysenbach; submitted 19.04.20; peer-reviewed by C Li, M Hu; comments to author 12.06.20; revised version received 25.06.20; accepted 26.07.20; published 09.09.20.

Please cite as:
Leng M, Zhao Y, Xiao H, Li C, Wang Z
Internet-Based Supportive Interventions for Family Caregivers of People With Dementia: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(9):e19468
URL: http://www.jmir.org/2020/9/e19468/
doi:10.2196/19468
PMID:32902388

Review

Effectiveness of Individual Real-Time Video Counseling on Smoking, Nutrition, Alcohol, Physical Activity, and Obesity Health Risks: Systematic Review

Judith Byaruhanga[1,2,3,4], MPH; Prince Atorkey[1,2,3,4], MPhil; Matthew McLaughlin[1,2,3,4], BSc; Alison Brown[1,2,3,4], BClinSc, BNutrDiet (Hons); Emma Byrnes[1,3], BSocSc (Hons); Christine Paul[1,3,4], PhD; John Wiggers[1,2,3,4], PhD; Flora Tzelepis[1,2,3,4], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton Heights, Australia

[4]Priority Research Centre for Health Behaviour, Faculty of Health & Medicine, University of Newcastle, Callaghan, Australia

Corresponding Author:
Judith Byaruhanga, MPH
Hunter New England Population Health
Locked Bag 10
Wallsend,
Australia
Phone: 61 249246454
Email: judith.byaruhanga@uon.edu.au

Abstract

Background: Real-time video communication technology allows virtual face-to-face interactions between the provider and the user, and can be used to modify risk factors for smoking, nutrition, alcohol consumption, physical activity, and obesity. No systematic reviews have examined the effectiveness of individual real-time video counseling for addressing each of the risk factors for smoking, nutrition, alcohol consumption, physical activity, and obesity.

Objective: This systematic review aims to examine the effectiveness of individually delivered real-time video counseling on risk factors for smoking, nutrition, alcohol consumption, physical activity, and obesity.

Methods: The MEDLINE (Medical Literature Analysis and Retrieval System Online), EMBASE (Excerpta Medica Database), PsycINFO, Cochrane Register of Controlled Trials, and Scopus databases were searched for eligible studies published up to November 21, 2019. Eligible studies were randomized or cluster randomized trials that tested the effectiveness of individual real-time video communication interventions on smoking, nutrition, alcohol, physical activity, and obesity in any population or setting; the comparator was a no-intervention control group or any other mode of support (eg, telephone); and an English-language publication.

Results: A total of 13 studies were eligible. Four studies targeted smoking, 3 alcohol, 3 physical activity, and 3 obesity. In 2 of the physical activity studies, real-time video counseling was found to significantly increase physical activity when compared with usual care at week 9 and after 5 years. Two obesity studies found a significant change in BMI between a video counseling and a documents group, with significantly greater weight loss in the video counseling group than the in-person as well as the control groups. One study found that those in the video counseling group were significantly more likely than those in the telephone counseling group to achieve smoking cessation. The remaining studies found no significant differences between video counseling and telephone counseling or face-to-face counseling for smoking cessation, video counseling and face-to-face treatment on alcohol consumption, video counseling and no counseling for physical activity, and video counseling and face-to-face treatment on BMI. The global methodological quality rating was moderate in 1 physical activity study, whereas 12 studies had a weak global rating.

Conclusions: Video counseling is potentially more effective than a control group or other modes of support in addressing physical inactivity and obesity and is not less effective in modifying smoking and alcohol consumption. Further research is required to determine the relative benefits of video counseling in terms of other policy and practice decision-making factors such as costs and feasibility.

(J Med Internet Res 2020;22(9):e18621) doi:10.2196/18621

KEYWORDS

telehealth; videoconferencing; smoking cessation; diet; alcohol drinking; physical activity; obesity; mobile phone

Introduction

Background

Tobacco use, poor nutrition, risky alcohol consumption, physical inactivity, and obesity are the leading modifiable health risks that can cause noncommunicable diseases, including cardiovascular disease, chronic respiratory disease, cancer, stroke, and diabetes [1]. Globally, it is estimated that there are 1.1 billion tobacco smokers, and tobacco use is responsible for the death of 8 million people each year [2]. Harmful alcohol consumption is responsible for 3 million deaths and causes >200 chronic and acute diseases [3]. Globally, in 2016, an estimated 0.9 million injury deaths and 52.4 million injury disability-adjusted life years (DALYs) were attributable to alcohol [3]. Similarly, poor diet accounts for 11 million adult deaths [4], of which 3 million annual deaths are attributed to excess salt or sodium intake [5], 2 million deaths per year are attributable to diets low in fruits and vegetables, and 3 million deaths are attributable to low intake of whole grains [4,6]. Insufficient physical activity causes 5.3 million premature deaths annually [7].

Real-time video communication, also known as videoconferencing, telehealth, or telecare [8], is a scalable and accessible intervention delivered over the internet via a video camera connected to a computer, smartphone, or tablet [9]. Real-time video communication is available to the 3.9 billion people who have access to the internet worldwide [10] and have a device with a video camera. Video communication software such as Skype can be downloaded for free by internet users and is widely used for personal or professional communication every day [11]. Real-time video technology allows a real-time, virtual face-to-face interaction between the provider and the user [12] at any time of the day in any location with internet access [13]. Another advantage of real-time video counseling is that it provides a mode for delivering individual counseling that allows counselors to respond to the client's verbal and nonverbal cues, unlike telephone support, which is an audio-only intervention, or written materials. Real-time video counseling is supported by the media richness theory, which conceptualizes that real-time video counseling ranks highly as a rich mode of communication as it provides virtual face-to-face support and advisers are able to respond to nonverbal cues, which is not possible in all forms of behavioral support [14]. There is an opportunity for instant clarification of doubts or identifying reluctance or enthusiasm in both the voice and body language, consequently minimizing chances of being misunderstood. Real-time video counseling also eliminates travel time and associated costs of face-to-face interventions [15], improves discretion and comfort as the video call can be taken in a preferred private place to avoid the potential stigma associated with clinic visits [15,16], and has widespread reach. Real-time video counseling has the potential to be used on a large scale to target health risks attributable to smoking, nutrition, alcohol consumption, physical activity, and obesity [9,10].

Individual counseling is used by service providers to deliver support for smoking cessation [17], nutrition [18], physical activity [19], obesity [19], and alcohol consumption [20]. Individual counseling is primarily delivered in person or via the telephone by service providers [21-24], and these modes have been found to be effective in improving health risks attributable to smoking, nutrition, alcohol consumption, physical activity, and obesity [17-19,25-27]. The advantages of individual counseling include convenient scheduling for both the client and service provider, focused treatment with personalized feedback, and a high level of confidentiality as clients can disclose themselves in private compared with group counseling [28].

The capability, opportunity, motivation, and behavior (COM-B) model by Michie et al [29] provides a theoretical framework to examine the effectiveness of real-time video counseling on health risks for smoking, nutrition, alcohol consumption, physical activity, and obesity. The COM-B model suggests that behavior is a result of 3 factors: capability (psychological or physical), opportunity (physical or social), and motivation (reflective or automatic) [29]. Real-time video counseling may maximize capability, opportunity, and motivation to encourage behavior change in the following ways. First, a real-time video counseling intervention for health risks of smoking, nutrition, alcohol consumption, physical activity, and obesity can assist participants in realizing their capability by enhancing their knowledge about effective behavior change methods, situations, and environments that act as triggers for their behaviors. Second, real-time video counseling may remove barriers such as the distance and time to travel to access face-to-face treatment, thereby increasing the opportunity for behavior change. Third, real-time video counseling can include motivational interviewing and cognitive behavioral therapy techniques to increase motivation to improve factors for smoking, nutrition, alcohol consumption, physical activity, and obesity [29].

One systematic review has examined the effectiveness of various technology-based interventions on smoking cessation, including real-time video counseling; however, the review only included studies of participants with low socioeconomic status or disadvantaged populations [30]. This review identified only 1 study that found no significant difference between video counseling and telephone counseling on smoking cessation [31]. A Cochrane systematic review examined the effectiveness of real-time video counseling for smoking cessation only and found limited evidence that suggested no difference between video counseling and telephone counseling [32]. However, this systematic review excluded studies that measured smoking cessation <6 months postbaseline [32]. To the best of our knowledge, there are no other systematic reviews that have examined the effectiveness of individual real-time video support

for addressing each of the risk factors for smoking, nutrition, alcohol consumption, physical activity, and obesity.

Objective

This systematic review aimed to examine the effectiveness of individual real-time video counseling on health risks for smoking, nutrition, alcohol, physical inactivity, and obesity relative to (1) a no-intervention control group or (2) other modes of intervention delivery.

Methods

Narrative Review

This narrative review follows the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33] and was completed as per the protocol registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number: CRD42017071885). Meta-analyses were not undertaken because of the heterogeneity between studies (eg, clinical vs nonclinical populations) and the small number of studies examining the effectiveness of real-time video counseling for each risk factor for smoking, nutrition, alcohol consumption, physical activity, and obesity.

Search Strategy

The electronic databases Cochrane Register of Controlled Trials (via Cochrane Library), MEDLINE (Medical Literature Analysis and Retrieval System Online; from 1946), EMBASE (Excerpta Medica dataBASE; from 1947), PsycINFO (from 1806), and Scopus were searched from inception to retrieve studies published up to November 21, 2019, that described a real-time video counseling intervention (eg, video conferencing or video consultation or telehealth or telemedicine) for modifying health risks for smoking, nutrition, alcohol consumption, physical activity, and obesity. The reference lists of included trials were also manually searched to retrieve any other relevant studies.

The database search consisted of focused text word searches and medical subject heading searches. The search terms were divided into 3 groups: (1) smoking, nutrition, alcohol, physical activity, and obesity behavior (ie, tobacco use, nutrition, alcohol drinking, physical activity, obesity, healthy lifestyle, lifestyle), (2) video communication intervention (ie, telemedicine, videoconferencing, remote consultation, Skype, Viber, webcam, Talky Core, WhatsApp, FaceTime, Messenger, Google Hangouts), and (3) study design (ie, randomized controlled trial, cluster randomized trial). Textbox 1 outlines the search strategy.

Textbox 1. Search strategy.

- Nicotine/
- Tobacco/
- exp “Tobacco Use Cessation”/
- exp “Tobacco Use”/
- (Cigar* or smok* or tobacco or nicotine).tw.
- 1 or 2 or 3 or 4 or 5
- exp Healthy Lifestyle/
- exp Life Style/
- (lifestyle* or life style*).tw.
- nutrition*.mp.
- exp Fruit/
- exp Vegetables/
- (fruit* or vegetable*).tw.
- 7 or 8 or 9 or 10 or 11 or 12 or 13
- exp Alcohol Drinking/
- exp Alcoholism/ or exp Drinking Behavior/
- exp Alcoholic Intoxication/
- (Alcohol* or drinking).tw.
- 15 or 16 or 17 or 18
- exp Exercise/
- physical activity.mp.
- exp Sedentary Lifestyle/
- (physical activit* or physical inactivit*).tw.
- (exercise* or Sport*).tw.
- 20 or 21 or 22 or 23 or 24
- exp Overweight/
- Obes*.tw.
- 26 or 27
- exp Telemedicine/
- exp Videoconferencing/
- Remote Consultation/
- (skype or viber or webcam or talky core or whatsapp or facetime or messenger or google* hangouts).mp. (mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms)
- ((real time or realtime) adj3 (counsel* or support* or therap* or conference or consult*)).tw.
- (remote adj3 (communicat* or consult*)).tw.
- 29 or 30 or 31 or 32 or 33 or 34
- 6 or 14 or 19 or 25 or 28
- 35 and 36
- exp Randomized Controlled Trial/
- exp Randomized Controlled Trials as Topic/
- exp Clinical Trial/
- exp Clinical Trials as Topic/

- exp Random Allocation/
- Random*.tw.
- Trial.tw.
- 38 or 39 or 40 or 41 or 42 or 43 or 44
- 37 and 45

Eligibility Criteria

Studies were included in this review if they met the following criteria:

1. Study design: randomized trials or cluster randomized trials. Randomized trials and cluster randomized trials were included because these designs are considered the gold standard for measuring effectiveness [34].
2. Study participants: any population (ie, general population, patients).
3. Setting: any setting, including community and health care settings.
4. Intervention: video communication was used as the mode to deliver individual, one-on-one support (ie, Skype, FaceTime, Facebook Messenger, WhatsApp, or any preferred individual real-time video communication platform).
5. Comparators: the comparators included a no-intervention control group or any other form of support to address the risk factors for smoking, nutrition, alcohol consumption, physical activity, and obesity, such as written materials, telephone counseling, web-based support, and face-to-face interventions.
6. Language: studies published in English.
7. Outcome measures: any measure of an individual's smoking (eg, smoking cessation, quit attempts), nutrition (eg, serves of fruit and/or vegetables, calories), alcohol (eg, number of standard drinks of alcohol), physical activity (eg, number of minutes of moderate or vigorous physical activity or metabolic equivalent [MET] minutes), or obesity (eg, BMI, waist circumference).

Study Selection

After removing duplicate records, 2 authors (JB and FT, PA, or MM) independently screened the titles and abstracts of all records using either EndNote or Covidence. Papers that did not meet the eligibility criteria were excluded. Two reviewers independently examined the full text of the papers that were deemed eligible or whose eligibility was uncertain based on the title and abstract screening. Two reviewers met and discussed any discrepancies until a consensus was reached. The reasons for exclusion were recorded for all full text papers assessed that were ineligible.

Data Extraction of Study Characteristics

Two authors (JB and AB or EB) independently extracted the following data from eligible studies: authors and country, years data collected, study design, sample characteristics, recruitment method, eligibility criteria, participation rate, treatment conditions, the video intervention received, retention at follow-up, outcome measures, the comparators, and costs. All discrepancies were resolved between the 2 reviewers through discussion, and a third reviewer (FT) was consulted when necessary.

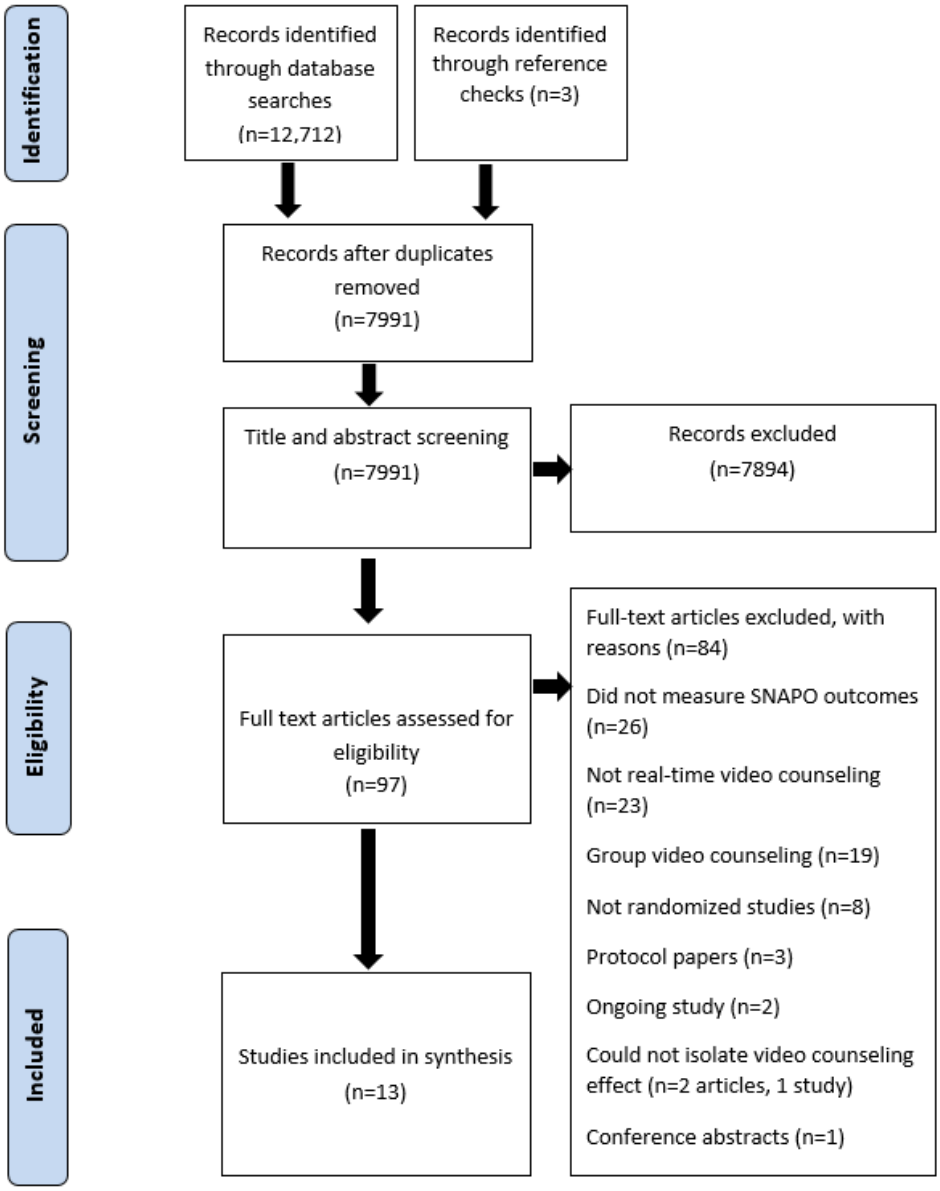
Methodological Quality Assessment

The quality assessment of each included study was assessed independently by 2 reviewers (JB and FT). The Quality Assessment Tool for Quantitative Studies developed by the Effective Public Health Practice Project was used to assess methodological quality [35] according to the instructions described in the Quality Assessment Tool for Quantitative Studies Dictionary [36]. The Quality Assessment Tool for Quantitative Studies assesses randomized and nonrandomized trials in relation to 6 components: selection bias, study design, confounders, blinding, data collection methods, withdrawals, and dropouts. Each study was rated as *strong*, *moderate*, or *weak* on each of these components. An overall global rating was then assigned to each study, with studies classified as *strong* (no weak ratings), *moderate* (1 weak rating), or *weak* (2 weak ratings).

Results

After removing duplicates, a total of 7991 records were screened. Of these, 7894 records were excluded at the title and abstract screening stage, and 97 full text records were assessed for eligibility (Figure 1). A total of 84 of the 97 full text records were excluded for the following reasons: 26 did not measure smoking, nutrition, alcohol consumption, physical activity, and obesity outcomes; 23 did not use any form of video counseling intervention; 19 involved group video counseling and not individual video counseling [37-54]; 3 were protocol papers [55-57]; 8 were not randomized studies [58-65]; 2 studies were ongoing [32,66]; 1 was a conference abstract [67]; and 2 studies described a multicomponent intervention, and it was not possible to isolate the effect of real-time video counseling [68,69]. The remaining 13 eligible studies were included in the review. Figure 1 presents the PRISMA diagram for screening and selection.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of the screening and selection process. SNAPO: smoking, nutrition, alcohol consumption, physical activity, and obesity.



Study Characteristics

Four studies focused on individual video counseling for smoking cessation [31,70-72], 3 studies focused on alcohol consumption [73-75], 3 studies focused on physical activity [76-78], and 3 studies focused on obesity [79-81]. No trial examined the effectiveness of real-time video counseling on nutrition. Most trials were funded by grants from the government [76,77,79-81] or university [71,73]. One study reported receiving no funding [75], 1 study was funded by the Craig H. Neilsen Foundation [78], and 1 study was funded by CureApp Inc [72]. All 13 studies were published from 2010 onward.

Smoking Cessation

Three randomized trials that examined the effectiveness of real-time video counseling compared with telephone counseling for smoking cessation were conducted in the United States [31,70,71]. One trial conducted in Japan compared real-time video counseling with face-to-face counseling [72]. The studies followed-up participants for 3 [70], 6 [71,72], or 12 months [31]. One trial focused on a clinical population of women with HIV [71], whereas 3 studies included nonclinical populations [31,70,72]. The sample sizes were 49 [70,71], 115 [72], and 566 participants [31]. Two studies recruited women only [70,71], whereas another study had a majority of female participants (65%) [31] and 1 study had a majority of male participants (81%) [72]. Across the 4 trials, the mean age was 45 (SD 11.7) to 55 years (SD 11) [31,70-72]. The participation rates were 36.66% (566/1544) [31], 52% (49/94) [71], 64% (49/77) [70], and unclear for 1 study [72]. The retention rates at follow-up were 87.6% (496/566) at 12 months [31], 55% (27/49) [71] and 97.4% (112/115) [72] at 6 months, and 78% (38/49) at 3 months [70].

Multimedia Appendix 1 [31,70-72] provides a detailed description of the study characteristics. The trial in rural United States (Kansas) recruited participants from 20 primary care clinics and through community-based activities (eg, radio

interviews, health fairs, and religious organizations) [31], whereas another trial recruited Korean-American women through web-based communities and newspaper advertisements [70]. The US trial among women living with HIV had participants referred by health workers, professional health networks, advertisements on free websites, and Craiglist [71], whereas Nomura et al [72] recruited participants from community clinics or centers.

One trial delivered the intervention through 4 individually tailored sessions at the clinic [31], whereas participants were offered 5 internet-based video counseling calls in another study [72]. In 2 studies, participants received up to 8 individual video counseling calls at home [70,71]. Participants from the rural US study also received written materials on smoking cessation and pharmacotherapy [31], whereas studies with Korean-American women and women living with HIV offered nicotine patches to their participants [70,71]. Participants were followed-up at 3, 6, and 12 months for 1 study [31] and at the end of the intervention and at 3 and 6 months for another study [71]. In the study with Korean-American women, participants were followed-up for 3 months [70], and in the study conducted in Japan, participants were followed-up at 3 and 6 months [72].

Three studies reported prolonged abstinence, 1 at 3 months [70], 1 at 6 months [71], and 1 at 12 months [31]. One study reported continuous abstinence between 9 and 12 weeks and 9 and 24 weeks [72]. Three studies reported 7-day point prevalence abstinence outcomes at 1, 2, and 3 months [70]; at the end of the intervention; at 3 and 6 months [71]; and at 12 months [31]. Only the rural US study reported provider costs for real-time video counseling and telephone counseling interventions [31].

Alcohol Use

Multimedia Appendix 2 [73-75] provides a detailed description of the study characteristics. One study was conducted in Denmark [73] and the other 2 studies were conducted in the United States [74,75]. Participants were recruited from public outpatient alcohol clinics [73], a web-based research participation system for undergraduate students [75], and community supervision offices for people with substance abuse [74]. One study included a clinical population with an alcohol dependence syndrome [73], whereas the other 2 studies were conducted with nonclinical populations [74,75]. The sample sizes were 51 [75], 71 [73], and 127 [74] across the studies. The proportion of men was high in 2 studies, specifically 73% (52/71) [73] and 81.1% (103/127) [74], whereas the majority were women (60.8%) in 1 study [75]. The mean age was 19 (SD not reported) [75] and 47 years (SD 12.8) [73], with a median age of 30.5 years [74]. The participation rate was 63% [73] and 73% [74] across 2 studies and unclear in 1 study [75]. One trial compared video counseling (telehealth) with face-to-face treatment only [75]. The other 2 trials compared individual video conferencing plus face-to-face support (treatment as usual) with face-to-face only support (treatment as usual) [73,74].

In 1 study, participants received up to 5 sessions with the therapist via videoconference and were followed-up at 3 months [74], whereas another study offered between 1 and 3 sessions a week at the initial stages, followed by 1 session every other week for about 7 months, and participants were followed-up until 12 months [73]. In the third study, participants received only 2 sessions and were followed-up at 1, 2, and 3 months [75].

Alcohol consumption was measured at 3, 6, and 12 months in 1 study [73], at 3 months in another study [74], and at 1, 2, and 3 months in the third study [75]. The costs of the video intervention and face-to-face support (treatment as usual) were not reported in any of these 3 studies [73-75].

Physical Activity

Three trials examined the effectiveness of individual video counseling to increase physical activity. The trials were conducted in Australia [76], the United States [77], and Canada [78]. Participants were recruited from print and web advertising in 1 study [76], whereas in the other 2 studies, participants were recruited from their primary care provider [77] or from outpatient rehabilitation hospitals in Montreal, a local adapted fitness center, an organization representing persons with spinal cord injury (SCI), pre-existing databases of previous research participants, and social media platforms [78]. Two physical activity trials were conducted with clinical populations, specifically people with paraplegia [78] and people with diabetes [77], whereas 1 study was conducted with a nonclinical population [76]. The sample sizes across the 3 studies were 24 [78], 154 [76], and 1650 participants [77]. In 2 trials, most participants were female (117/154, 76.0% [76] and 1037/1650, 62.84% [77]), whereas in the third study, the majority were male 73% (16/24) [78]. The average BMI was 31 kg/m^2 [76] and 32 kg/m^2 [77] in 2 studies, and BMI was not reported in the Canadian study [78]. Across the studies, the mean age ranged from 51.64 (SD 12.3) [78] to 70.9 (SD 6.63) years [77]. One trial compared real-time video counseling plus computer-tailored advice with computer-tailored web-based physical activity intervention only (advice tailored to an individual with graphs and text) and to a waitlist control group [76]. The intervention participants in this trial received tailored physical activity advice plus video counseling every 2 weeks for 8 weeks (“My Activity Coach”) [76]. These were four 10-min coaching sessions with a behavioral expert using a web-based video-calling program (Skype) compared with computer-tailored physical activity advice only and a waitlist control. Participants were followed-up at week 9 and at 6 months [76]. In the second study [77], the intervention comprised video counseling (home video calls) with a diabetes educator conducted every 4 to 6 weeks for self-management, which was compared with face-to-face care (usual clinic-based care), and participants were followed-up for 5 years [77]. The third trial compared video counseling (intervention) versus regular routine (control) in adults with SCI in Canada [78]. The intervention comprised 1 leisure time physical activity (LTPA) counseling session per week for 8 weeks, resulting in a total of 8 counseling sessions [78]. Participants were then followed-up at weeks 6 and 10 [78].

One trial assessed physical activity in minutes per week at week 9 and 6 months [76], another assessed the rate of decline in physical activity in older participants over a 5-year period [77], and the third trial assessed total LTPA at 6 and 10 weeks [78]. One trial reported a retention rate of 92% at 10 weeks [78], and 2 trials reported retention rates of <50% [76,77]. No cost

information was provided for any of the studies. Multimedia Appendix 3 [76-78] provides a detailed description of the study characteristics.

Obesity

Multimedia Appendix 4 [79-81] provides a description of the study characteristics examining the effectiveness of video counseling on obesity. Three trials [79-81] used individually delivered real-time video counseling to target obesity. The trials targeted clinical populations with lifestyle conditions such as diabetes [79,80], hypertension [80], and overweight or obesity (BMI>25 kg/m^2) [79-81]. One study was conducted in Japan [80], another study in Denmark [79], and the third in the United States [81]. In 1 study, the participants were recruited through telephone calls from outpatient departments [79] and in another study via community advertisements [80], whereas it was unclear how participants were recruited in the third study [81]. The sample sizes were 30 [81], 68 [80], and 165 [79]. The mean age was 66 years (SD 1.7) [80] and 58 years (SD 9.3) [79] in 2 studies and ranged from 42.2 to 44.5 years across the 3 groups in the third study [81]. The majority of participants were male 63.8% (106/166) in 1 trial [79] and female 65% (44/68) in another trial [80], whereas the gender distribution was not reported in the third trial [81]. The trials reported BMI and physical activity outcomes at 3 months [80], BMI and waist to hip ratio at 8 and 14 months [79], and physical activity and body weight loss over 12 weeks [81].

The video counseling intervention was compared with either individualized documented reports (individualized written reports at 3 time points addressing lifestyle modifications) [80] or usual care (face-to-face) [79] or face-to-face or a control group that received no feedback from mobile health devices and no health coaching sessions [81]. In 1 study, real-time video consultations were delivered 3 times in 3 months [80]. In another study, they used video add-ons in usual clinic-based care (every 3-6 months) with a health care center nurse via a tablet for 32 weeks [79]. In the third study, the video counseling intervention participants received health coaching educational materials and weekly individualized videoconferencing by a multidisciplinary team (registered dietitian, exercise physiologist, and medical doctor) based on data uploaded over the 12-week intervention [81]. None of the trials provided any information on the cost of the interventions.

Effectiveness of Real-Time Video Counseling on Smoking Cessation

In the nonclinical populations [31,70,72], Richter et al found no significant difference between video counseling and telephone counseling for self-reported 7-day point prevalence abstinence at 3 months and 6 months and reported no significant difference between video counseling (9.8%) and telephone counseling (12%) in biochemically verified 7-day point prevalence abstinence and prolonged abstinence (video 8.1% and telephone 7.6%) at 12 months [31]. Kim et al [70] reported no significant difference between biochemically validated 7-day point prevalence abstinence in the video counseling arm (33.3%) compared with the telephone counseling arm (28%) at 3 months. Prolonged abstinence also did not differ significantly between the video counseling arm (29.2%) and the telephone counseling arm (28%) [70]. Nomura et al [72] found no significant difference between video counseling and face-to-face sessions for biochemically validated continuous abstinence rate between weeks 9 to 12 (video 81.0%; face-to-face 78.9%) and weeks 9 to 24 (video 74.1%; face-to-face 71.9%).

In a study conducted with women living with HIV, a clinical population, the video counseling group was significantly more likely than the telephone counseling group to achieve biochemically verified point prevalence abstinence at 3 months (video 33.3%; telephone 4.8%) and 6 months postquitting (video 38.1%; telephone 4.8%) [71]. This study also found that those in the video counseling group were significantly more likely than the telephone counseling group to achieve a 6-month prolonged abstinence (video 33.3%; telephone 4.8%) [71].

Effectiveness of Real-Time Video Counseling on Alcohol Use

Two studies were conducted in a nonclinical population [74,75]. In 1 study [74], compared with usual care (social service clinician) only, the real-time video communication group did not significantly differ on any alcohol consumption, days of drinking, drinks per week, and days experiencing alcohol problems at 3 months. In the second study, there was no significant difference in the change in Alcohol Use Disorders Identification Test scores between the video support group and the face-to-face support group from baseline to 1 month posttreatment and 1 to 3 months posttreatment [75]. Similarly, there was no significant difference in Rutgers Alcohol Problem Index (RAPI) scores between the groups at 1-month follow-up and the decrease in RAPI scores from baseline to 1 month and between 1- and 3-month follow-ups [75].

Only a single study was conducted in a clinical population (alcohol dependent) [73]. Tarp et al [73] found no significant difference between video counseling options and usual face-to-face care in the change from baseline to 12 months in the number of days of alcohol consumption in the past month and days of excessive alcohol consumption in the past month.

Effectiveness of Real-Time Video Counseling on Physical Activity

One study for physical activity was conducted among a nonclinical population [76]. In this study, there was a significant change in physical activity (minutes per week) from baseline to week 9 between the tailoring and video-coaching intervention for physical activity and the control group, but there was no significant change between the tailoring and video-coaching intervention and the tailoring-only intervention [76]. From baseline to 6 months, the change in physical activity (minutes per week) did not significantly differ between the tailoring plus video-coaching intervention and either of the other groups [76].

Two of the physical activity studies were conducted among a clinical population [77,78].

In the study by Weinstock et al [77] among people with diabetes, there was a significantly lower rate of decline in physical activity over time in the video counseling group than in the usual care group. In the second study among people with paraplegia, Chemtob et al found that compared with the control group, the

video counseling group reported greater total minutes of LTPA at 6 weeks (Hedge g=0.87) and 10 weeks (Hedge g=0.85) [78]. For moderate and vigorous physical activity, moderate effect sizes were found at 6 weeks (Hedge g=0.52) and small effect sizes were found at 10 weeks (Hedge g=0.34) favoring the video counseling group over the control group [78].

Effectiveness of Real-Time Video Counseling on Obesity

All studies examining the effectiveness of real-time video counseling on obesity were conducted with clinical populations [79-81]. One study compared video counseling with usual care and found no changes in BMI or waist to hip ratio [79]. The second study found a significant change in BMI from preintervention to postintervention (3 months) between the video counseling intervention and the individualized monthly document reports group but no significant change between the groups in average steps per day from preintervention to postintervention [80]. The third study by Johnson et al [81] found that the video counseling group achieved significantly greater weight loss from baseline to 12 weeks than the in-person group and the control group. This study also reported that the video counseling group had significantly higher steps per day than the in-person group at week 4 and the control group at weeks 6, 8, 9, and 11 [81].

Satisfaction With Real-Time Video Counseling for Smoking Cessation

Two studies [31,71] compared the satisfaction of real-time video counseling for smoking cessation with telephone counseling. In 1 study with a nonclinical population [31], those in the video counseling group (97%) were significantly more likely to recommend the program to family and friends than those in the telephone counseling arm (91.9%), but no between-group differences were found for other satisfaction measures. In the other study with a clinical population, there was no significant difference in mean satisfaction scores between the video counseling and the telephone counseling groups [71].

Satisfaction With Real-Time Video Counseling for Alcohol Use

Of the 3 [73-75] studies on alcohol consumption, 1 study [75] assessed satisfaction. This study compared treatment satisfaction between the video counseling group and the face-to-face support group in a nonclinical population and found no significant difference between the 2 groups for client satisfaction questionnaire scores at either session 1 or session 2 [75].

Satisfaction With Real-Time Video Counseling for Physical Activity

One study on physical activity in a nonclinical population [76] reported on the satisfaction between tailoring and video coaching compared with tailoring-only. Alley et al [76] found no significant difference between these groups on program satisfaction scores.

Satisfaction With Real-Time Video Counseling for Obesity

All 3 studies conducted with clinical populations [79-81] that focused on obesity did not assess satisfaction with video counseling compared with the comparator used.

Methodological Quality Assessment for Real-Time Video Counseling Studies

Table 1 outlines the methodological quality ratings for each study across the 6 components (selection bias, study design, confounders, blinding, data collection methods, withdrawals, and dropouts) and the overall global rating. In terms of the global rating, only 1 study was rated as moderate (a physical activity trial) [77], whereas 12 studies had a weak global rating (4 smoking cessation trials [31,70-72], 3 alcohol trials [73-75], 2 physical activity trials [76,78], and 3 obesity trials [79-81]).

Table 1. Methodological quality assessment of eligible studies.

Study	Selection bias	Study design	Confounders	Blinding	Data collection methods	Withdrawals and dropouts	Global rating
Smoking							
Kim et al [70]	Weak	Strong	Weak	Weak	Strong	Moderate	Weak
Kim et al [71]	Weak	Strong	Weak	Weak	Strong	Weak	Weak
Nomura et al [72]	Weak	Strong	Strong	Weak	Strong	Strong	Weak
Richter et al [31]	Weak	Strong	Strong	Weak	Strong	Strong	Weak
Alcohol							
King et al [75]	Weak	Strong	Moderate	Weak	Strong	Weak	Weak
Staton-Tindall et al [74]	Moderate	Strong	Weak	Weak	Weak	Strong	Weak
Tarp et al [73]	Weak	Strong	Strong	Weak	Strong	Moderate	Weak
Physical activity							
Alley et al [76]	Weak	Strong	Weak	Weak	Strong	Weak	Weak
Chemtob et al [78]	Weak	Strong	Weak	Moderate	Strong	Strong	Weak
Weinstock et al [77]	Weak	Strong	Strong	Moderate	Moderate	Strong	Moderate
Obesity							
Hansen et al [79]	Weak	Strong	Strong	Weak	Strong	Strong	Weak
Homma et al [80]	Weak	Strong	Weak	Weak	Strong	Strong	Weak
Johnson et al [81]	Weak	Strong	Strong	Weak	Strong	Strong	Weak

Most studies (n=12) were rated as weak for selection bias because the participation rate was <60% or unclear [31,70-73,75-81]. Blinding of either the outcome assessor or the participant was also another component where most (n=11) studies were rated as weak [31,70-76,79-81]. Regarding confounders, 6 studies had a weak rating because <60% of the potential confounders were controlled for [80] or it was unclear whether potential confounders were controlled for [70,71,74,76,78]. Data collection was only weak in a single study [74] because it was unclear whether the tools used were reliable. Three studies were rated as weak in relation to withdrawals and dropouts because they reported low overall retention rates of 30% [75] and 38% [76] or low retention for one arm of the study (48%) [71].

Discussion

Principal Findings

This is the first review to examine the effectiveness of individual real-time video counseling on smoking, nutrition, alcohol consumption, physical activity, and obesity. This review focused on real-time video communication technology, an emerging intervention delivery mode. The overall results suggest that video counseling is neither more nor less effective in modifying smoking and alcohol consumption but may have particular benefits for addressing physical inactivity and obesity. Given that the effectiveness of video counseling was similar to conventional methods used to treat smoking and alcohol consumption and that many individuals with nicotine dependence or alcohol dependence may not join and complete conventional treatment [82,83], video counseling provides another option to engage people with nicotine dependence or alcohol dependence who are unlikely to use conventional treatment or drop out of such support. If real-time video counseling is at least equally effective to existing treatments such as face-to-face interventions, then the smoking, nutrition, alcohol consumption, physical activity, and obesity program providers should consider including video counseling as an additional option into their services. The importance of a variety of delivery modes has been demonstrated during the COVID-19 pandemic, where access to face-to-face services has been restricted, whereas in contrast, real-time video counseling for risks for smoking, nutrition, alcohol consumption, physical activity, and obesity is sustainable in this context. The cost of video counseling compared with other modes of delivery is difficult to determine because it was reported in only 1 smoking cessation study [31], which required participants to travel to the clinic to receive video sessions (instead of receiving video sessions at home).

Of the 4 studies that examined the effectiveness of video counseling on smoking cessation [31,70-72], only 1 study reported a significant difference between video counseling and telephone counseling at the 3- and 6-month follow-up, which favored the video counseling group [71]. All studies that focused on smoking cessation were comparative effectiveness trials, and there is currently no evidence available on the effectiveness of real-time video counseling compared with a no-intervention or minimal support (eg, written self-help materials) control group for smoking cessation. The global methodological quality rating of the 4 studies that assessed the effectiveness of real-time

video counseling for smoking cessation was weak, suggesting that the methodological rigor of the evidence needs to be improved, particularly in relation to blinding. However, given the nature of trials that examine the effectiveness of real-time video counseling, blinding would be difficult [84]. Additionally, 3 of the 4 studies were conducted in specific populations, such as Korean-American women [70], women living with HIV [71], and rural smokers [31], hence limiting the generalizability of the findings. Given that quitlines provide telephone counseling as part of their standard practices [85,86], and 2 studies report no differences between telephone counseling and video counseling for smoking cessation [31,70] whereas 1 study suggests that video counseling for smoking cessation is superior to telephone counseling [71], quitline providers could consider expanding their routine services to include real-time video counseling.

The evidence in 3 studies indicated that there was no significant difference between real-time video counseling and face-to-face counseling (usual care) for reducing alcohol consumption [73-75]. All studies had a weak global rating with small sample sizes and low retention rates, which resulted in limited power to detect any differences. Moreover, one of the studies included a largely white population (98%) [74] and therefore may have limited generalizability with respect to other cultures and populations, such as those in low- and middle-income countries. Nonetheless, given that real-time video counseling overcomes barriers associated with face-to-face treatment for alcohol consumption such as time and distance [87], and no differences were found between face-to-face treatment (usual care) and video consultations [73-75], service providers could consider offering real-time video counseling as an additional option for modifying alcohol consumption.

Real-time video counseling was found to significantly increase physical activity when compared with usual care at week 9 [76] and after 5 years [77]. However, given the limitations in methodological quality and the paucity of research in this field, further randomized trials examining the effectiveness of real-time video counseling on physical activity are warranted. Given that the existing studies focused on high-income countries, included only obese or diabetic populations, and included participants who were predominantly women, white, and highly educated, the generalizability of the findings to other populations may be limited. Despite the limited evidence, the existing research suggests that real-time video counseling is more effective than usual care for improving physical activity. Therefore, physical activity service providers could consider offering real-time video counseling as part of their routine practice.

Two studies that focused on obesity reported a significant change in BMI from preintervention to 3 months between the video counseling intervention and the individualized monthly document reports group [80], and the video counseling group achieved significantly greater weight loss from baseline to 12 weeks than the in-person group and control group [81]. Only 1 study found no changes in BMI and waist to hip ratio between the video add-on group and the face-to-face treatment group (usual clinic-based care) [79]. Two of the 3 studies that focused on obesity also reported physical activity outcomes. There was a significant difference for 1 study reporting on increasing steps per day for video counseling compared with the control group that favored the video counseling group [81], whereas 1 study found no difference in change in average steps per day between video counseling and individualized documented reports [80]. All 3 studies were rated as weak for selection bias and therefore were unlikely to be representative of the target population. There is some evidence to suggest that real-time video counseling is effective for obesity. Further randomized trials assessing the effectiveness of real-time video counseling with robust methodological quality on obesity are required.

It is worth noting that 7 studies [71,73,77-81] have focused on clinical populations (1 HIV [smoking trial] [71], 3 diabetes [1 physical activity and 2 obesity trials] [77,79,80], 1 alcohol dependence syndrome [alcohol trial] [73], 1 paraplegia [physical activity] [78], and 3 obesity [obesity trials] [79-81]). Four of these studies in clinical populations found that video counseling was superior to the comparator [71,77,80,81], whereas video counseling was as effective as the comparator in 3 trials [73,78,79]. Six studies [31,70,72,74-76] focused on nonclinical populations. Five of the studies with nonclinical populations reported video counseling to be as effective as the comparator group [31,70,72,74,75], whereas video counseling was superior to the comparator in 1 study in the short term but not in the longer term [76]. To expand the evidence, future research is needed to examine the effectiveness of real-time video counseling for smoking, nutrition, alcohol consumption, physical activity, and obesity behaviors in both clinical and nonclinical populations.

Four studies (1 smoking [31], 1 alcohol [74], and 2 physical activity [76,77]) were either conducted exclusively in rural areas or rural and/or regional areas were targeted along with urban locations as part of recruitment. One study with rural residents reported no difference between video counseling and telephone counseling for smoking cessation [31] and another study found no difference between video counseling and face-to-face support on alcohol consumption [74]. In studies that targeted rural and/or regional areas along with urban locations, real-time video counseling was found to significantly increase physical activity compared with usual care at week 9 [76] and after 5 years [77]. Given that rural populations may face challenges accessing services because of distance, real-time video counseling, which is either as effective or more effective than control or comparator interventions, may overcome barriers to accessing smoking, nutrition, alcohol consumption, physical activity, and obesity services in rural locations.

Satisfaction with video counseling was compared with a comparator group in 2 smoking trials [31,71], 1 alcohol trial [75], and 1 physical activity trial [76]. Three of these 4 studies reported no significant differences between real-time video counseling and the comparator group in terms of satisfaction [71,75,76], whereas 1 smoking cessation trial reported that those in the video counseling group were more likely to recommend the program to family and friends than those in the telephone counseling arm [31]. Overall, these results suggest that in terms of satisfaction, those offered real-time video counseling to address smoking, alcohol consumption, and physical activity risks are at least as satisfied with this program as those offered

other conventional methods. This provides further support for the potential of real-time video counseling to be integrated into existing preventive care programs.

Limitations

Although a comprehensive search strategy was conducted, the studies included were disproportionate across smoking, nutrition, alcohol consumption, physical activity, and obesity outcomes. Namely, there was no intervention targeting nutrition and only 4 studies targeting smoking, 3 studies targeting alcohol consumption, 3 studies targeting physical activity, and 3 studies targeting obesity. The lack of studies limits the conclusions that can be made and highlights the need for more trials assessing the effectiveness of individual, real-time video counseling that target these behaviors. Additionally, some studies that were not published in a peer-reviewed journal or not written in English were excluded, and some studies may have been missed through limitations in the searched databases [88]. Another limitation is that all included studies were conducted in high-income countries; therefore, the findings may not be generalizable to populations in developing countries or those with a diverse socioeconomic status and cultural background. Furthermore, more than half of the studies had a sample size of <100 participants [70,71,73,75,78,80,81], which may have resulted in inadequate statistical power to detect differences between groups. Additionally, in terms of methodological quality, 12 of the 13 studies had a global rating of weak, with improvements needed particularly for selection bias and blinding. Furthermore, the quality assessment for each study was based on the information reported in the publication by the authors [36], which may have had an impact on quality assessment.

This review highlights the need for more research trials examining the effectiveness of video counseling for health risks for smoking, nutrition, alcohol consumption, physical activity, and obesity. Future research should assess the effectiveness of video counseling for each health risk behavior in various populations (eg, general population, high-risk groups, and minority groups), settings (eg, health care settings, community settings, rural and remote locations), countries (eg, low- and middle-income), and cultures (eg, culturally and linguistically diverse groups, indigenous) to build upon the evidence-base and improve the generalizability of the findings. Studies examining the effectiveness of real-time video counseling for health factors of smoking, nutrition, alcohol consumption, physical activity, and obesity should consider having a larger sample size to increase the power to detect differences between groups, include populations with diverse socioeconomic and cultural backgrounds, and reduce selection bias through random selection and blind assessors and participants where possible. Future research could also examine the effectiveness of real-time video counseling for other behaviors such as sleep, health care seeking behaviors, adherence to treatments, and mental health.

Such evidence is important for informing the practices of public health prevention programs and health practitioners. Real-time video consultations have been successfully used by health practitioners for various patient-clinician consultations of long-term conditions such as heart failure, depression, schizophrenia, stroke, asthma, spinal cord injury, and chronic pain [89]. Similarly, this review suggests that health practitioners could extend the use of real-time video consultations to address the risks of smoking, nutrition, alcohol consumption, physical activity, and obesity with clients. Public health programs such as quitlines and other telephone or face-to-face services that aim to modify the risks of smoking, nutrition, alcohol consumption, physical activity, and obesity could also consider including the option for clients to choose to have support delivered via real-time video consultations. The choice to utilize real-time videoconferencing may be influenced by many factors such as client preference and funding available to providers and conditions of such funding. The use of video communication technology to provide health care services during the COVID-19 pandemic illustrates the sustainability of real-time video counseling for the risks of smoking, nutrition, alcohol consumption, physical activity, and obesity and the accessibility and reach of this intervention.

Conclusions

This review focused on effectiveness, costs, and satisfaction, factors that contribute to decision making regarding the mode by which care is delivered to clients. Policy makers and service providers also take into account other factors when making a decision about whether to integrate an intervention into their routine practices, such as feasibility, each from a provider and a client perspective. Further research is required to determine the relative benefits of video counseling in terms of these other policy and practice decision-making factors.

Acknowledgments

The School of Medicine and Public Health at the University of Newcastle, Hunter New England Population Health, and Hunter Medical Research Institute provided infrastructure support. FT was supported by a Cancer Institute New South Wales Early Career Fellowship (15/ECF/1-44) followed by a National Health & Medical Research Council Career Development Fellowship (APP1143269). The sponsors had no role in the study design; in the collection, analyses, and interpretation of data; in the writing of the manuscript; and in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of studies examining the effectiveness of video counseling on smoking cessation .

[DOCX File , 18 KB - jmir_v22i9e18621_app1.docx]

Multimedia Appendix 2
Characteristics of studies examining the effectiveness of video counseling on alcohol consumption.
[DOCX File , 17 KB - jmir_v22i9e18621_app2.docx]

Multimedia Appendix 3
Characteristics of studies examining the effectiveness of video counseling on physical activity.
[DOCX File , 17 KB - jmir_v22i9e18621_app3.docx]

Multimedia Appendix 4
Characteristics of studies examining the effectiveness of video counseling on obesity.
[DOCX File , 16 KB - jmir_v22i9e18621_app4.docx]

References

1. Noncommunicable Diseases. World Health Organisation. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2020-02-24] [WebCite Cache ID https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases]
2. Who Report on the Global Tobacco Epidemic, 2019: Offer Help to Quit Tobacco Use. World Health Organisation. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/326043/9789241516204-eng.pdf?ua=1 [accessed 2020-08-31]
3. World Health Organisation. Global Status Report on Alcohol and Health 2018. Geneva, Switzerland: World Health Organization; 2018.
4. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
5. Research W. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. World Cancer Research Fund/American Institute for Cancer Research. 2018. URL: https://www.wcrf.org/dietandcancer [accessed 2020-03-26]
6. Micha R, Shulkin ML, Peñalvo JL, Khatibzadeh S, Singh GM, Rao M, et al. Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the nutrition and chronic diseases expert group (NutriCoDE). PLoS One 2017;12(4):e0175149 [FREE Full text] [doi: 10.1371/journal.pone.0175149] [Medline: 28448503]
7. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
8. Roberts S, Spain B, Hicks C, London J, Tay S. Telemedicine in the northern territory: an assessment of patient perceptions in the preoperative anaesthetic clinic. Aust J Rural Health 2015 Jun;23(3):136-141. [doi: 10.1111/ajr.12140] [Medline: 25615954]
9. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012 Jun;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
10. Number of Internet Users Worldwide From 2005 to 2019 (in Millions). Statistica. 2019. URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ [accessed 2019-12-03]
11. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
12. Wade VA, Karnon J, Elshaug AG, Hiller JE. A systematic review of economic analyses of telehealth services using real time video communication. BMC Health Serv Res 2010 Aug 10;10:233 [FREE Full text] [doi: 10.1186/1472-6963-10-233] [Medline: 20696073]
13. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
14. Daft R, Lengel R. Information richness: a new approach to managerial behavior and organization design Research in organizational behavior. J Mark Res 1984;6:-. [doi: 10.21236/ada128980]
15. Copeland J, Martin G. Web-based interventions for substance use disorders: a qualitative review. J Subst Abuse Treat 2004 Mar;26(2):109-116. [doi: 10.1016/S0740-5472(03)00165-X] [Medline: 15050088]
16. Rochlen AB, Zack JS, Speyer C. Online therapy: review of relevant definitions, debates, and current empirical support. J Clin Psychol 2004 Mar;60(3):269-283. [doi: 10.1002/jclp.10263] [Medline: 14981791]
17. Lancaster T, Stead L. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]

18. Au LE, Whaley SE, Gurzo K, Meza M, Rosen NJ, Ritchie LD. Evaluation of online and in-person nutrition education related to salt knowledge and behaviors among special supplemental nutrition program for women, infants, and children participants. J Acad Nutr Diet 2017 Sep;117(9):1384-1395. [doi: 10.1016/j.jand.2016.12.013] [Medline: 28196620]
19. O'Hara BJ, Phongsavan P, Venugopal K, Eakin EG, Eggins D, Caterson H, et al. Effectiveness of australia's get healthy information and coaching service: translational research with population wide impact. Prev Med 2012 Oct;55(4):292-298. [doi: 10.1016/j.ypmed.2012.07.022] [Medline: 22885323]
20. Freyer-Adam J, Baumann S, Haberecht K, Tobschall S, Bischof G, John U, et al. In-person alcohol counseling versus computer-generated feedback: results from a randomized controlled trial. Health Psychol 2018 Jan;37(1):70-80. [doi: 10.1037/hea0000556] [Medline: 28967769]
21. Borland R, Segan C, Livingston P, Owen N. The effectiveness of callback counselling for smoking cessation: a randomized trial. Addiction 2001 Jun;96(6):881-889. [doi: 10.1046/j.1360-0443.2001.9668819.x] [Medline: 11399219]
22. Owen L. Impact of a telephone helpline for smokers who called during a mass media campaign. Tob Control 2000 Jun;9(2):148-154 [FREE Full text] [doi: 10.1136/tc.9.2.148] [Medline: 10841850]
23. Lindqvist H, Forsberg LG, Forsberg L, Rosendahl I, Enebrink P, Helgason AR. Motivational interviewing in an ordinary clinical setting: a controlled clinical trial at the Swedish national tobacco quitline. Addict Behav 2013 Jul;38(7):2321-2324 [FREE Full text] [doi: 10.1016/j.addbeh.2013.03.002] [Medline: 23584193]
24. Skov-Ettrup LS, Dalum P, Bech M, Tolstrup JS. The effectiveness of telephone counselling and internet- and text-message-based support for smoking cessation: results from a randomized controlled trial. Addiction 2016 Jul;111(7):1257-1266. [doi: 10.1111/add.13302] [Medline: 26748541]
25. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and Web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Dec;9(4):660-682. [doi: 10.1007/s11764-015-0442-2] [Medline: 25757733]
26. Desroches S, Lapointe A, Ratté S, Gravel K, Légaré F, Turcotte S. Interventions to enhance adherence to dietary advice for preventing and managing chronic diseases in adults. Cochrane Database Syst Rev 2013 Feb 28(2):CD008722 [FREE Full text] [doi: 10.1002/14651858.CD008722.pub2] [Medline: 23450587]
27. Stead LF, Hartmann-Boyce J, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2013 Aug 12(8):CD002850. [doi: 10.1002/14651858.CD002850.pub3] [Medline: 23934971]
28. McMahon G, Palmer S. Handbook of Counselling Psychology. New York, USA: Routledge; 2014.
29. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
30. Boland V, Stockings E, Mattick R, McRobbie H, Brown J, Courtney R. The methodological quality and effectiveness of technology-based smoking cessation interventions for disadvantaged groups: a systematic review and meta-analysis. Nicotine Tob Res 2018 Feb 7;20(3):276-285. [doi: 10.1093/ntr/ntw391] [Medline: 28034998]
31. Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. J Med Internet Res 2015 May 8;17(5):e113 [FREE Full text] [doi: 10.2196/jmir.3975] [Medline: 25956257]
32. Tzelepis F, Paul C, Williams C, Gilligan C, Regan T, Daly J, et al. Real-time video counselling for smoking cessation. Cochrane Database Syst Rev 2019 Oct 29;2019(10):1-13. [doi: 10.1002/14651858.CD012659.pub2] [Medline: 31684699]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
34. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018 Dec;125(13):1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
35. Quality Assessment Tool for Quantitative Studies Dictionary. EPHPP – McMaster Evidence Review & Synthesis Centre. 2010. URL: https://merst.ca/wp-content/uploads/2018/02/qualilty-assessment-dictionary_2017.pdf [accessed 2019-01-16]
36. Quality Assessment Tool for Quantitative Studies. National Collaborating Centre for Methods and Tools. 2017. URL: https://www.nccmt.ca/knowledge-repositories/search/14 [accessed 2019-01-16]
37. Azar KM, Aurora M, Wang EJ, Muzaffar A, Pressman A, Palaniappan LP. Virtual small groups for weight management: an innovative delivery mechanism for evidence-based lifestyle interventions among obese men. Transl Behav Med 2015 Mar;5(1):37-44 [FREE Full text] [doi: 10.1007/s13142-014-0296-6] [Medline: 25729451]
38. Davis AM, Sampilo M, Gallagher KS, Dean K, Saroja MB, Yu Q, et al. Treating rural paediatric obesity through telemedicine vs telephone: outcomes from a cluster randomized controlled trial. J Telemed Telecare 2016 Mar;22(2):86-95 [FREE Full text] [doi: 10.1177/1357633X15586642] [Medline: 26026186]
39. Davis AM, Sampilo M, Gallagher KS, Landrum Y, Malone B. Treating rural pediatric obesity through telemedicine: outcomes from a small randomized controlled trial. J Pediatr Psychol 2013 Oct;38(9):932-943 [FREE Full text] [doi: 10.1093/jpepsy/jst005] [Medline: 23428652]
40. Ehlers DK, Huberty JL, de Vreede G. Can an evidence-based book club intervention delivered via a tablet computer improve physical activity in middle-aged women? Telemed J E Health 2015 Feb;21(2):125-131. [doi: 10.1089/tmj.2013.0360] [Medline: 25526014]

41. Wild B, Hünnemeyer K, Sauer H, Hain B, Mack I, Schellberg D, et al. A 1-year videoconferencing-based psychoeducational group intervention following bariatric surgery: results of a randomized controlled study. Surg Obes Relat Dis 2015;11(6):1349-1360. [doi: 10.1016/j.soard.2015.05.018] [Medline: 26421929]
42. Nelson E, Yadrich DM, Thompson N, Wright S, Stone K, Adams N, et al. Telemedicine support groups for home parenteral nutrition users. Nutr Clin Pract 2017 Dec;32(6):789-798. [doi: 10.1177/0884533617735527] [Medline: 29016235]
43. Nolan RP, Upshur RE, Lynn H, Crichton T, Rukholm E, Stewart DE, et al. Therapeutic benefit of preventive telehealth counseling in the community outreach heart health and risk reduction trial. Am J Cardiol 2011 Mar 1;107(5):690-696. [doi: 10.1016/j.amjcard.2010.10.050] [Medline: 21215382]
44. Tsai LL, McNamara RJ, Moddel C, Alison JA, McKenzie DK, McKeough ZJ. Home-based telerehabilitation via real-time videoconferencing improves endurance exercise capacity in patients with COPD: the randomized controlled TeleR study. Respirology 2017 May;22(4):699-707 [FREE Full text] [doi: 10.1111/resp.12966] [Medline: 27992099]
45. Huennemeyer K, Teufel M, Hain B, Rieber N, Sauer H, Rudofsky G. The Base for a Balanced Life - The Randomized Controlled Base (Bariatric Surgery and Education) Study. In: 16th Congress of the International Federation for the Surgery. 2011 Presented at: IFS'11; August 31-September 3, 2011; Hamburg, Germany.
46. Wu G, Keyes L, Callas P, Ren X, Bookchin B. Comparison of telecommunication, community, and home-based Tai Chi exercise programs on compliance and effectiveness in elders at risk for falls. Arch Phys Med Rehabil 2010 Jun;91(6):849-856. [doi: 10.1016/j.apmr.2010.01.024] [Medline: 20510973]
47. Hwang R, Morris NR, Mandrusiak A, Bruning J, Peters R, Korczyk D, et al. Cost-utility analysis of home-based telerehabilitation compared with centre-based rehabilitation in patients with heart failure. Heart Lung Circ 2019 Dec;28(12):1795-1803. [doi: 10.1016/j.hlc.2018.11.010] [Medline: 30528811]
48. Hwang R, Bruning J, Morris NR, Mandrusiak A, Russell T. Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. J Physiother 2017 Apr;63(2):101-107 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.017] [Medline: 28336297]
49. Porter LS, Gao X, Lyna P, Kraus W, Olsen M, Patterson E, et al. Pilot randomized trial of a couple-based physical activity videoconference intervention for sedentary cancer survivors. Health Psychol 2018 Sep;37(9):861-865. [doi: 10.1037/hea0000608] [Medline: 30138021]
50. Clark DO, Keith N, Weiner M, Xu H. Outcomes of an RCT of videoconference vs in-person or in-clinic nutrition and exercise in midlife adults with obesity. Obes Sci Pract 2019 Apr;5(2):111-119 [FREE Full text] [doi: 10.1002/osp4.318] [Medline: 31019728]
51. West DS, Stansbury M, Krukowski RA, Harvey J. Enhancing group-based internet obesity treatment: a pilot RCT comparing video and text-based chat. Obes Sci Pract 2019 Dec;5(6):513-520 [FREE Full text] [doi: 10.1002/osp4.371] [Medline: 31890241]
52. Rosas LG, Lv N, Azar KM, Xiao L, Hooker SP, Lewis MA, et al. Hombre: a randomized controlled trial to compare two approaches to weight loss for overweight and obese latino men (hombres con opciones para mejorar el bienestar y bajar el riesgo de enfermedades crónicas; men with choices to improve well-being and decrease chronic disease risk). Contemp Clin Trials 2018 May;68:23-34. [doi: 10.1016/j.cct.2018.02.019] [Medline: 29505867]
53. Mascarenhas MN, Chan JM, Vittinghoff E, van Blarigan EL, Hecht F. Increasing physical activity in mothers using video exercise groups and exercise mobile apps: randomized controlled trial. J Med Internet Res 2018 May 18;20(5):e179 [FREE Full text] [doi: 10.2196/jmir.9310] [Medline: 29776899]
54. Chai LK, Collins CE, May C, Ashman A, Holder C, Brown LJ, et al. Feasibility and efficacy of a web-based family telehealth nutrition intervention to improve child weight status and dietary intake: a pilot randomised controlled trial. J Telemed Telecare 2019 Jul 31:- epub ahead of print. [doi: 10.1177/1357633X19865855] [Medline: 31364474]
55. Lynch CP, Williams JS, Ruggiero KJ, Knapp RG, Egede LE. Tablet-aided behavioral intervention effect on self-management skills (TABLETS) for diabetes. Trials 2016 Mar 22;17:157 [FREE Full text] [doi: 10.1186/s13063-016-1243-2] [Medline: 27005766]
56. Zanaboni P, Dinesen B, Hjalmarsen A, Hoaas H, Holland AE, Oliveira CC, et al. Long-term integrated telerehabilitation of COPD patients: a multicentre randomised controlled trial (iTrain). BMC Pulm Med 2016 Aug 22;16(1):126 [FREE Full text] [doi: 10.1186/s12890-016-0288-z] [Medline: 27549782]
57. Cox NS, McDonald CF, Alison JA, Mahal A, Wootton R, Hill CJ, et al. Telerehabilitation versus traditional centre-based pulmonary rehabilitation for people with chronic respiratory disease: protocol for a randomised controlled trial. BMC Pulm Med 2018 May 15;18(1):71 [FREE Full text] [doi: 10.1186/s12890-018-0646-0] [Medline: 29764393]
58. Chen AC, Rosenthal DN, Couch SC, Berry S, Stauffer KJ, Brabender J, et al. Healthy hearts in pediatric heart transplant patients with an exercise and diet intervention via live video conferencing-design and rationale. Pediatr Transplant 2019 Feb;23(1):e13316. [doi: 10.1111/petr.13316] [Medline: 30393915]
59. Vincze L, Rollo ME, Hutchesson MJ, Callister R, Collins CE. VITAL change for mums: a feasibility study investigating tailored nutrition and exercise care delivered by video-consultations for women 3-12 months postpartum. J Hum Nutr Diet 2018 Jun;31(3):337-348. [doi: 10.1111/jhn.12549] [Medline: 29543356]

60. Marhefka SL, Turner D, Lockhart E, Rivara A, Wang W, Shuter J. Meeting our patients 'where they are': video-group smoking cessation for people living with HIV. J Assoc Nurses AIDS Care 2018;29(2):338-344 [FREE Full text] [doi: 10.1016/j.jana.2017.09.007] [Medline: 29033093]
61. Coles N, Patel BP, Li P, Cordeiro K, Steinberg A, Zdravkovic A, et al. Breaking barriers: adjunctive use of the Ontario telemedicine network (OTN) to reach adolescents with obesity living in remote locations. J Telemed Telecare 2020 Jun;26(5):271-277. [doi: 10.1177/1357633X18816254] [Medline: 30526258]
62. Onezi HA, Khalifa M, El-Metwally A, Househ M. The impact of social media-based support groups on smoking relapse prevention in Saudi Arabia. Comput Methods Programs Biomed 2018 Jun;159:135-143. [doi: 10.1016/j.cmpb.2018.03.005] [Medline: 29650308]
63. Fleischman A, Hourigan SE, Lyon HN, Landry MG, Reynolds J, Steltz SK, et al. Creating an integrated care model for childhood obesity: a randomized pilot study utilizing telehealth in a community primary care setting. Clin Obes 2016 Dec;6(6):380-388 [FREE Full text] [doi: 10.1111/cob.12166] [Medline: 27863024]
64. van Straaten MG, Cloud BA, Morrow MM, Ludewig PM, Zhao KD. Effectiveness of home exercise on pain, function, and strength of manual wheelchair users with spinal cord injury: a high-dose shoulder program with telerehabilitation. Arch Phys Med Rehabil 2014 Oct;95(10):1810-7.e2 [FREE Full text] [doi: 10.1016/j.apmr.2014.05.004] [Medline: 24887534]
65. Frueh BC, Henderson S, Myrick H. Telehealth service delivery for persons with alcoholism. J Telemed Telecare 2005;11(7):372-375. [doi: 10.1258/135763305774472060] [Medline: 16238840]
66. Byaruhanga J, Tzelepis F, Paul C, Wiggers J, Byrnes E, Bowman J. The short-term effectiveness of real-time video counseling on smoking cessation among smokers residing in regional and remote areas. J Clin Oncol 2018;2017(5):-. [doi: 10.1002/14651858.CD012659]
67. Parameswaran V, Josan K, Winterbottom J, Shearer J, Khandelwal A, Nallamshetty S, et al. Abstract P337: cardioclick an innovative telehealth approach to lifestyle intervention in high risk south Asians. Circulation 2019 Mar 05;139(Suppl_1):-. [doi: 10.1161/circ.139.suppl_1.p337]
68. Alencar MK, Johnson K, Mullur R, Gray V, Gutierrez E, Korosteleva O. The efficacy of a telemedicine-based weight loss program with video conference health coaching support. J Telemed Telecare 2019 Apr;25(3):151-157. [doi: 10.1177/1357633X17745471] [Medline: 29199544]
69. Alencar M, Johnson K, Gray V, Mullur R, Gutierrez E, Dionico P. Telehealth-based health coaching increases m-health device adherence and rate of weight loss in obese participants. Telemed J E Health 2020 Mar;26(3):365-368. [doi: 10.1089/tmj.2019.0017] [Medline: 30994410]
70. Kim S, Sitthisongkram S, Bernstein K, Fang H, Choi W, Ziedonis D. A randomized controlled trial of a videoconferencing smoking cessation intervention for Korean American women: preliminary findings. Int J Womens Health 2016;8:453-462 [FREE Full text] [doi: 10.2147/IJWH.S109819] [Medline: 27660494]
71. Kim S, Darwish S, Lee SA, Sprague C, DeMarco RF. A randomized controlled pilot trial of a smoking cessation intervention for US women living with HIV: telephone-based video call vs voice call. Int J Womens Health 2018;10:545-555 [FREE Full text] [doi: 10.2147/IJWH.S172669] [Medline: 30288127]
72. Nomura A, Tanigawa T, Muto T, Oga T, Fukushima Y, Kiyosue A, et al. Clinical efficacy of telemedicine compared to face-to-face clinic visits for smoking cessation: multicenter open-label randomized controlled noninferiority trial. J Med Internet Res 2019 Apr 26;21(4):e13520 [FREE Full text] [doi: 10.2196/13520] [Medline: 30982776]
73. Tarp K, Bojesen AB, Mejldal A, Nielsen AS. Effectiveness of optional videoconferencing-based treatment of alcohol use disorders: randomized controlled trial. JMIR Ment Health 2017 Sep 29;4(3):e38 [FREE Full text] [doi: 10.2196/mental.6713] [Medline: 28963093]
74. Staton-Tindall M, Havens JR, Webster JM, Leukefeld C. METelemedicine: a pilot study with rural alcohol users on community supervision. J Rural Health 2014;30(4):422-432 [FREE Full text] [doi: 10.1111/jrh.12076] [Medline: 24800699]
75. King SC, Richner KA, Tuliao AP, Kennedy JL, McChargue DE. A comparison between telehealth and face-to-face delivery of a brief alcohol intervention for college students. Subst Abus 2019 Oct 23:1-9 epub ahead of print. [doi: 10.1080/08897077.2019.1675116] [Medline: 31644389]
76. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. Web-based video-coaching to assist an automated computer-tailored physical activity intervention for inactive adults: a randomized controlled trial. J Med Internet Res 2016 Aug 12;18(8):e223 [FREE Full text] [doi: 10.2196/jmir.5664] [Medline: 27520283]
77. Weinstock RS, Brooks G, Palmas W, Morin PC, Teresi JA, Eimicke JP, et al. Lessened decline in physical activity and impairment of older adults with diabetes with telemedicine and pedometer use: results from the IDEATel study. Age Ageing 2011 Jan;40(1):98-105. [doi: 10.1093/ageing/afq147] [Medline: 21081539]
78. Chemtob K, Rocchi M, Arbour-Nicitopoulos K, Kairy D, Fillion B, Sweet S. Using tele-health to enhance motivation, leisure time physical activity, and quality of life in adults with spinal cord injury: A self-determination theory-based pilot randomized control trial. Psychol Sport Exerc 2019 Jul;43:243-252. [doi: 10.1016/j.psychsport.2019.03.008]
79. Hansen CR, Perrild H, Koefoed BG, Zander M. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017 Jun;176(6):727-736. [doi: 10.1530/EJE-16-0811] [Medline: 28325823]

80. Homma S, Imamura H, Nakamura T, Fujimura K, Ito Y, Maeda Y, et al. A comparative study on the effectiveness of one-way printed communication versus videophone interactive interviews on health promotion. J Telemed Telecare 2016 Jan;22(1):56-63. [doi: 10.1177/1357633X15587436] [Medline: 26026183]
81. Johnson KE, Alencar MK, Coakley KE, Swift DL, Cole NH, Mermier CM, et al. Telemedicine-based health coaching is effective for inducing weight loss and improving metabolic markers. Telemed J E Health 2019 Feb;25(2):85-92 [FREE Full text] [doi: 10.1089/tmj.2018.0002] [Medline: 29847222]
82. Cunningham JA, Kypri K, McCambridge J. The use of emerging technologies in alcohol treatment. Alcohol Res Health 2011;33(4):320-326 [FREE Full text] [Medline: 23580017]
83. Fleming M, Manwell LB. Brief intervention in primary care settings. A primary treatment method for at-risk, problem, and dependent drinkers. Alcohol Res Health 1999;23(2):128-137 [FREE Full text] [Medline: 10890807]
84. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol 2004 Jul;57(6):543-550. [doi: 10.1016/j.jclinepi.2003.12.010] [Medline: 15246122]
85. Greenhalgh E, Stillman S. Cessation Assistance: Telephone- and Internet-based Interventions. Tobacco in Australia. 2020. URL: https://www.tobaccoinaustralia.org.au/chapter-7-cessation/7-14-methods-services-and-products-for-quitting-te [accessed 2020-09-02]
86. Ossip-Klein DJ, McIntosh S. Quitlines in North America: evidence base and applications. Am J Med Sci 2003 Oct;326(4):201-205. [doi: 10.1097/00000441-200310000-00010] [Medline: 14557735]
87. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
88. Sterne J, Egger M, Moher D. Addressing Reporting Biases. Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1. Chichester, UK: Cochrane Collaboration; 2011.
89. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]

Abbreviations

COM-B: capability, opportunity, motivation, and behavior
DALY: disability-adjusted life year
LTPA: leisure time physical activity
MET: metabolic equivalent
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews
RAPI: Rutgers Alcohol Problem Index
SCI: spinal cord injury

Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by M allman-Farinelli, A Nomura, S Lippke, T Muto; comments to author 12.06.20; revised version received 18.07.20; accepted 26.07.20; published 11.09.20.

Please cite as:
Byaruhanga J, Atorkey P, McLaughlin M, Brown A, Byrnes E, Paul C, Wiggers J, Tzelepis F
Effectiveness of Individual Real-Time Video Counseling on Smoking, Nutrition, Alcohol, Physical Activity, and Obesity Health Risks: Systematic Review
J Med Internet Res 2020;22(9):e18621
URL: http://www.jmir.org/2020/9/e18621/
doi:10.2196/18621
PMID:32915156

Review

Identification of Patient Perceptions That Can Affect the Uptake of Interventions Using Biometric Monitoring Devices: Systematic Review of Randomized Controlled Trials

Alexander Perlmutter[1,2], MPhil, MPH; Mehdi Benchoufi[3], MD; Philippe Ravaud[3], MD, PhD; Viet-Thi Tran[3], MD, PhD

[1]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, United States

[2]School of Global Public Health, New York University, New York, NY, United States

[3]UMR1153 (METHODS team), Centre de Recherche en Epidemiologie et StatistiqueS, Institut national de la santé et de la recherche médicale, Paris, France

Corresponding Author:
Alexander Perlmutter, MPhil, MPH
Department of Epidemiology
Mailman School of Public Health
Columbia University
Room 720.10
722 W 168th St
New York, NY, 10032
United States
Phone: 1 703 336 9067
Email: asp2183@cumc.columbia.edu

Abstract

Background: Biometric monitoring devices (BMDs) are wearable or environmental trackers and devices with embedded sensors that can remotely collect high-frequency objective data on patients' physiological, biological, behavioral, and environmental contexts (for example, fitness trackers with accelerometer). The real-world effectiveness of interventions using biometric monitoring devices depends on patients' perceptions of these interventions.

Objective: We aimed to systematically review whether and how recent randomized controlled trials (RCTs) evaluating interventions using BMDs assessed patients' perceptions toward the intervention.

Methods: We systematically searched PubMed (MEDLINE) from January 1, 2017, to December 31, 2018, for RCTs evaluating interventions using BMDs. Two independent investigators extracted the following information: (1) whether the RCT collected information on patient perceptions toward the intervention using BMDs and (2) if so, what precisely was collected, based on items from questionnaires used and/or themes and subthemes identified from qualitative assessments. The two investigators then synthesized their findings in a schema of patient perceptions of interventions using BMDs.

Results: A total of 58 RCTs including 10,071 participants were included in the review (the median number of randomized participants was 60, IQR 37-133). BMDs used in interventions were accelerometers/pedometers (n=35, 60%), electrochemical biosensors (eg, continuous glucose monitoring; n=18, 31%), or ecological momentary assessment devices (eg, carbon monoxide monitors for smoking cessation; n=5, 9%). Overall, 26 (45%) trials collected information on patient perceptions toward the intervention using BMDs and allowed the identification of 76 unique aspects of patient perceptions that could affect the uptake of these interventions (eg, relevance of the information provided, alarm burden, privacy and data handling, impact on health outcomes, independence, interference with daily life). Patient perceptions were unevenly collected in trials. For example, only 5% (n=3) of trials assessed how patients felt about privacy and data handling aspects of the intervention using BMDs.

Conclusions: Our review showed that less than half of RCTs evaluating interventions using BMDs assessed patients' perceptions toward interventions using BMDs. Trials that did assess perceptions often only assessed a fraction of them. This limits the extrapolation of the results of these RCTs to the real world. We thus provide a comprehensive schema of aspects of patient perceptions that may affect the uptake of interventions using BMDs and which should be considered in future trials.

Trial Registration: PROSPERO CRD42018115522; https://tinyurl.com/y5h8fjgx

(J Med Internet Res 2020;22(9):e18986) doi:10.2196/18986

KEYWORDS

systematic review; patient perceptions; biometric monitoring device; randomized controlled trials; accelerometer; pedometer; ecological momentary assessment; electrochemical biosensor; adoption; uptake; real-world

Introduction

Biometric monitoring devices (BMDs) are wearable or environmental trackers and devices with embedded sensors that can remotely collect high-frequency objective data on patients' physiological, biological, behavioral, and environmental contexts [1]. In recent years, there has been a surge of therapeutic interventions using BMDs to monitor patients' health and treatment response to reactively adjust patients' care "just in time" [1-7]. The development of these innovative interventions using BMDs has raised great interest from governments, payers, care providers, and patients given their potential to transform the delivery of care from intermittent clinical visits with clinicians to remote and continuous management of patients, at scale, in real time [2,7-10].

Despite promising results, the real-world effectiveness of interventions using BMDs depends on patients' uptake, engagement, and adherence to these interventions [11]. For example, there is evidence of low patient engagement in the first large-scale implementations of digital monitoring strategies (eg, 90% incomplete follow-up for MyHeart Counts; 55% incomplete follow-up data for the Healthy Pregnancy Research Program) [12,13].

The literature on reasons explaining the poor uptake of these interventions, specifically on patients' perceptions that can affect the uptake of interventions using BMDs is limited to the following: (1) small-sized pilot studies with short follow-ups [14-16], (2) surveys that explore stated preferences from patients [11,17], and (3) more rarely, objective assessment of patients' perceptions toward these interventions in the clinical trials evaluating them (eg, via questionnaires). As a result, it is still unclear which specific patient perceptions should be measured in the clinical trials evaluating interventions using BMDs to inform their potential uptake. Such knowledge would strengthen inference about the potential external validity of results and benefit the planning of future trials.

In this study, we aimed to systematically review recent RCTs evaluating interventions using BMDs to understand whether and how patients' perceptions toward these technologies were considered.

Methods

We uploaded a prespecified protocol in November 2018 on PROSPERO (The International Prospective Register of Systematic Reviews; CRD42018115522). We followed standard procedures for systematic reviews and reported processes and results according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [18].

Data Sources and Searches

We systematically searched PubMed for eligible studies published in MEDLINE between January 1, 2017, and December 31, 2018. These eligibility dates were chosen to provide a sample of recent trials reflecting the current state of science on interventions using BMDs. The search equation had no language restrictions and was derived from the Cochrane Highly Sensitive Search Strategy with a filter for randomized controlled trials Medical Subject Heading (MeSH) terms, and free-text words pertaining to digital, mobile, and electronic health keywords identified during a pilot phase (Multimedia Appendix 1).

Study Selection

We included published primary reports of RCTs in humans that assessed the efficacy of an intervention using BMDs (ie, interventions using wearables, trackers, or sensors/biosensors—for combined home and mobile use—that have the capability to collect and transmit data for the purposes of improving a patient's health or preventing disease onset) [9,19,20]. When an intervention involved multiple components including some not related to BMDs, we focused on the component(s) involving BMDs. We excluded interventions utilizing telemedicine/telehealth (eg, videoconferencing), SMS text messages sent to mobile phones, and exclusively smartphone apps [21]. We excluded protocols, observational studies, and reviews. We also excluded publications evaluating interventions that were confined to a doctor's office (eg, virtual reality headset intervention for the treatment of social anxiety disorder), and publications on interventions targeting clinicians rather than patients.

One investigator (AP) screened titles and abstracts for irrelevant publications. AP confirmed the eligibility of all screened-in studies based on the articles' full-text and the reasons for not meeting eligibility.

Data Extraction

One investigator (AP) used a standardized form to extract from the articles (and supplementary material and referenced sources if necessary) the general characteristics of trials (authors, title, journal, publication year, number of participants randomized, technology being assessed). When possible, we also reviewed the trial's entry in a public clinical trial registry (eg, ClinicalTrials.gov) using information available in published articles. We assessed whether some outcomes measuring patients' perceptions could be registered but not reported in published articles.

Two investigators (AP, MB) used a standardized form to independently extract data on how patients' perceptions toward interventions using BMDs were assessed. These data included the following: (1) whether the trial collected information on patients' perceptions toward the interventions using BMDs, (2)

whether the information collected was a study outcome (primary or secondary), (3) how this information was measured (eg, using questionnaires, interviews, focus groups, or a combination of these), and (4) which patient perceptions were collected. This latter extraction was based on the review of all items from questionnaires used to assess patients' perceptions toward interventions using BMDs, and/or themes and subthemes from qualitative assessments (ie, interviews and focus groups), if available. All items, themes and subthemes extracted were then compiled into a comprehensive list of patient perceptions toward the BMDs that were assessed in the included trials. Therefore, the list provided information on patient perceptions toward interventions using BMDs that may affect their uptake from both researchers' (from the standardized questionnaires used in the RCTs) and patients' perspectives (eg, from the qualitative assessments obtained in the RCTs).

Data Synthesis and Analysis

General Characteristics of RCTs

We summarized the characteristics of included trials with frequencies (proportions) for categorical variables and medians and interquartile ranges (IQR) for continuous variables.

Schema of Patients' Perceptions That Could Affect the Uptake of Interventions Using BMDs

Two investigators (AP and MB) independently organized the list of patient perceptions toward interventions using BMDs by critically examining the wording of the extracted content and context. First, they excluded general assessments (eg, whether the device was acceptable or helpful, in general) and restricted the list to specific patient perceptions toward interventions using BMDs that could affect the uptake of interventions. Second, they grouped similar patient perceptions (eg, "easy to use" and "I thought this system was easy to use" were grouped together as "easy to use"). Finally, they organized these perceptions into a schema of specific aspects of patient perceptions. Disagreements were collaboratively settled with a third investigator (VTT).

How RCTs and Validated Scales Cover the Schema of Patient Perceptions

We investigated how the trials included in this review covered the schema of patients' perceptions toward interventions using BMDs by mapping the specific aspects of patients' perceptions measured in each trial to the overarching categories and subcategories of the schema.

Similarly, we investigated how comprehensively the validated scales used in the included trials covered the schema by mapping specific aspects of patients' perceptions from each validated questionnaire to the categories and subcategories of the schema.

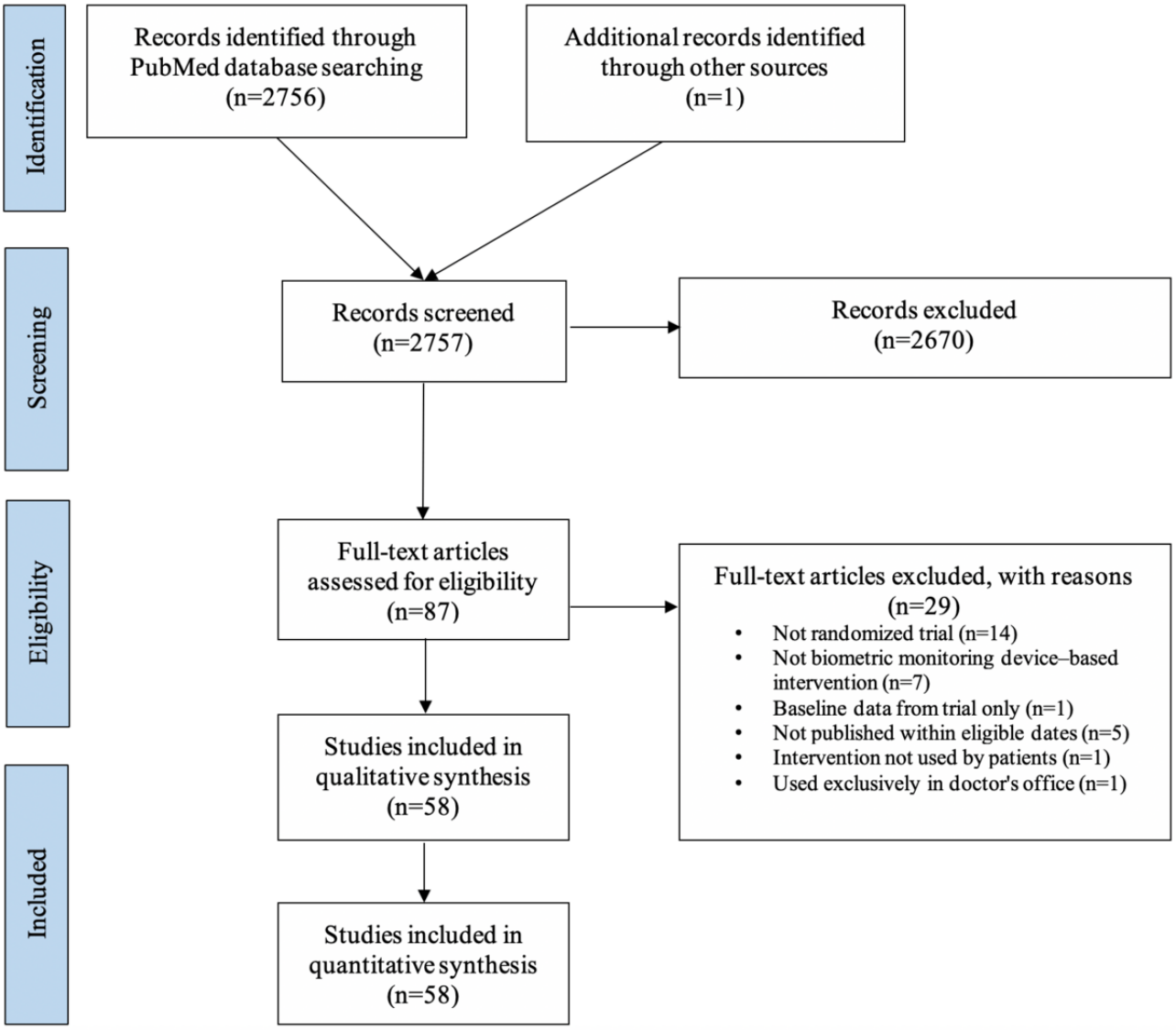
Members of the public were not involved in the design of this systematic review or the interpretation of the results.

Results

General Characteristics of RCTs

In total, 58 RCTs that randomized 10,071 participants were included in the review (Figure 1, Multimedia Appendix 2). RCTs randomized a median of 60 participants (IQR 37-133). Trials involved patients with diabetes (n=12, 21%), cancer (n=5, 9%), or healthy (or at-risk) primary patients (n=15, 26%). Trials were mostly funded by nonprofit sources (n=40, 69%); there were 5 (9%) trials that did not report their funding source. Most trials were single-center (n=54, 93%) and tested a commercialized technology (n=47, 81%).

Figure 1. PRISMA flowchart. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



BMDs used in interventions were mainly accelerometers/pedometers (eg, Fitbit; n=35, 60%), electrochemical biosensors (eg, continuous glucose monitoring devices; n=18, 31%), or ecological momentary assessment devices (n=5, 9%) that were either worn (eg, blood pressure monitor) or unworn (eg, carbon monoxide measurement monitors designed for smoking cessation; Table 1).

In total, 28 (48%) and 26 (45%) of the 58 included RCTs discussed and collected information on patients' perceptions about the intervention using BMDs, respectively. Overall, 20 (34%) trials explicitly stated that the collected perceptions were trial outcomes. All 26 trials that collected perceptions reported how they were collected (eg, questionnaire, interview, focus group): 18 (31%) trials used a questionnaire, with 5 (9%) reporting that they used a validated instrument (Table 2).

Table 1. Characteristics of the 58 included trials (N=58)[a].

Characteristic	Trials
Number of patients randomized, median (IQR)	60 (37-133)
Type of biometric monitoring device[b], n (%)	
Accelerometer/pedometer	35 (60)
Electrochemical biosensor	18 (31)
Ecological momentary assessment/attachable	5 (9)
Therapeutic area, n (%)	
Diabetes	12 (21)
Improving physical activity (primary prevention)	12 (21)
Improving diet (primary prevention)	3 (5)
Cardiovascular diseases (including stroke)	10 (17)
Cancer	5 (9)
Rheumatologic diseases	5 (9)
Smoking/alcohol cessation	3 (5)
Respiratory diseases	3 (5)
Weight management	2 (3)
Neurological diseases	2 (3)
Gastrointestinal diseases	1 (2)
Single or multicenter, n (%)	
Single center trial	54 (93)
Multicenter trial	4 (7)
Use of a commercial biometric monitoring device , n (%)	
Yes	47 (81)
No	10 (17)
Unknown	1 (2)
Funding, n (%)	
Nonprofit (government, university, nonprofit nongovernmental organization)	40 (69)
For-profit (pharmaceutical industries)	8 (14)
Mixed	5 (9)
Not reported	5 (9)

[a]Percentages may not equal 100% due to rounding.

[b]Many of these biometric monitoring devices were used in addition to a smartphone application.

Table 2. Collection, discussion, and reporting of patient perceptions toward biometric monitoring devices in the 58 included trials.

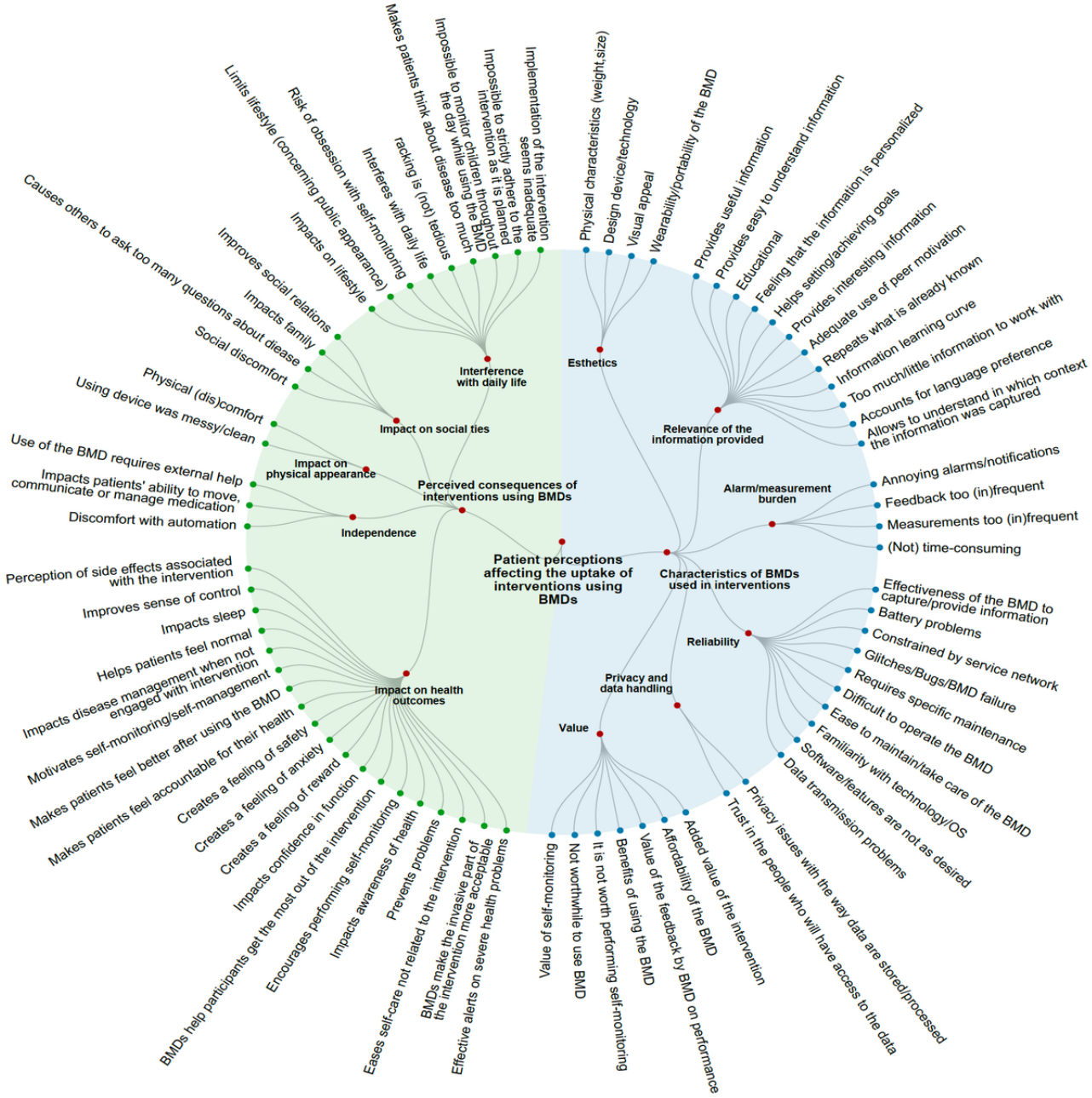
Collection, discussion, and reporting of patient perceptions	Studies, n (%)
Discussed at least one patient perception	28 (48)
Collected at least one patient perception	26 (45)
With a questionnaire	18 (31)
With face-to-face interviews	2 (3)
With focus group	1 (2)
By combining multiple collection modalities	5 (9)
Patient perception was reported as a trial outcome	20 (34)

Schema of Patients' Perceptions That Could Affect the Uptake of Interventions Using BMDs

Among the 26 trials that collected patients' perceptions toward the intervention using BMDs, 23 (39%) evaluated specific patient perceptions of the intervention that could affect the uptake (ie, 3 collected only general satisfaction with or acceptability of the BMD).

We identified 76 unique specific aspects of patients' perceptions toward interventions using BMDs that could affect their uptake. These aspects of perceptions were grouped into two overarching categories: (1) patient perceptions toward characteristics of BMDs used in interventions (n=39, 51%) and (2) perceived consequences of interventions using BMDs (n=37, 49%; Figure 2).

Figure 2. Schema of the 76 specific patient perceptions that could affect the uptake of interventions using BMDs. Specific perceptions are represented by blue nodes around the figure and organized in subcategories (outer red nodes) and major categories (inner red nodes dividing the circle into green and blue sections). BMD: biometric monitoring device.



Patients' Perceptions Toward Characteristics of BMDs Used in Interventions

Patients' perceptions toward the characteristics of BMDs were related to the following:

1. Esthetics, which describes the look and feel of the BMD. For instance, in one trial, patients were asked to assess whether the BMD was attractive and visually appealing. A total of 6 (10%) trials measured this patient perception [22].
2. Relevance of the information provided, which describes how well the patient feels he or she can interact with or use the information that the BMD delivers. For instance, in one trial, participants reported that potential further development of the BMD could include "more interesting content" on a

web-based mobile service related to their use of a wrist-worn physical activity monitor [23]. A total of 15 (26%) trials measured this patient perception.

3. Alarm/measurement burden, which describes patients' views about the BMDs' features, such as alarm frequency or how frequently a measurement occurs. For instance, one trial asked patients to rate how much they agreed with the statement "Alarms too often for no good reason" [24]. A total of 6 (10%) trials measured this patient perception.
4. Reliability, which describes whether patients feel that the BMD used in the intervention can function properly (battery, connectivity, maintenance). For instance, one trial reported that "two participants discontinued using the Fitbit…because of battery problems" [25]. A total of 7 (12%) trials measured this patient perception.
5. Privacy and data handling, which describes how much patients feel that their privacy is protected and how accountable the people/organizations with whom their data is shared will use it for genuinely medical reasons. For instance, one trial asked patients how much they agreed with the statement "My privacy was protected when I used the system" [26]. A total of 3 (5%) trials measured this patient perception.
6. Value, which describes what patients can accept to forgo in terms of time or money for the intervention. For instance, one trial asked how much patients agreed with the statement "The effort of using this technology/method is worthwhile for me" [27]. A total of 10 (17%) trials measured this patient perception.

Perceived Consequences of Interventions Using BMDs

Patients' perceptions related to the potential consequences of the interventions using BMDs involved the following:

1. Perceived impact on health outcomes, which describes how the intervention may impact the patients' health, disease, or response to treatment. For instance, one trial asked patients how much they agreed with the statement "Has helped to control diabetes better even when not wearing it" [24]. A total of 12 (21%) trials measured this patient perception.
2. Independence, which describes how the BMD may impact patients' dependence on others or automation to conduct tasks. For instance, one trial asked participants how much they agreed with the statement "I felt that I needed someone's help to be able to use the system" [26]. A total of 9 (16%) trials measured this patient perception.
3. Perceived impact on their physical appearance, which describes patients' views about how the BMD can impact their appearance or make them feel (physically). For instance, a questionnaire in one trial asked patients, "How physically uncomfortable was wearing the bracelet?" [28]. A total of 11 (19%) trials measured this patient perception.
4. Social ties, which describes how patients feel the intervention using the BMD makes them engage with other people and vice versa. For instance, one trial asked participants how much they agreed with the statement "Has caused more family arguments" [24]. A total of 6 (10%) trials measured this patient perception.
5. Interference, which describes how the intervention using the BMD interferes with daily life or alleviates daily stressors, and how patients feel about modifying their lifestyle to use the BMD. For instance, one trial asked patients how much they agreed with the statement "Causes too many hassles in daily life" [29]. A total of 12 (21%) trials measured this patient perception.

How RCTs and Validated Scales Cover the Schema of Patient Perceptions

Of the 23 trials that collected at least one specific aspect of a patient perception, 18 (78%) covered both perceptions toward characteristics of BMDs and perceptions of potential consequences of the intervention. Trials covered a median of 4 of the schema's 11 subcategories (IQR 3-6, maximum 9). Furthermore, 8 of the trials covered 5 or more of the subcategories (Figure 3).

Figure 3. Patients' perceptions toward interventions using BMDs collected in the included trials (n=58). All RCTs included in the current systematic review are shown around the figure by the first author's last name. RCTs that collected at least one specific patient perception toward BMDs are shown in light green shading (category indicating patients' perceptions toward characteristics of BMDs used in interventions) or beige shading (category indicating patients' perceptions of consequences of interventions using BMDs). Gray shading corresponds to RCTs not collecting a specific patient perception toward the intervention using BMDs. Colored nodes in the interior of the figure correspond to subcategories of patient perceptions toward interventions using BMDs according to the schema in Figure 2. BMD: biometric monitoring device; RCT: randomized controlled trial.

In the included trials, we identified four validated scales to measure patient perceptions toward the intervention using a BMD:

1. The 44-item Continuous Glucose Monitoring Satisfaction Questionnaire, used in two trials [24,29], covered 9/39 perceptions toward characteristics of BMDs and 17/37 perceptions of potential consequences of interventions using BMDs.
2. The 29-item Tele-healthcare Satisfaction Questionnaire used in one trial [27] covered 6/39 perceptions toward characteristics of BMDs and 5/37 perceptions of potential consequences of interventions using BMDs.
3. The 16-item Marshfield Usability Survey used in one trial [26] covered 4/39 perceptions toward characteristics of BMDs and 2/37 perceptions of potential consequences of interventions using BMDs.
4. The 17-item questionnaire adapted from Vandelanotte et al [30] used in one trial [31] covered 5/39 perceptions toward characteristics of BMDs and 3/37 perceptions of potential consequences of interventions using BMDs.

Discussion

In this systematic review, we assessed how patients' perceptions toward interventions using BMDs were assessed in recent RCTs.

Our results highlight that less than half of trials collected patients' perceptions toward the intervention. Among trials that did, most only partially covered the potential patient perceptions that could affect the uptake of interventions using BMDs. For example, only 5% of included trials assessed how patients felt with the privacy and data handling aspects of the intervention using BMDs. As a result, this creates an information gap regarding the potential uptake and implementation of these interventions [32,33].

Further, our work enabled the identification of a comprehensive list of 76 specific aspects of patients' perceptions toward interventions using BMDs that could affect their uptake, coming from both investigators' insights (through the analysis of the questionnaires used in the trials) and patients' perspectives (through the inclusion of results from qualitative inquiries collected during trials). Our findings may help researchers developing new interventions using BMDs consider and address all aspects that could impact the uptake of their interventions.

To our knowledge, this is the first study to provide a comprehensive schema of patients' perceptions toward interventions using BMDs. Our findings fit the empirical examples [34-37] of theoretical models about patient perspectives' relationships with technology adoption [38,39] in that patients express views concerning ease of use, lack of privacy, enjoyment, motivation, and social influence. Our work is also more nuanced, emphasizing patients' views about device affordability, reliability, relevance of information and content, value, and interference (with daily life), among many others.

Our first major result is that less than half of the trials in this review collected patient perceptions toward interventions using BMDs. These patients' views are crucial to knowing whether the interventions would function in real-world settings and measuring them is the only way to get an insight into the potential uptake of these interventions in the real world [32,33]. In particular, we advocate against equating retention in trials with BMD adoption because retention is affected by the context of research.

Our second major result is that the patients' perceptions toward interventions using BMDs collected in trials are numerous. Our results highlight that no scale used to measure patients' perceptions toward interventions using BMDs offered a comprehensive assessment of the potential uptake of the interventions. Our schema of patient perceptions provides an empirical framework for helping guide implementation of the results of trials using BMDs, with the ultimate goal of wide-scale adoption in real-world settings. For example, it may serve the development of a new measurement tool for future trials.

Our findings complement the existing literature exploring the factors that may affect the uptake of BMDs and interventions using BMDs in health care, which was mainly composed of the following: (1) small-scale qualitative studies and theoretical models of technology adoption, (2) small-scale pilot studies testing the BMDs in controlled environments, and (3) surveys exploring stated preferences from patients. Individually, these studies did not capture the abundance and context of patients' views toward BMDs and their adoption. For instance, theoretical models of technology adoption were not necessarily health care–specific. Pilot studies of interventions using BMDs often have short follow-up periods, and views expressed about BMDs may not be generalizable because of the limited sample size [14-16]. Qualitative studies or surveys explore stated preferences from patients [11,17] and often explore the general perceptions of people rather than their experience with specific BMDs in their own daily lives. Finally, there are some clinical trials that were included in our review in which patients' perceptions toward BMDs were evaluated. However, unlike individual studies, this review organized the patients' perceptions from all trials into a single schema. To our knowledge, our results present the most comprehensive assessment of patients' perceptions toward BMDs that exists, which will help investigators and sponsors refine interventions to improve patients' uptake and engagement.

Our study has some limitations. First, our inclusion criteria limited this review to RCTs (ie, preliminary observational pilot studies were excluded). However, we argue that these pilot studies evaluating new interventions using BMDs do not usually include many participants. Second, the schema we created is one of ostensibly multiple schemas that could have been created. Even though the systematically executed extraction would aid other investigators' attempts to reproduce our findings, other investigators could create a different schema than ours, based on their experience. Third, we may have missed trials using BMDs by virtue of these devices being novel. Medical Subject Heading (MeSH) terms may not have been assigned yet or the assigned MeSH terms may not have included the ones from our search. Although our review was probably missing some trials, this would not have changed our main results that a large number of patient perceptions may affect the uptake of interventions using BMDs and that most trials did not adequately cover all of them. Fourth, we only searched one database as a trade-off between feasibility and potential impact on results. As our work is a methodological review describing the characteristics of RCTs evaluating interventions using BMDs, we do not need the same exhaustivity as a meta-analysis to evaluate a therapeutic intervention; thus, the omission of some studies published in journals not indexed in MEDLINE is unlikely to change the results. Fifth, screening of search results' titles and abstracts was conducted by only one investigator (AP) instead of multiple assessors and could have resulted in the omission of some eligible trials.

A large number of patient perceptions can impact the uptake of a particular intervention using BMDs, help predict their real-world adoption, and guide the implementation of such interventions in routine clinical care. However, only a few of these perceptions are measured and only in fewer than half of clinical trials. Our review provides a simple schema of 11 important subcategories that comprehensively cover the factors that may affect the adoption of interventions using BMDs and could guide the development of future interventions. Future research should consider how intervention and BMD characteristics relate to perceptions toward interventions involving BMDs.

Acknowledgments

The authors wish to thank Ms Elise Diard for creating figures for this manuscript, and Dr Silvia Martins for her steadfast support of AP's research interests. This study was funded by a grant from the National Institutes of Health, National Institute on Drug Abuse (T32DA031099).

Authors' Contributions

AP, PR, and VTT designed the systematic review; AP screened and extracted content from publications; AP and MB confirmed the eligibility of publications; AP and VTT wrote the manuscript and created the schema; and PR helped edit the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 16 KB - jmir_v22i9e18986_app1.docx]

Multimedia Appendix 2
References of included studies.
[DOCX File , 20 KB - jmir_v22i9e18986_app2.docx]

Multimedia Appendix 3
Characteristics of the 58 included trials by whether RCT reports collected general and specific information on patients' perceptions about the intervention using BMDs.
[DOCX File , 21 KB - jmir_v22i9e18986_app3.docx]

References

1. Arnerić SP, Cedarbaum J, Khozin S, Papapetropoulos S, Hill D, Ropacki M, et al. Biometric monitoring devices for assessing end points in clinical trials: developing an ecosystem. Nat Rev Drug Discov 2017 Oct;16(10):736. [doi: 10.1038/nrd.2017.153] [Medline: 28935908]
2. Elenko E, Underwood L, Zohar D. Defining digital medicine. Nat Biotechnol 2015 May 12;33(5):456-461. [doi: 10.1038/nbt.3222] [Medline: 25965750]
3. Green EM, van Mourik R, Wolfus C, Heitner SB, Dur O, Semigran MJ. Machine learning detection of obstructive hypertrophic cardiomyopathy using a wearable biosensor. NPJ Digit Med 2019 Jun 24;2(1):57 [FREE Full text] [doi: 10.1038/s41746-019-0130-0] [Medline: 31304403]
4. Carreiro S, Fang H, Zhang J, Wittbold K, Weng S, Mullins R, et al. iMStrong: Deployment of a Biosensor System to Detect Cocaine Use. J Med Syst 2015 Dec 21;39(12):186 [FREE Full text] [doi: 10.1007/s10916-015-0337-9] [Medline: 26490144]
5. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The Future of Wearable Technologies and Remote Monitoring in Health Care. American Society of Clinical Oncology Educational Book 2019 May(39):115-121. [doi: 10.1200/edbk_238919]
6. Singh R, Lewis B, Chapman B, Carreiro S, Venkatasubramanian K. A Machine Learning-based Approach for Collaborative Non-Adherence Detection during Opioid Abuse Surveillance using a Wearable Biosensor. Biomed Eng Syst Technol Int Jt Conf BIOSTEC Revis Sel Pap 2019 Feb;5:310-318 [FREE Full text] [doi: 10.5220/0007382503100318] [Medline: 30993266]
7. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]
8. Ajana B. Digital health and the biopolitics of the Quantified Self. Digit Health 2017 Feb;3:2055207616689509 [FREE Full text] [doi: 10.1177/2055207616689509] [Medline: 29942580]
9. Rich E, Miah A. Mobile, wearable and ingestible health technologies: towards a critical research agenda. Health Sociology Review 2016 Jul 28;26(1):84-97. [doi: 10.1080/14461242.2016.1211486]
10. Ruckenstein M, Schüll ND. The Datafication of Health. Annu Rev Anthropol 2017 Oct 23;46(1):261-278. [doi: 10.1146/annurev-anthro-102116-041244]
11. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019 Jun 14;2(1):53 [FREE Full text] [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
12. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of Obtaining Measures of Lifestyle From a Smartphone App: The MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2017 Jan 01;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]

13. Radin J, Steinhubl S, Su A, Bhargava H, Greenberg B, Bot B, et al. The Healthy Pregnancy Research Program: transforming pregnancy research through a ResearchKit app. NPJ Digit Med 2018;1:45 [FREE Full text] [doi: 10.1038/s41746-018-0052-2] [Medline: 31304325]
14. Manini TM, Mendoza T, Battula M, Davoudi A, Kheirkhahan M, Young ME, et al. Perception of Older Adults Toward Smartwatch Technology for Assessing Pain and Related Patient-Reported Outcomes: Pilot Study. JMIR Mhealth Uhealth 2019 Mar 26;7(3):e10044 [FREE Full text] [doi: 10.2196/10044] [Medline: 30912756]
15. Grando MA, Bayuk M, Karway G, Corrette K, Groat D, Cook CB, et al. Patient Perception and Satisfaction With Insulin Pump System: Pilot User Experience Survey. J Diabetes Sci Technol 2019 Nov 05;13(6):1142-1148 [FREE Full text] [doi: 10.1177/1932296819843146] [Medline: 31055947]
16. Kropff J, DeJong J, Del Favero S, Place J, Messori M, Coestier B, AP@home consortium. Psychological outcomes of evening and night closed-loop insulin delivery under free living conditions in people with Type 1 diabetes: a 2-month randomized crossover trial. Diabet Med 2017 Feb 28;34(2):262-271 [FREE Full text] [doi: 10.1111/dme.13268] [Medline: 27696520]
17. Mosconi P, Radrezza S, Lettieri E, Santoro E. Use of Health Apps and Wearable Devices: Survey Among Italian Associations for Patient Advocacy. JMIR mHealth uHealth 2019 Jan 15;7(1):e10242 [FREE Full text] [doi: 10.2196/10242] [Medline: 30664455]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
19. Zhang D, Liu Q. Biosensors and bioelectronics on smartphone for portable biochemical detection. Biosens Bioelectron 2016 Jan 15;75:273-284. [doi: 10.1016/j.bios.2015.08.037] [Medline: 26319170]
20. Sezgin E, Yildirim S, Özkan-Yildirim S, Sumuer E, editors. Current and Emerging mHealth Technologies: Adoption, Implementation, and Use. Cham, Switzerland: Springer International Publishing; 2018.
21. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of Digital Health Studies Registered in ClinicalTrials.gov. JAMA Intern Med 2019 Jun 01;179(6):838-840 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7235] [Medline: 30801617]
22. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
23. Leinonen A, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, et al. Feasibility of Gamified Mobile Service Aimed at Physical Activation in Young Men: Population-Based Randomized Controlled Study (MOPO). JMIR Mhealth Uhealth 2017 Oct 10;5(10):e146 [FREE Full text] [doi: 10.2196/mhealth.6675] [Medline: 29017991]
24. Beck RW, Riddlesworth T, Ruedy K, Ahmann A, Bergenstal R, Haller S, DIAMOND Study Group. Effect of Continuous Glucose Monitoring on Glycemic Control in Adults With Type 1 Diabetes Using Insulin Injections: The DIAMOND Randomized Clinical Trial. JAMA 2017 Jan 24;317(4):371-378. [doi: 10.1001/jama.2016.19975] [Medline: 28118453]
25. Heron N, Kee F, Mant J, Reilly PM, Cupples M, Tully M, et al. Stroke Prevention Rehabilitation Intervention Trial of Exercise (SPRITE) - a randomised feasibility study. BMC Cardiovasc Disord 2017 Dec 12;17(1):290 [FREE Full text] [doi: 10.1186/s12872-017-0717-9] [Medline: 29233087]
26. Lakshminarayan K, Westberg S, Northuis C, Fuller CC, Ikramuddin F, Ezzeddine M, et al. A mHealth-based care model for improving hypertension control in stroke survivors: Pilot RCT. Contemp Clin Trials 2018 Jul;70:24-34 [FREE Full text] [doi: 10.1016/j.cct.2018.05.005] [Medline: 29763657]
27. Carpinella I, Cattaneo D, Bonora G, Bowman T, Martina L, Montesano A, et al. Wearable Sensor-Based Biofeedback Training for Balance and Gait in Parkinson Disease: A Pilot Randomized Controlled Trial. Arch Phys Med Rehabil 2017 Apr;98(4):622-630.e3. [doi: 10.1016/j.apmr.2016.11.003] [Medline: 27965005]
28. Barnett NP, Celio MA, Tidey JW, Murphy JG, Colby SM, Swift RM. A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. Addiction 2017 Jun 22;112(6):1025-1035 [FREE Full text] [doi: 10.1111/add.13767] [Medline: 28107772]
29. Abraham MB, Nicholas JA, Smith GJ, Fairchild JM, King BR, Ambler GR, PLGM Study Group. Reduction in Hypoglycemia With the Predictive Low-Glucose Management System: A Long-term Randomized Controlled Trial in Adolescents With Type 1 Diabetes. Diabetes Care 2018 Feb;41(2):303-310. [doi: 10.2337/dc17-1604] [Medline: 29191844]
30. Vandelanotte C, De Bourdeaudhuij I. Acceptability and feasibility of a computer-tailored physical activity intervention using stages of change: project FAITH. Health Educ Res 2003 Jun 01;18(3):304-317. [doi: 10.1093/her/cyf027] [Medline: 12828232]
31. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR mHealth uHealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
32. Barnard KD, Pinsker JE, Oliver N, Astle A, Dassau E, Kerr D. Future artificial pancreas technology for type 1 diabetes: what do users want? Diabetes Technol Ther 2015 May;17(5):311-315. [doi: 10.1089/dia.2014.0316] [Medline: 25629627]

33. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and Facilitators of Engagement With Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
34. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Industr Mngmnt & Data Systems 2015 Oct 19;115(9):1704-1723. [doi: 10.1108/IMDS-03-2015-0087]
35. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: An empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: 10.1016/j.ijmedinf.2015.12.010] [Medline: 26878757]
36. Lunney A, Cunningham NR, Eastin MS. Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. Computers in Human Behavior 2016 Dec;65:114-120. [doi: 10.1016/j.chb.2016.08.007]
37. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
38. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A Systematic Review of the Technology Acceptance Model in Health Informatics. Appl Clin Inform 2018 Jul 15;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
39. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

Abbreviations

BMD: biometric monitoring device
MeSH: Medical Subject Heading
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 30.03.20; peer-reviewed by M Falahee, J Tavares; comments to author 01.07.20; revised version received 23.07.20; accepted 26.07.20; published 11.09.20.

Please cite as:
Perlmutter A, Benchoufi M, Ravaud P, Tran VT
Identification of Patient Perceptions That Can Affect the Uptake of Interventions Using Biometric Monitoring Devices: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2020;22(9):e18986
URL: http://www.jmir.org/2020/9/e18986/
doi:10.2196/18986
PMID:32915153

Review

Use of Decision Support Tools to Empower Pregnant Women: Systematic Review

Elin Ngo[1], MSc; Maria Bich-Thuy Truong[1], MSc; Hedvig Nordeng[1,2], MSc, PhD

[1]PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, University of Oslo, Oslo, Norway

[2]Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

Corresponding Author:
Elin Ngo, MSc
PharmacoEpidemiology and Drug Safety Research Group
Department of Pharmacy
University of Oslo
Postboks 1068 Blindern
Oslo, 0316
Norway
Phone: 47 93849866
Email: e.t.p.ngo@farmasi.uio.no

Abstract

Background: Women face many health-related decisions during pregnancy. Digitalization, new technology, and a greater focus on empowering patients have driven the development of patient-centered decision support tools.

Objective: This systematic review provides an overview of studies investigating the effect of patient-centered decision support tools for pregnant women.

Methods: We searched 5 online databases, MEDLINE, EMBASE, Web of Science, PsycINFO, and Scopus, from inception to December 1, 2019. Two independent researchers screened titles, abstracts, and full-texts against the inclusion criteria. All studies investigating the effect of patient-centered decision support tools for health-related issues among pregnant women were included. Study characteristics and results were extracted using the review management tool Rayyan and analyzed according to topic, type of decision support tools, control group, outcome measurements, and results.

Results: The 25 eligible studies covered a range of health topics, including prenatal screening (n=10), gestational diabetes and weight gain (n=7), lifestyle (n=3), blood pressure and preeclampsia (n=2), depression (n=1), asthma (n=1), and psychological well-being (n=1). In general, the use of decision support tools increased women's knowledge, and recording symptoms enhanced satisfaction with maternity care.

Conclusions: The opportunities created by digitalization and technology should be used to develop innovative patient-centered decision support tools tailored to support pregnant women. Effect on clinical outcomes should be documented.

(J Med Internet Res 2020;22(9):e19436) doi:10.2196/19436

KEYWORDS

decision support tools; pregnancy; mobile application; empowerment

Introduction

Background

Patient-centered decision support tools are developed to involve patients in their own health-related decisions by (1) clearly stating the decisions that need to be made, (2) providing information about the options, outcomes, risks, and benefits, and (3) clarifying personal values. Decision support tools aim to complement, not replace, counseling from health care providers. The goal is to empower patients to make the decisions that are best for themselves and improve communication with their care providers [1,2].

Patient involvement in decision making varies among patient groups but is especially common among young women [3], coinciding with the time in life at which they become pregnant and, for many women, face completely new health-related decisions. In particular, decisions about medication use in pregnancy may be challenging, as it requires handling the unique task of weighing the benefits and risks of treatment for themselves against the benefits and risks for their unborn child.

These situations are not uncommon, as over 60% of pregnant women use medications at least once during pregnancy [4-6].

Prior studies [7] have shown that pregnant women actively seek information to enable them to make decisions about medication use in pregnancy. First time pregnant women are more likely to seek information about medications and health-related problems during pregnancy than women who have previously had children [8-10]. Despite the frequent use of the internet, pregnant women tend not to discuss the information they have retrieved online with their health care providers [11]. Provision of tailored and credible information though a decision support tool may have the potential to empower and improve informed decision making among pregnant women [12].

The last literature review [13] on patient-centered tools to support women's decisions during pregnancy was published in 2012. Since then, there has been an increased focus on digitalization and novel tools to empower patients. An updated literature review could help identify knowledge gaps concerning patient-centered decision support tools for pregnant women [14,15].

Objective

The aim of this systematic review was to identify studies evaluating the efficacy of patient-centered decision support tools for pregnant women and provide guidance for future research and the development of new, efficient tools.

Methods

Literature Search Strategy

The following online databases were searched from inception to January 18, 2019: MEDLINE, EMBASE, Web of Science, PsycINFO, and Scopus. An updated search was conducted December 1, 2019. Each database was searched using a customized search strategy (Multimedia Appendix 1). The following keywords or MeSH terms (Medical Subject Headings) were used for the database search: *pregnancy*, *parturition*, *prenatal care*, *antenatal care*, *mobile application*, *mobile health*, *decision support techniques*, *choice behavior*, *patient education*, *decision making*, *satisfaction*, *quality of life*, and *knowledge*.

Selection of Studies

The studies were selected in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [16].

Type of Study

Randomized controlled trials, cohort studies, register-based studies, and case-control studies were eligible for inclusion. Reviews, nonoriginal studies, Delphi studies, editorials, commentaries, letters to the editor, animal studies, and conference papers or abstracts were excluded. Full-texts in English were included in this review. Moreover, full texts in Norwegian, Swedish, or Danish were included, as the authors could fluently read papers in these languages.

Type of Participants

All studies focusing on women who used one or several patient-centered decision support tools during pregnancy regarding health- or pregnancy-related issues were included in this review. Studies evaluating decision support tools for use in the prepregnancy period, postpartum period, or delivery-related (eg, support during birth, cesarean delivery, mode of birth after cesarean section, or breech position) were excluded.

Type of Intervention

All types of tools (digital or paper-based) developed to support women's health-related decisions by providing tailored information to her situation or recordings in pregnancy were included.

Type of Control Group

Participants in the control group were pregnant women who received standard prenatal care or used a different decision support tool than the participants in the intervention group. A control group was not required in descriptive studies.

Types of Outcome Measures

Outcome measures that assessed the women's knowledge, satisfaction, decision making, quality of life, use experience, behaviors, or control of clinical measures in pregnancy were included.

Study Selection and Data Extraction

All studies identified from the 5 databases were saved in reference management software (EndNote X8.1). Duplicates were removed, and the remaining studies were uploaded to free online systematic review data management software (Rayyan) [17]. First, the 2 researchers (EN and MT) independently screened titles and abstracts against the inclusion criteria, and disagreements were discussed until consensus was reached. The full-texts included from the previous round were then independently screened and categorized by the same researchers using EndNote and Excel (Microsoft Inc). At this step, excluded studies were categorized as (1) full-text not available, (2) foreign language, (3) wrong publication type, (4) wrong study design, (5) the study did not investigate the use of a decision support tool, or (6) the study did not include pregnant women or irrelevant outcome (eg, delivery, cesarean section, and economic analyses).

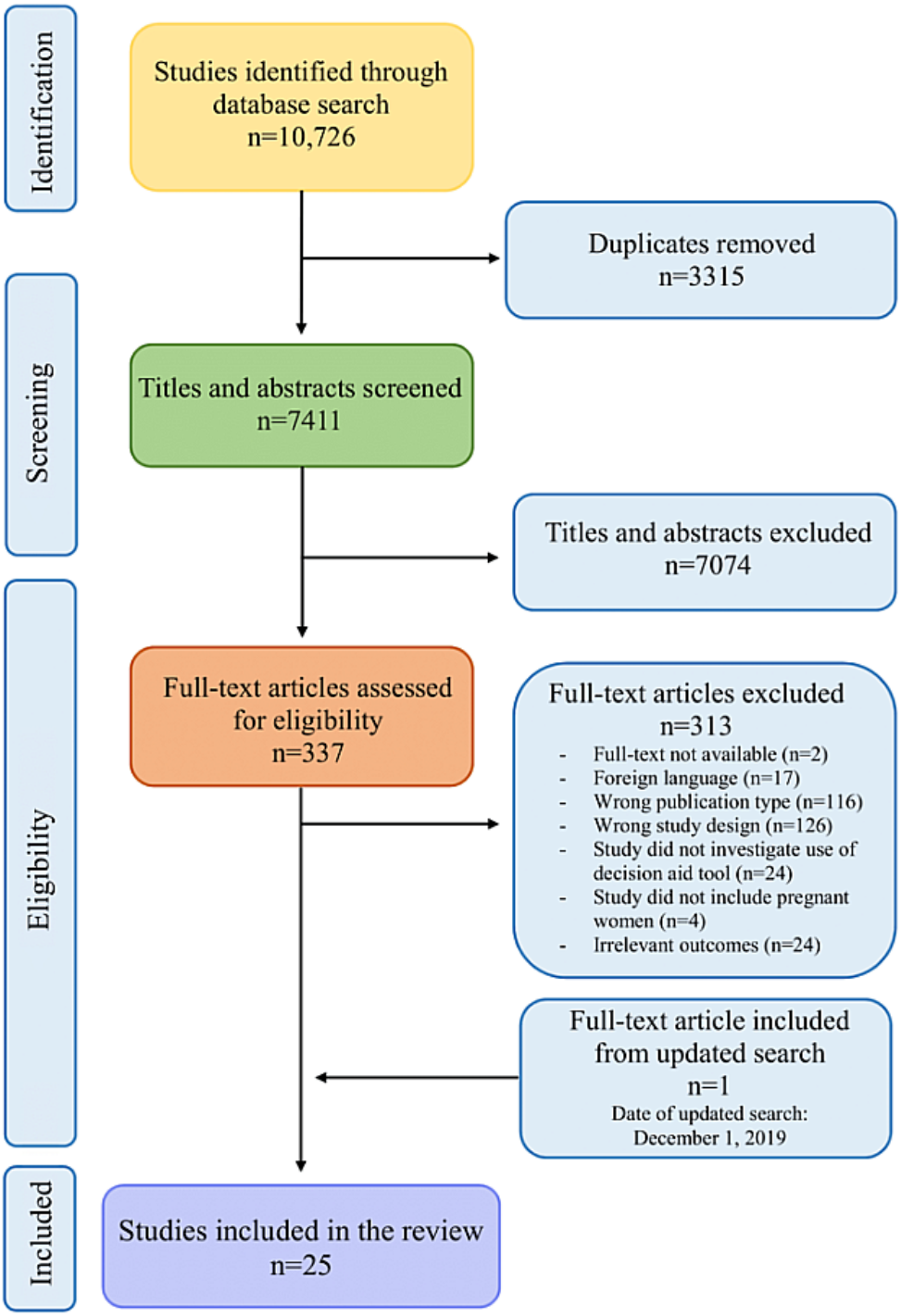
The studies included after the full-text screening were analyzed using a data extraction form (Multimedia Appendix 2). Information extracted from the studies included information about the study design, population, setting, method of recruitment, type of intervention or decision support tool, control group, outcomes measure, and results. Findings were grouped into major topics such as prenatal screening, gestational diabetes and weight gain, lifestyle, blood pressure and preeclampsia, depression, asthma, and physiological well-being.

Results

Search Findings

A total of 10,726 studies were initially identified in the first search (January 18, 2019) from the 5 online databases, with 7411 remaining after the deletion of duplicates. Of these, 7074 studies were excluded based on titles and abstracts, and 337 full-texts were screened for eligibility (Figure 1). The most common reason for exclusion was wrong study design (n=126).

Figure 1. Flowchart of the identification and selection of evaluated studies.



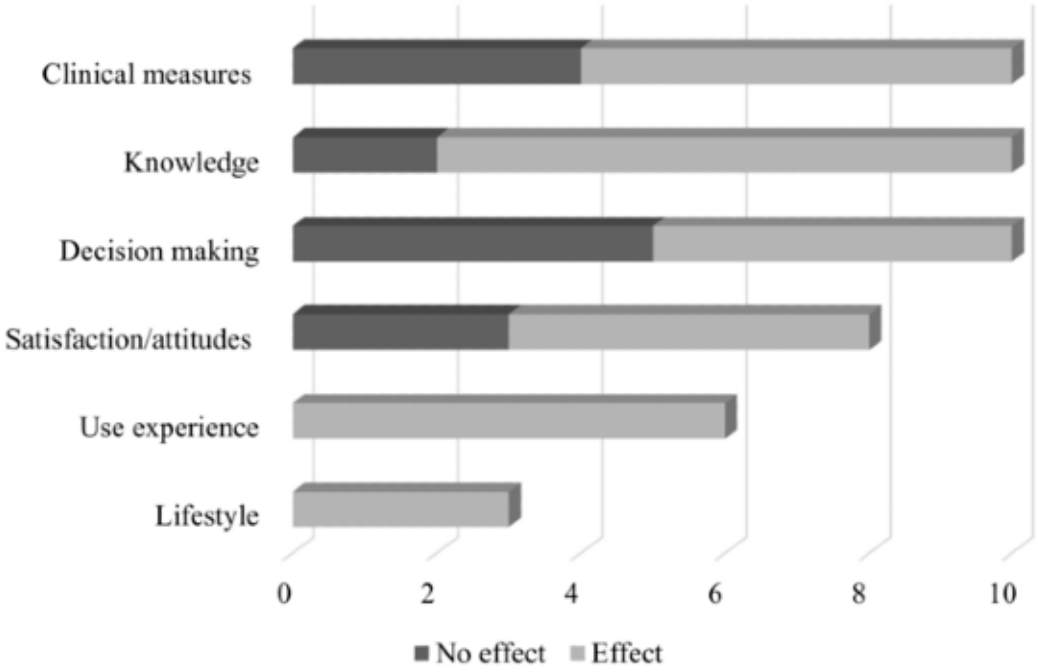
The updated search (December 1, 2019) identified 1221 new studies from the same databases as the first search. Of these, only 1 study was eligible for inclusion in this review after the screening process.

Included Studies

A total of 25 studies were included in this review, all in English. The studies covered 7 major topics: prenatal screening, gestational diabetes and weight gain, blood pressure and preeclampsia, lifestyle, depression, asthma, and physiological well-being (Multimedia Appendix 3). The decision support

tools were provided either as digital tools (webpage, mobile app, video, SMS text messages, n=24) or as written educational material (n=1). Outcome measures included in the digital decision support tools were clinical measures (n=10), knowledge level (n=10), decision making (n=10), satisfaction or attitudes (n=8), use experience (n=6), and lifestyle (n=3). One paper-based decision support tool investigated the effect on knowledge (n=1), attitudes (n=1), decision making (n=1), and clinical measures (n=2) (Figure 2). Several studies used multiple instruments for measuring the same outcome. The total number of outcome measures may thus exceed the number of studies included.

Figure 2. Effect of digital decision support tools.



Effect of Patient-Centered Decision Support as Interventions

Prenatal Screening

Ten studies [18-27] evaluated the effect of a patient-centered decision support tool on women's decisions about performing prenatal screening for genetic disorders and birth defects. Pregnant women at ≤26 gestational weeks were included in these studies. One study [24] did not have a cut-off on gestational weeks. Nine decision support tools were digital and one was provided as written material. The outcomes measured in these studies were knowledge (n=9), decision making (n=11), satisfaction or attitudes (n=6), clinical measures (n=3), and use experience (n=1).

Overall, women who used a decision support tool had higher knowledge scores than the control group and knew about the risks and benefits of genetic screening in pregnancy (Multimedia Appendix 3). Independent of the type of decision support tool, the results show decreased decisional conflict for women in the intervention group compared to those in standard care. This indicated that women using decision support tools felt more informed and were more aware of the risk and expected outcome of each option when compared to their counterparts [19-21,23,27]. Women using decision support tools also had better knowledge scores [19,21-23,25-27], except for in 2 studies [20,24] which showed no effect on knowledge. Both digital and paper-based decision support tools showed no difference in attitudes and frequency of completing screening (digital: 32%; paper-based: 15%; P=.087) [19,23,25,27].

Gestational Diabetes and Weight Gain

Seven studies [28-33] investigated the effect of using decision support tools on blood glucose level control for pregnant women with gestational diabetes. Two studies [28,34] evaluated the effect on gestational weight gain in general and among women with gestational diabetes. The evaluated decision support tools were apps (n=4), web-based tools (n=2), and SMS text message–based (n=1). Outcome measures were knowledge level (n=1), satisfaction (n=2), use experience (n=2), blood glucose level control (n=3), and weight control (n=2).

Women using an app to record blood glucose level readings daily, in addition to receiving SMS text messages from their doctor with advice when readings were abnormal, reported more blood glucose level readings than women who recorded their blood glucose level readings in a paper diary (app: 3.8; paper diary: 2.6 recordings per day) [30]. The vast majority of women with diabetes using the apps felt more satisfied with the care they received [29]. Women receiving tailored advice online (about blood glucose) from their care provider also had a better understanding of the risks related to gestational weight gain for themselves (tailored advice: 34%; control: 21%; P=.044) and the fetus (tailored advice: 62%; control: 38%; P=.001) [31].

Women using apps as decision support tool showed no difference or improvements in in blood glucose level control [28,30]. However, women who used a web-chat with direct

contact and feedback from their health care providers had significant lower fasting blood glucose level (web-chat and feedback: 4.3; control: 5.3; $P<.001$) and 2-hour postprandial blood glucose (web-chat and feedback: 5.8; control: 6.9; $P<.001$) [33]. They also felt they had more control of their symptoms and a better overview of their blood glucose when using a decision support tool as a supplement to standard care [32].

Lifestyle

Three studies [35-37] investigated the effect of decision support tools on alcohol consumption and smoking cessation during pregnancy. The tools were an app [35], a web-based tool [36], and an SMS text message–based tool [37].

A computer-tailored letter providing information about the risk of alcohol use in pregnancy had no effect on women's refrainment from alcohol use after 3 months when compared to standard care. They did, however, refrain from alcohol to a larger extent after 6 months (computer-tailored letter: 78%; standard care: 55%, $P=.04$) [36]. Providing SMS text messages with general pregnancy information also resulted in a decreased alcohol consumption in pregnancy compared to maternity care alone (SMS text messages: 3.5%; standard maternity care: 1.1%; $P<.098$) [37].

Blood Pressure and Preeclampsia

Two studies [38,39] investigated the effect of an app on blood pressure readings and knowledge about preeclampsia. Women using the app recorded their blood pressure and shared the information with their care provider more frequently [38]. They also had significantly higher knowledge scores than women not using the app (app user: 78.1; control: 15.8; $P<.001$) [39].

Depression

A recently published study [40] investigated the effect of a mood tracking and alert app among pregnant women with depression on mood and depressive symptoms measured by the Patient Health Questionnaire 9 [41]. The app also provided information about mental health and physical activity and alerted prenatal providers when depressive symptoms were worsening. All women in the study also had access to a patient portal that provided an overview of upcoming appointments and clinical results and which could be used to request prescription refills. Women in the intervention group recorded depressive symptoms an average 5.3 days per week. Their health care providers were more likely to mention mental health at check-ups ($P=.02$), and women using the app had a higher rate of referral to a mental health specialist ($P=.03$) [40].

Asthma

One study [42] investigated the effect of an app on asthma symptoms during pregnancy. In that study, 58% of the women had moderate to severe asthma. Women in the intervention group received a chronic obstructive pulmonary disease measurement device (COPD ‑ 6) in addition to an app for recording symptoms and medication use weekly, as well as with weekly feedback. Women in the intervention groups had better control of symptoms (Asthma Control Questionnaire: –0.30 vs. 0.06, $P=.02$), and quality of life (Asthma Quality-of-life Questionnaire score: 0.51 vs. –0.22, $P=.002$) after 6 months [42].

Psychological Well-Being

One study [43] investigated the use of a decision support tool and its effect on psychological well-being. Women received SMS text messages with information tailored to their gestational week, 2 times per week from gestational week 28 onward. Women receiving these SMS text messages had lower anxiety scores (2.8 vs. 4.9, $P=.002$) and higher confidence scores (8.9 vs. 7.8, $P=.001$) than women receiving standard care only [43].

Discussion

Main Findings

This systematic review provides an updated overview of current knowledge regarding patient-centered decision support tools for women during pregnancy. The 25 studies included more than 5000 women covering a broad range of health conditions in pregnancy. The majority of studies investigated the effect of a decision support tool in relation to prenatal screening (10/25, 40%) or gestational diabetes and weight gain during pregnancy (7/25, 28%). In general, the decision support tools were found to increase the women's knowledge and enhance communication with health care providers. Digital decision support tools also seemed to be more convenient and led to more recorded clinical data than what was recorded by paper-based tools.

Interestingly, almost all decision support tools, both digital and written material, increased the women's knowledge compared to knowledge received through standard care [19,21-27,31]. However, the majority of women participating in the studies were highly educated, and had been pregnant before; thus, they may not be representative of the general pregnant population. In addition, knowledge scores were most commonly measured immediately after the intervention was given or within 6 weeks. Therefore, whether gained knowledge lasted over time is unknown. One study [20] found no difference in knowledge between women receiving genetic counseling about prenatal screening with and without a supplementary app. The fact that both groups received a high-standard intervention such as genetic counseling could possibly explain why there was no additional benefit of the app on knowledge scores. Taken together, these results indicate that decision support during pregnancy, regardless of whether it is written or digital, may be a useful complement to standard antenatal care when specialized counseling is less available. It is still important to bear in mind that women receiving a consultation in advance may have been influenced to read more, which may have affected the results.

The studies included in this review show the potential of a patient-centered decision support tool to promote communication between health care providers and women. Women who frequently used digital support tools were more likely to bring their recordings to their health care provider. They were also more satisfied with the care they received and discussed their concerns with the health care provider to a greater extent than their counterparts did [27,29,31,38,40]. This indicates that women are more likely to discuss their problems

with their health care providers when they are knowledgeable about the topic [44-46]. It should be noted that many of the studies included samples of women of higher sociodemographic status than that of the general population of pregnant women. This may have caused a selection bias of potentially more resourceful or motivated women, limiting the generalizability of the findings to all pregnant women.

Interpretation in Light of Other Evidence

The use of decision support tools, in general, improves patient knowledge, make them better informed, and makes their choices and options clearer [47,48]. This review shows that this also applies to pregnant women. Mobile apps and decision support tools are increasingly used for self-management in many different chronic diseases that women of reproductive age have, such as migraine and diabetes, but high-quality decision support tools developed specifically for pregnancy are, to a large degree, still lacking. Moreover, there is clear potential for developing decision support tools to support decisions about medications in pregnancy. Nausea and vomiting in pregnancy, pain and self-managed conditions such as heartburn and constipation are examples where digital treatment algorithms may yet prove to be useful.

Our findings expand on and support earlier reviews that reported the potential benefits of decision support tools for decisions related to pregnancy. Both Say et al [49] and Dugas et al [13] advocated the potential for decision support tools to improve obstetric care. Our review included more studies that were recent (since 2012), even though our inclusion criteria were focused on decision support tools used only by women during pregnancy. More decision support tools after 2012 are electronic, as apps and web-based. The opportunities created by digitalization and technology should be used to develop innovative patient-centered decision support tools tailored to support pregnant women. Furthermore, the studies in our review covered a wider range of topics during pregnancy, but coverage of the most common topics regarding women's health during pregnancy was still lacking (eg, decision support tools for nausea and vomiting in pregnancy).

What Makes a Good Decision Support Tool for Pregnant Women?

The most effective decision support tools for pregnant women shared some common features. First, digital decision support tools seem more convenient if evidenced-based and if relevant information from different sources can be assembled in one app. This will avoid multiple or conflicting information sources, which has previously been an important concern among pregnancy women [50].

Second, digital tools that enable pregnant women to share recordings with their health care providers and get real-time feedback seem to be the most useful [18,29,32]. Such tools enable individually tailored information and improve communication during pregnancy. This is in line with previous findings on weight gain in pregnancy showing that specific and tailored information is more effective than general information [34].

Lastly, digital decision support tools were more convenient for recording symptoms than spiral notebooks. Women using digital support tools recorded their symptoms more frequently [38]. An earlier study [51] comparing the use of digital tools and spiral notebooks in general also reported that digital tools are potentially more accurate. This indicates that future development of decision support tools should focus and invest in digital tools.

A Supplement, Not a Replacement

Even with increased technology, there is still a gap in the development of patient-centered decision support tools for pregnancy-related conditions. Given that women have high information needs and the potential that decision support tools have in empowering them, we expect this can be a valuable supplement for both women and their health care providers during prenatal care. Given that women were more satisfied with and were more likely to discuss their health problems with their care providers [30,31,38,40], it seems plausible that patient-centered decision support tools may promote healthier pregnancies and reduce the burden on health care services, with little extra cost after development. Decision support tools do not replace health care providers but provide additional relevant clinical information, supporting women to make better decisions together with their health care providers.

The sparseness of studies evaluating the effect of decision support tools, especially on clinical outcomes, stands in great contrast to the number of apps targeting pregnant women. This highlights the importance of developing and testing decision support tools for pregnant women. Only tools that are of high quality and that are efficient should be promoted.

Limitations

This literature review has some limitations that should be taken into consideration when interpreting the results. First, there were few patient-centered decision support tools within each topic, and the diversity of outcome measures made it challenging to draw overall conclusions. Second, the individual studies overrepresented women with higher sociodemographic status, and the majority of pregnant women included in the studies were of a white ethnic background. Third, a number of studies had a low number of participants, and the women who consented to the studies may have been motivated to participate, which can cause a selection bias and give more positive results than what would be achieved in the typical target population.

Studies including decision support tools used by health care providers, decision support tools regarding childbirth, maternal and fetal health outcomes, and decision tools used in the postpartum period were excluded. An expanded review including these outcomes and topics should be assessed in future studies and may provide greater insight into the field.

Conclusion

Despite the technological possibilities, the focus on patient involvement, and documented information needs, few heterogeneous studies have been performed on the effect of decision support tools in pregnancy. These few studies, however, have demonstrated the potential benefit to knowledge, perception, confidence in decision making, and communication

between the women and their health care providers. More decision support tools should be developed and tailored to meet the needs of pregnant patients. The effect of such tools on clinical outcomes should be tested before recommending them or implementing them as a supplement in routine maternity care.

Acknowledgments

The authors would like to thank the librarians at the University of Oslo Science Library for their feedback on the literature search and for providing the studies that were not available online. The authors also wish to thank the University of Oslo Medical Library for their valuable course on the use of the review management tool Rayyan.

The study was funded by Foundation Dam through the Norwegian Women's Public Health Association. HN is funded by the European Research Council Starting Grant DrugsInPregnancy (grant number 639377).

Authors' Contributions

EN, MT, and HN designed the study. EN and MT performed the systematic search and conducted the main analysis. EN drafted the first version of the manuscript. EN, MT, and HN contributed to the interpretation of results and critical appraisal of the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 41 KB - jmir_v22i9e19436_app1.docx]

Multimedia Appendix 2
Extraction sheet.
[DOCX File , 25 KB - jmir_v22i9e19436_app2.docx]

Multimedia Appendix 3
The characteristics of studies included in this review.
[DOCX File , 39 KB - jmir_v22i9e19436_app3.docx]

References

1. DrugTherapeutics Bulletin. An introduction to patient decision aids. BMJ 2013 Jul 23;347:f4147. [doi: 10.1136/bmj.f4147] [Medline: 23881944]
2. The Ottawa Hospital. Patient Decision Aids. 2019. URL: https://decisionaid.ohri.ca/ [accessed 2019-12-02]
3. Say R, Murtagh M, Thomson R. Patients' preference for involvement in medical decision making: A narrative review. Patient Education and Counseling 2006 Feb;60(2):102-114. [doi: 10.1016/j.pec.2005.02.003]
4. Irvine L, Flynn RW, Libby G, Crombie IK, Evans JM. Drugs Dispensed in Primary Care During Pregnancy. Drug Safety 2010;33(7):593-604. [doi: 10.2165/11532330-000000000-00000]
5. Lupattelli A, Spigset O, Twigg MJ, Zagorodnikova K, Mårdby AC, Moretti ME, et al. Medication use in pregnancy: a cross-sectional, multinational web-based study. BMJ Open 2014 Feb 17;4(2):e004365. [doi: 10.1136/bmjopen-2013-004365]
6. Navaro M, Vezzosi L, Santagati G, Angelillo IF. Knowledge, attitudes, and practice regarding medication use in pregnant women in Southern Italy. PLoS ONE 2018 Jun 19;13(6):e0198618. [doi: 10.1371/journal.pone.0198618]
7. Lynch MM, Squiers LB, Kosa KM, Dolina S, Read JG, Broussard CS, et al. Making Decisions About Medication Use During Pregnancy: Implications for Communication Strategies. Matern Child Health J 2017 Sep 12;22(1):92-100. [doi: 10.1007/s10995-017-2358-0]
8. Hämeen-Anttila K, Jyrkkä J, Enlund H, Nordeng H, Lupattelli A, Kokki E. Medicines information needs during pregnancy: a multinational comparison. BMJ Open 2013 Apr 26;3(4):e002594. [doi: 10.1136/bmjopen-2013-002594]
9. Kamali S, Ahmadian L, Khajouei R, Bahaadinbeigy K. Health information needs of pregnant women: information sources, motives and barriers. Health Info Libr J 2017 Nov 13;35(1):24-37. [doi: 10.1111/hir.12200]
10. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
11. Ceulemans M, Van Calsteren K, Allegaert K, Foulon V. Beliefs about medicines and information needs among pregnant women visiting a tertiary hospital in Belgium. Eur J Clin Pharmacol 2019 Mar 4;75(7):995-1003. [doi: 10.1007/s00228-019-02653-w]

12. Lynch MM, Amoozegar JB, McClure EM, Squiers LB, Broussard CS, Lind JN, et al. Improving Safe Use of Medications During Pregnancy: The Roles of Patients, Physicians, and Pharmacists. Qual Health Res 2017 Oct 03;27(13):2071-2080. [doi: 10.1177/1049732317732027]
13. Dugas M, Shorten A, Dubé E, Wassef M, Bujold E, Chaillet N. Decision aid tools to support women's decision making in pregnancy and birth: A systematic review and meta-analysis. Social Science & Medicine 2012 Jun;74(12):1968-1978. [doi: 10.1016/j.socscimed.2012.01.041]
14. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29:a1655. [doi: 10.1136/bmj.a1655]
15. Grewal A, Kataria H, Dhawan I. Literature search for research planning and identification of research problem. Indian J Anaesth 2016;60(9):635. [doi: 10.4103/0019-5049.190618]
16. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700-b2700. [doi: 10.1136/bmj.b2700]
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
18. Åhman A, Sarkadi A, Lindgren P, Rubertsson C. 'It made you think twice' - an interview study of women's perception of a web-based decision aid concerning screening and diagnostic testing for fetal anomalies. BMC Pregnancy Childbirth 2016 Sep 13;16(1):267 [FREE Full text] [doi: 10.1186/s12884-016-1057-y] [Medline: 27619366]
19. Beulen L, van den Berg M, Faas BH, Feenstra I, Hageman M, van Vugt JM, et al. The effect of a decision aid on informed decision-making in the era of non-invasive prenatal testing: a randomised controlled trial. Eur J Hum Genet 2016 May 18;24(10):1409-1416. [doi: 10.1038/ejhg.2016.39]
20. Carlson LM, Harris S, Hardisty EE, Hocutt G, Vargo D, Campbell E, et al. Use of a novel computerized decision aid for aneuploidy screening: a randomized controlled trial. Genet Med 2018 Sep 14;21(4):923-929. [doi: 10.1038/s41436-018-0283-2]
21. Kuppermann M, Norton ME, Gates E, Gregorich SE, Learman LA, Nakagawa S, et al. Computerized Prenatal Genetic Testing Decision-Assisting Tool. Obstetrics & Gynecology 2009;113(1):53-63. [doi: 10.1097/aog.0b013e31818e7ec4]
22. Kuppermann M, Pena S, Bishop JT, Nakagawa S, Gregorich SE, Sit A, et al. Effect of Enhanced Information, Values Clarification, and Removal of Financial Barriers on Use of Prenatal Genetic Testing. JAMA 2014 Sep 24;312(12):1210. [doi: 10.1001/jama.2014.11479]
23. Nagle C, Gunn J, Bell R, Lewis S, Meiser B, Metcalfe S, et al. Use of a decision aid for prenatal testing of fetal abnormalities to improve women's informed decision making: a cluster randomised controlled trial [ISRCTN22532458]. BJOG: Int J O & G 2008 Feb;115(3):339-347. [doi: 10.1111/j.1471-0528.2007.01576.x]
24. Skjøth MM, Draborg E, Lamont RF, Pedersen CD, Hansen HP, Ekstrøm CT, et al. Informed choice about Down syndrome screening - effect of an eHealth tool: a randomized controlled trial. Acta Obstet Gynecol Scand 2015 Sep 18;94(12):1327-1336. [doi: 10.1111/aogs.12758]
25. Rothwell E, Johnson E, Wong B, Rose N, Latendresse G, Altizer R, et al. The Use of a Game-Based Decision Aid to Educate Pregnant Women about Prenatal Screening: A Randomized Controlled Study. Amer J Perinatol 2018 Aug 14;36(03):322-328. [doi: 10.1055/s-0038-1667371]
26. Yee LM, Wolf M, Mullen R, Bergeron AR, Cooper Bailey S, Levine R, et al. A randomized trial of a prenatal genetic testing interactive computerized information aid. Prenat Diagn 2014 Mar 18;34(6):552-557. [doi: 10.1002/pd.4347]
27. Björklund U, Marsk A, Levin C, Öhman SG. Audiovisual information affects informed choice and experience of information in antenatal Down syndrome screening – A randomized controlled trial. Patient Education and Counseling 2012 Mar;86(3):390-395. [doi: 10.1016/j.pec.2011.07.004]
28. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2018 Nov 7;42(6):709-714. [doi: 10.1007/s40618-018-0975-0]
29. Hirst JE, Mackillop L, Loerup L, Kevat DA, Bartlett K, Gibson O, et al. Acceptability and User Satisfaction of a Smartphone-Based, Interactive Blood Glucose Management System in Women With Gestational Diabetes Mellitus. J Diabetes Sci Technol 2014 Oct 30;9(1):111-115. [doi: 10.1177/1932296814556506]
30. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the Efficacy of a Mobile Phone-Based Blood Glucose Management System With Standard Clinic Care in Women With Gestational Diabetes: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71. [doi: 10.2196/mhealth.9512]
31. McDonald SD, Park CK, Pullenayegum E, Bracken K, Sword W, McDonald H, et al. Knowledge translation tool to improve pregnant women's awareness of gestational weight gain goals and risks of gaining outside recommendations: a non-randomized intervention study. BMC Pregnancy Childbirth 2015 May 30;15(1):105 [FREE Full text] [doi: 10.1186/s12884-015-0534-z] [Medline: 25925384]
32. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021]

33. Yang P, Lo W, He Z, Xiao X. Medical nutrition treatment of women with gestational diabetes mellitus by a telemedicine system based on smartphones. J. Obstet. Gynaecol. Res 2018 May 23;44(7):1228-1234. [doi: 10.1111/jog.13669]
34. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: A pilot study. Patient Education and Counseling 2014 Nov;97(2):256-260. [doi: 10.1016/j.pec.2014.07.030]
35. Dotson JAW, Pineda R, Cylkowski H, Amiri S. Development and Evaluation of an iPad Application to Promote Knowledge of Tobacco Use and Cessation by Pregnant Women. Nursing for Women's Health 2017 Jun;21(3):174-185. [doi: 10.1016/j.nwh.2017.04.005]
36. van der Wulp NY, Hoving C, Eijmael K, Candel MJ, van Dalen W, De Vries H. Reducing Alcohol Use During Pregnancy Via Health Counseling by Midwives and Internet-Based Computer-Tailored Feedback: A Cluster Randomized Trial. J Med Internet Res 2014 Dec 05;16(12):e274. [doi: 10.2196/jmir.3493]
37. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012 Dec 26;12(1):1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
38. Ledford CJW, Womack JJ, Rider HA, Seehusen AB, Conner SJ, Lauters RA, et al. Unexpected Effects of a System-Distributed Mobile Application in Maternity Care: A Randomized Controlled Trial. Health Educ Behav 2017 Sep 16;45(3):323-330. [doi: 10.1177/1090198117732110]
39. Parsa S, Khajouei R, Baneshi MR, Aali BS. Improving the knowledge of pregnant women using a pre-eclampsia app: A controlled before and after study. International Journal of Medical Informatics 2019 May;125:86-90. [doi: 10.1016/j.ijmedinf.2019.03.001]
40. Hantsoo L, Criniti S, Khan A, Moseley M, Kincler N, Faherty LJ, et al. A Mobile Application for Monitoring and Management of Depressed Mood in a Vulnerable Pregnant Population. PS 2018 Jan;69(1):104-107. [doi: 10.1176/appi.ps.201600582]
41. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
42. Zairina E, Abramson MJ, McDonald CF, Li J, Dharmasiri T, Stewart K, et al. Telehealth to improve asthma control in pregnancy: A randomized controlled trial. Respirology 2016 Mar 31;21(5):867-874. [doi: 10.1111/resp.12773]
43. Jareethum R, Titapant V, Chantra T, Sommai V, Chuenwattana P, Jirawan C. Satisfaction of healthy pregnant women receiving short message service via mobile phone for prenatal support: A randomized controlled trial. J Med Assoc Thai 2008 May;91(4):458-463. [Medline: 18556852]
44. Song H, Cramer EM, McRoy S, May A. Information Needs, Seeking Behaviors, and Support Among Low-Income Expectant Women. Women & Health 2013 Nov;53(8):824-842. [doi: 10.1080/03630242.2013.831019]
45. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003]
46. McMullan M. Patients using the Internet to obtain health information: How this affects the patient–health professional relationship. Patient Education and Counseling 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006]
47. Hsieh Y, Brennan PF. What are pregnant women's information needs and information seeking behaviors prior to their prenatal genetic counseling? AMIA Annu Symp Proc 2005:355-359 [FREE Full text] [Medline: 16779061]
48. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014 Jan 28(1):CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]
49. Say R, Robson S, Thomson R. Helping pregnant women make better decisions: a systematic review of the benefits of patient decision aids in obstetrics. BMJ Open 2011 Dec 21;1(2):e000261-e000261. [doi: 10.1136/bmjopen-2011-000261]
50. Hämeen-Anttila K, Nordeng H, Kokki E, Jyrkkä J, Lupattelli A, Vainio K, et al. Multiple Information Sources and Consequences of Conflicting Information About Medicine Use During Pregnancy: A Multinational Internet-Based Survey. J Med Internet Res 2014 Feb 20;16(2):e60. [doi: 10.2196/jmir.2939]
51. Meltzer EO, Kelley N, Hovell MF. Randomized, Cross-Over Evaluation of Mobile Phone vs Paper Diary in Subjects with Mild to Moderate Persistent Asthma~!2008-07-02~!2008-08-08~!2008-09-05~!. TORMJ 2008 Oct 15;2(1):72-79. [doi: 10.2174/1874306400802010072]

Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by M Ceulemans, W Gyselaers; comments to author 12.06.20; revised version received 25.06.20; accepted 25.06.20; published 14.09.20.

Please cite as:
Ngo E, Truong MBT, Nordeng H
Use of Decision Support Tools to Empower Pregnant Women: Systematic Review
J Med Internet Res 2020;22(9):e19436
URL: http://www.jmir.org/2020/9/e19436/
doi:10.2196/19436
PMID:32924961

Review

Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review

Theresa Schachner[1], BSc, BA, MSc; Roman Keller[1,2], BSc, MSc; Florian v Wangenheim[1,2], Prof Dr

[1]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Future Health Technologies programme, Campus for Research Excellence and Technological Enterprise, Singapore-ETH Centre, Singapore

Corresponding Author:
Theresa Schachner, BSc, BA, MSc
Department of Management, Technology, and Economics
ETH Zurich
WEV G 228, Weinbergstr 56/58
Zurich
Switzerland
Phone: 41 446325209
Email: tschachner@ethz.ch

Abstract

Background: A rising number of conversational agents or chatbots are equipped with artificial intelligence (AI) architecture. They are increasingly prevalent in health care applications such as those providing education and support to patients with chronic diseases, one of the leading causes of death in the 21st century. AI-based chatbots enable more effective and frequent interactions with such patients.

Objective: The goal of this systematic literature review is to review the characteristics, health care conditions, and AI architectures of AI-based conversational agents designed specifically for chronic diseases.

Methods: We conducted a systematic literature review using PubMed MEDLINE, EMBASE, PyscInfo, CINAHL, ACM Digital Library, ScienceDirect, and Web of Science. We applied a predefined search strategy using the terms “conversational agent,” “healthcare,” “artificial intelligence,” and their synonyms. We updated the search results using Google alerts, and screened reference lists for other relevant articles. We included primary research studies that involved the prevention, treatment, or rehabilitation of chronic diseases, involved a conversational agent, and included any kind of AI architecture. Two independent reviewers conducted screening and data extraction, and Cohen kappa was used to measure interrater agreement.A narrative approach was applied for data synthesis.

Results: The literature search found 2052 articles, out of which 10 papers met the inclusion criteria. The small number of identified studies together with the prevalence of quasi-experimental studies (n=7) and prevailing prototype nature of the chatbots (n=7) revealed the immaturity of the field. The reported chatbots addressed a broad variety of chronic diseases (n=6), showcasing a tendency to develop specialized conversational agents for individual chronic conditions. However, there lacks comparison of these chatbots within and between chronic diseases. In addition, the reported evaluation measures were not standardized, and the addressed health goals showed a large range. Together, these study characteristics complicated comparability and open room for future research. While natural language processing represented the most used AI technique (n=7) and the majority of conversational agents allowed for multimodal interaction (n=6), the identified studies demonstrated broad heterogeneity, lack of depth of reported AI techniques and systems, and inconsistent usage of taxonomy of the underlying AI software, further aggravating comparability and generalizability of study results.

Conclusions: The literature on AI-based conversational agents for chronic conditions is scarce and mostly consists of quasi-experimental studies with chatbots in prototype stage that use natural language processing and allow for multimodal user interaction. Future research could profit from evidence-based evaluation of the AI-based conversational agents and comparison thereof within and between different chronic health conditions. Besides increased comparability, the quality of chatbots developed for specific chronic conditions and their subsequent impact on the target patients could be enhanced by more structured development and standardized evaluation processes.

(J Med Internet Res 2020;22(9):e20701) doi:10.2196/20701

KEYWORDS

artificial intelligence; conversational agents; chatbots; healthcare; chronic diseases; systematic literature review

Introduction

Conversational agents or chatbots are computer systems that imitate natural conversation with human users through images and written or spoken language [1]. This paper focuses on conversational agents that deploy intelligent software or artificial intelligence (AI), which is increasingly used for applications in credit scoring [2], marketing strategies [3], and medical image analysis in radiology [4].

There are several ways of defining AI, as discussed by Russel and Norvig [5] in 1995. Their commonality is that AI describes algorithms that artificially emulate human cognitive and behavioral thought processes and are instantiated in software programs. Since then, the number of definitions had risen with the growing number of AI applications [6]. There are several specific understandings of AI such as by De Bruyn et al [7], who define AI as software that can "autonomously generate new constructs and knowledge structures" [7]. More general approaches describe and distinguish between weak AI, strong AI, and artificial general intelligence (AGI). Coined by John Searle in 1980, the term weak AI describes software that appears intelligent by mimicking specific human cognitive processes such as image recognition or natural language processing [8]. Strong AI denotes software that truly possesses intelligence without mimicking it [8]. AGI as an expansion of these terms designates true intelligence for all human cognitive processes instead of just for individual tasks [9,10]. For this paper, we adopt the understanding of weak AI when talking about AI-based conversational agents; the algorithms implemented in the conversational agent software each mimic distinct and narrowly restricted human cognitive processes.

The latest advances in AI allow for increasingly natural interactions between humans and their machine agent counterparts [11,12]. This emulated human-machine communication becomes more complex and sophisticated, especially through advancements in machine learning with the application of neural networks [13-15]. This is reflected in the rising number of conversational agents that aim at human-like exchanges [16] in fields such as e-commerce, travel, tourism, and health care [17-19]. Well-known examples of such intelligent chatbots are Microsoft's Cortana, Amazon's Alexa, or Apple's Siri [12].

The focus on the human-machine relationship was present from the very beginning in the history of chatbots; the rule-based software program ELIZA [20] was designed to take on the role of a psychotherapist in order to mimic a patient-centered Rogerian psychotherapy exchange. Developed in 1966 by Joseph Weizenbaum, it was then followed by PARRY, another mental health care–related chatbot developed in 1972 [21]. While ELIZA played the role of the therapist, PARRY took on the part of a schizophrenic patient [20,21]. Even though ELIZA passed a restricted Turing Test—a machine intelligence test with the success criterion of whether a human can distinguish a machine from a human during a conversation [22]—it was a rule-based and pre-scripted software program [23]. Similarly, other early forms of the then-called chatterbots such as Psyxpert, an expert system for disease diagnosis support written in Prolog [24] or SESAM-DIABETE, an expert system for diabetic patient education written in Lisp [25], followed a rule-based approach. ALICE (Artificial Linguistic Internet Computer Entity), in 1995, was the first computer system to use natural language processing for the interpretation of user input [12].

Since then, increasingly efficient access to and storage of data, decreasing hardware costs, and eased access to cloud-based services improved the development of AI architecture [26]. These advances gave rise to a more standardized deployment of natural language processing, voice recognition, natural language generation, and the like within chatbot development [11,12].

In health care, such AI-based conversational agents have demonstrated multiple benefits for disease diagnosis, monitoring, or treatment support in the last two decades [1,19,27,28]. They are used as digital interventions to deliver cost-efficient, scalable, and personalized medical support solutions that can be delivered at any time and any place via web-based or mobile apps [29-31]. Research studies have investigated a variety of AI-based conversational agents for different health care applications such as providing information to breast cancer patients [32]; providing information about sex, drugs, and alcohol to adolescents [33]; self-anamnesis for therapy patients [34]; assistance for health coaching to promote a healthy lifestyle [35]; or smoking cessation [36].

This paper focuses on one of the most urgent health care challenges of the 21st century—the rise of chronic conditions [37]. Chronic diseases are one of the leading drivers for reduced quality of life and increased economic health care expenses through repeated hospitalization, disability, and treatment expenditures [38]. In the United States alone, they affected over 50% of adults in 2016 and accounted for 86% of health care spending [37]. Hvidberg et al [39] and others defined chronic conditions as ailments that are anticipated to last at least 12 or more months, lead to functional limitations, and require continuous medical support [40,41]. As such, they require fundamentally different prevention, treatment, and management approaches than acute conditions, which are episodic, allow for general solutions, and can be treated within health care sites [37]. In contrast, chronic conditions require challenging lifestyle and behavioral changes, frequent self-care, and ongoing and personalized treatment that go beyond traditional health care sites and reach personal settings [37,42,43]. AI-based conversational agents provide suitable, personalized, and affordable digital solutions to react to these challenges and slow down individual disease deterioration to delay premature death.

Systematic literature reviews investigated a variety of contexts of health care chatbots such as the role of conversational agents in health care in general [1] and in mental health [44], aspects of personalization of health care chatbots [45], as well as technical aspects of AI systems and architectures of

conversational agents in health care [11]. However, there is surprisingly little systematic information on the application of AI-based conversational agents in health care for chronic diseases. This paper closes the gap. The objective of this paper is to identify the state of research of AI-based conversational agents in health care for chronic diseases. We extract stable findings and structures by outlining conversational agent characteristics, their underlying AI architectures, and health care applications. Additionally, we outline gaps and important open points that serve as guidelines for future research.

Methods

Reporting Standards

We performed a systematic literature review and followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [46]. The review protocol is available in the Multimedia Appendix 1.

Search Strategy

The search was conducted electronically during February 2020, using PubMed MEDLINE, EMBASE, PyscInfo, CINAHL, ACM Digital Library, ScienceDirect, and Web of Science. These databases were chosen as they cover relevant aspects in medicine and technology and have been used in other systematic literature reviews covering similar topics [1,45]. The search was updated by additional abstracts retrieved through various Google alerts covering different combinations of the search term until April 2020. The reference lists of other relevant literature reviews and articles were screened for additional articles. The process of query construction was initially informed by the first author's experience in the investigated areas and extended by incorporating associated terms such as synonyms, acronyms, and commonly known terms of the same context. The final search term included an extensive list of items describing the constructs "conversational agent," "healthcare," and "artificial intelligence" to ensure exhaustive coverage of the search space. The complete overview of the search terms for each construct is available in Multimedia Appendix 2. An exemplary search strategy is shown for PubMed MEDLINE in Table 1.

Table 1. The search strategy used in PubMed MEDLINE.

Search category	Search terms
Health care	"healthcare" OR "digital healthcare" OR "digital health" OR "health" OR "mobile health" OR "mHealth" OR "mobile healthcare"
Conversational agents	"conversational agent" OR "conversational agents" OR "conversational system" OR "conversational systems" OR "dialog system" OR "dialog systems" OR "dialogue systems" OR "dialogue system" OR "assistance technology" OR "assistance technologies" OR "relational agent" OR "relational agents" OR "chatbot" OR "chatbots" OR "digital agent" OR "digital agents" OR "digital assistant" OR "digital assistants" OR "virtual assistant" OR "virtual assistants"
Artificial intelligence	"artificial intelligence" OR "AI" OR "natural language processing" OR "NLP" OR "natural language understanding" OR "NLU" OR "machine learning" OR "deep learning" OR "neural network" OR "neural networks"
Combined	1 AND 2 AND 3

Selection Criteria

We included studies if they (1) were primary research studies that involved the prevention, treatment, or rehabilitation of chronic diseases; (2) involved a conversational agent; and (3) included any kind of artificial intelligence technique such as natural language understanding or deep learning for data processing.

Articles were excluded if they (1) involved only non-AI software architecture; (2) involved purely Wizard of Oz–based studies where the dialogue between human and conversational agent was mimicked by a human rather than performed by the conversational agent; (3) addressed health conditions and diseases that cannot conclusively be referred to as chronic diseases, general health, or any form of prechronic health conditions such as general well-being for the prevention of chronic diseases; (4) addressed chronic health conditions on a general level without specifying a disease or if the chronic disease only played a minor role for the study or was only mentioned in a few sentences.

Furthermore, we excluded studies without specific applications of conversational agents or where the application of the conversational agent for chronic diseases was only mentioned as a possibility or in a couple of sentences. We also excluded non-English papers, conference papers, workshop papers, literature reviews, posters, PowerPoint presentations, articles presented at doctoral colloquia, or if the article's full text was not accessible for the study authors.

Selection Process

All references that were identified through the searches were downloaded into Excel (Microsoft Corporation) and inserted in an Excel spreadsheet. Duplicates were removed. Screening was conducted by two independent reviewers in three phases, assessing first the article titles, followed by the abstracts, and finally the full texts. After each of these phases, Cohen kappa was calculated to measure interrater reliability between the researchers and determine the level of agreement [47]. Any disagreements were discussed and resolved in consensus.

Data Extraction

The two reviewers familiarized themselves with the identified articles and then independently extracted the contained information into an Excel spreadsheet with 30 columns containing information on the following aspects: (1) general information about the included studies, (2) health care/chronic conditions, (3) conversational agents, (4) AI, and (5) additional

study items such as conflict of interests or reported funding. We extracted data such as first author, year of publication, study design/type, study aim, conversational agent evaluation measures, main reported outcomes and findings, type of chronic condition, type of study participants, AI technique, AI system development, sources of funding, and conflicts of interest.

The full list can be seen in Multimedia Appendix 3. The extracted data were synthesized narratively. Quality of studies was not assessed in this analysis due to the diversity of analyzed studies. Any inconsistencies after the individual data extractions were discussed and resolved in consensus agreement.

Risk of Methodological Bias

The author team engaged in extensive discussion about the selection of an appropriate tool to assess methodological biases of the included studies, given the variety of study designs and the diversity of reported evaluation measures.

After extensive research in relevant journals, we decided to follow the approach of Maher et al [48], who devised a risk assessment tool based on the Consolidated Standards of Reporting Trials (CONSORT) checklist [49]. The tool developed by Maher et al [48] contains all 25 items from the CONSORT checklist and assigns scores of 1 or 0 to each item per study, indicating whether the item was satisfactorily fulfilled or not in the respective study. Lower scores imply higher risk of methodological bias and the inverse for higher scores. Whereas the CONSORT checklist was originally developed for controlled trials, we concluded that most of its criteria are applicable. We adapted the tool by Maher et al [48] by allowing scoring from 0 to 1 in order to more precisely assess the achieved score of each checklist item per study.

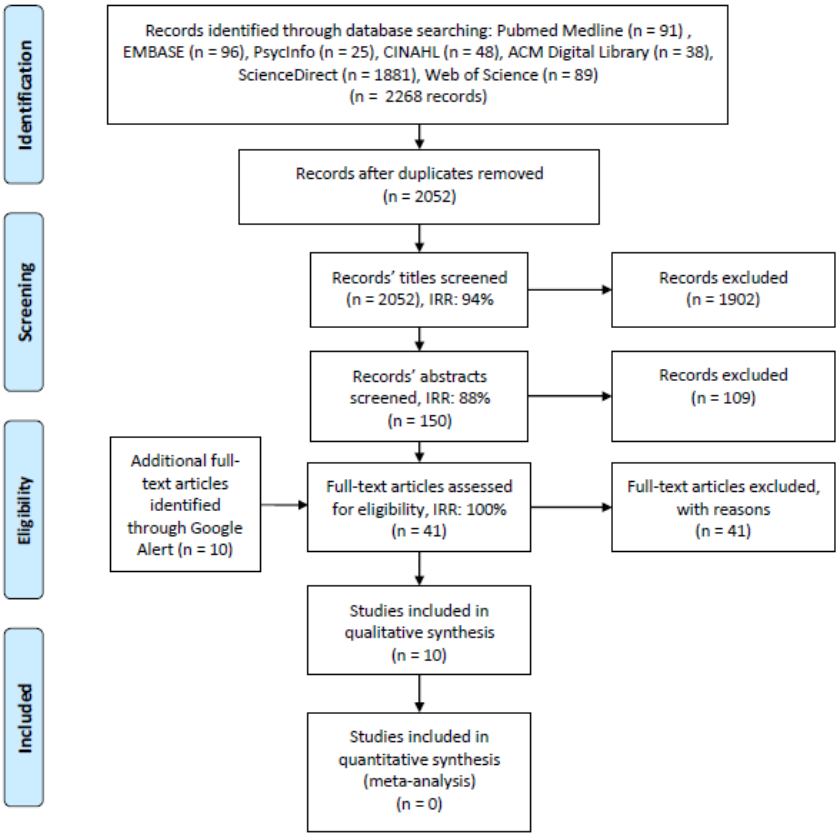
The authors independently familiarized themselves with the assessment tool and rated each study individually. Cohen kappa was calculated to assess interrater reliability between the two assessments and scored at 79%; the majority of disagreement concerned generalizability and sample size guidelines. Discrepancies were discussed and resolved in consensus. For details on the risk bias tool used and the authors' ratings, see Multimedia Appendix 4.

Results

Selection and Inclusion of Studies

In all, 2052 deduplicated citations from electronic databases were screened (Figure 1). Of these, 1902 papers were excluded during the title and abstract screening processes, respectively, leaving 41 papers eligible for full-text screening. The search was updated at full-text stage by 10 additional papers identified through Google Alerts, making 51 papers eligible for full-text screening. On reading the full texts, 41 papers were found to be ineligible for study inclusion. Ultimately, 10 papers were considered eligible for inclusion into our systematic literature review.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of included studies. Search updates were conducted until April 2020, with no additional papers being identified for inclusion. IRR: interrater reliability.



Characteristics of Included Studies

The full list of included studies can be seen in Table 2. Article publication dates ranged from 2010 to 2020, with 80% (8/10) papers published from 2016 onward. Four studies were conducted in the United States [50-53], 2 in Spain [54,55], and 1 each in Australia [56], Canada [57], United Kingdom [58], and Korea [59]. Most studies were quasi-experimental and involved users testing and evaluating the conversational agents [50,51,54,56-59]. Two studies were randomized controlled trials (RCTs) [52,53], and 1 was a proof-of-concept study [55].

Of the 10 studies, 4 aimed to design, develop, or evaluate a prototype conversational agent [50,51,58,59]. One study aimed to develop and implement a prototype architecture of a conversational agent [55]. Three studies aimed to only evaluate a specific conversational agent [52,53,56], and 1 study aimed to design, implement, and evaluate a specific conversational agent [57]. One study aimed to design and develop a domain-independent framework for the development of conversational agents and evaluate a corresponding prototype [54].

Three of 10 studies did not report on the sources of funding [54,56,57]. Seven studies reported no conflict of interest [50,51,54,55,57-59]. Two studies disclosed a relevant conflict of interest (see Multimedia Appendix 3) [52,53], and 1 study did not report upon conflict of interests [56].

Table 2. Overview and characteristics of included studies.

Study ID, study location, study design	Study aim	Main reported outcomes and findings	Type and number of study participants	Chronic condition addressed	Type of final target interaction recipient	Health/ application goal
Ferguson et al (2010), US, quasi-experimental	Design and development of prototype system	Prototype development for data collection, sufficient user engagement, development of working end-to-end spoken dialogue system for heart failure check-up	Heart failure patients (focus group: n= 9; survey: n=63)	Heart failure	Patients	Self-care support
Rhee et al (2014), US, quasi-experimental	Design and development of prototype system	High response rate for daily messages of adolescents (81%-97%), symptoms most common topic in adolescent-initiated messages, improvement of symptom and trigger awareness, promoted treatment adherence and sense of control, facilitation of adolescent-parent partnership	Adolescent asthma patient-parent dyads (n=15)	Asthma	Patient-parent dyads	Self-management tool
Griol and Callejas (2016), Spain, quasi-experimental	Design, development, and evaluation of domain-independent framework	Patient feedback: satisfactory system interaction, preference for multimodal interaction due to flexibility; caregiver feedback: positive assessment, perceived potential to stimulate cognitive abilities of patients	Alzheimer patients (n=25) and caregivers (n=6)	Alzheimer	Patients	Disease monitoring
Ireland et al (2016), Australia, quasi-experimental	Evaluation of chatbot	Positive overall impression, technical issues with speed of processing	Community members (n=33)	Parkinson/dementia	Patients	General conversation with Parkinson patients and facilitation of assessments; future: speech and communication therapy for patients
Fitzpatrick et al (2017), US, RCT[a]	Evaluation of fully automated conversational agent	Chatbot interaction significantly reduced depression and associated with high level of engagement and viewed as more favorable than information-only control comparison	Students (n=70)	Depression/anxiety	NA[b]	CBT[c]
Fulmer et al (2018), US, RCT	Evaluation of fully automated conversational agent	2 weeks of chatbot interaction with daily check-ins significantly reduced symptoms of depression, 4 weeks of chatbot interaction reduced symptoms of anxiety more than 2 weeks of chatbot interaction, chatbot interaction led to higher engagement and higher overall satisfaction than control intervention	Students (n=74)	Depression/anxiety	NA	Health support via different interventions such as CBT, mindfulness-based therapy
Easton et al (2019), UK, quasi-experimental	Co-design of prototype and acceptability assessment	Specification of 4 distinct self-management scenarios for patient support, positive engagement, AI[d]-based speech recognition did not work sufficiently - replacement with human wizard for video-based scenario testing	Co-design: COPD[e] patients (n=6), health professionals (n=5), video-based scenario testing: COPD patients (n=12)	COPD	Patients	Self-management tool

Study ID, study location, study design	Study aim	Main reported outcomes and findings	Type and number of study participants	Chronic condition addressed	Type of final target interaction recipient	Health/ application goal
Rose-Davis et al (2019), Canada, quasi-experimental	Design, implementation, and evaluation of prototype dialogue system	Implementation of AI-based extended model of argument into conversational agent prototype for delivering patient education, satisfactory feedback	Clinicians (n=6)	JIA[f]	Parents of patients	Patient education
Roca et al (2020), Spain, proof of concept	Development and prototype architecture implementation of chatbot	Development of prototype chatbot architecture based on microservices through the use of messaging platforms	Health care professionals (n=NA)	Variety of chronic diseases, specific example of psoriasis	Patients	Disease monitoring
Rehman et al (2020), Korea, quasi-experimental	Design, development, and evaluation of prototype chatbot	Algorithm performance: accuracy: 89%, precision: 90%, sensitivity: 89.9%, specificity: 94.9%, F-measure: 89.9%, good results in all user experience aspects, efficient disease prediction based on chief complaints	Students (n=33)	Diabetes, glaucoma	Patients	Disease diagnosis

[a]RCT: randomized controlled trial.

[b]NA: not available.

[c]CBT: cognitive behavioral therapy.

[d]AI: artificial intelligence.

[e]COPD: chronic obstructive pulmonary disease.

[f]JIA: juvenile idiopathic arthritis.

Evaluation Measures and Main Findings

Two studies assessed the technical performance of the conversational agents and reported consistently high performance measures of the conversational agent (accuracy: 89%; precision: 90%; sensitivity: 89.9%; specificity: 94.9%; F-measure: 89.9%) [59] as well as high message response rates (81% to 97%) [51].

In 7 studies, user experience was assessed. User experience was generally positive regarding the acceptability, understanding of the conversational agents, comprehensibility of the systems' responses, interaction rates, or content relevance [51,53,54,56-59].

Two RCTs reported on health-related outcomes and found that interaction with the conversational agents led to decreased symptoms of depression and anxiety compared with the control groups [52,53].

Four studies found high levels of engagement with the conversational agent or reported the conversational agent to be engaging [50,52,53,58]. One study found that the conversational agent improved awareness of disease symptoms and triggered and promoted treatment adherence [51].

One study reported that the developed conversational agent architecture was able to provide telemonitoring for chronic diseases [55]. The same study further received feedback of health professionals that the architecture provides a flexible solution for personalized monitoring services and data storage [55].

Health Care Characteristics

In the reviewed articles, psychological conditions were the most commonly addressed type of condition, which was the focus of 3 studies [52-54]. Other types of chronic conditions included respiratory [51,58], cardiovascular [50], nervous system [56], rheumatic [57], and autoimmune/eye conditions [59]. One study addressed various chronic diseases and outlined a specific example of an autoimmune disease [55]. More specifically, the addressed chronic conditions included depression and anxiety [52,53], heart failure [50], asthma [51], Alzheimer disease [54], Parkinson/dementia [56], chronic obstructive pulmonary disease (COPD) [58], juvenile idiopathic arthritis (JIA) [57], and diabetes/glaucoma [59]. One study addressed a variety of chronic diseases and delineated psoriasis as a specific example [55].

In 3 papers, students served as main study participants [52,53,59]. Disease-specific patients were involved in 3 studies [50,54,58]. Other types of study participants included patients' parents [51], caregivers [54], clinicians [57], health professionals [55,58], and community members [56].

Patients were the most common final targeted interaction recipients [50,54-56,58,59]. One study targeted the interaction for the use with patient-parent dyads [51], whereas 1 other study specifically targeted patients' parents [57]. Two studies did not provide further information on the targeted interaction recipients [52,53].

Self-care and self-management were the main health goals of the conversational agents in 3 studies [50,51,58], whereas 2 study agents were sought to assist in disease monitoring [54,55].

Other study health goals included general conversations with patients [56], cognitive behavioral therapy [52], patient education [57], and disease diagnosis [59]. One study reported health support via different interventions such as cognitive behavioral or mindfulness-based therapy [53].

Of the 10 studies, 2 aimed at further human involvement besides the targeted interaction recipients. One study additionally involved patients' parents as well as a certified asthma expert [51], and another study involved patients' caregivers [57].

Characteristics of Conversational Agents

Conversational agents were mostly used for data collection [50,54], coaching [52,53], diagnosis [55,59], and support [51,58] (see Table 3 for overview and characteristics of the conversational agents reported in the included studies). Education was the goal of one conversational agent [57] whereas another agent is currently built for data collection but it was anticipated that it may also have an educational and support purpose in future [56].

Different communication channels were used across the identified conversational agents. While two conversational agents use a smartphone app as their main communication channel [54,56], one study reports the general use of the mobile phone [51]. One agent uses a platform agnostic smartphone and desktop instant messenger app [52], and another agent uses a platform-specific application for Android and is usable on any smart Android device such as smartwatch, smartphone, tablet, laptop, and vendor-specific devices that contain a microphone and speaker and support Android [59]. Another agent employs a customizable platform that can be accessed via multiple communication channels such as Facebook, Slack, or short messaging services [53]. One agent uses a web browser as the main communication channel [58], while another agent is designed for communication channels such as messaging platforms or web interfaces [55]. The communication channel of two conversational agents was not specified in the papers [50,57].

The dialogue initiative of 4 conversational agents was held by the user [54,55,57,59], whereas 4 conversational agents used a mixed approach which means that both the user and the system were able to initiate the conversation [50-52,56]. Two studies did not report upon the dialogue initiative [53,58].

A total of 6 studies used a multimodal interaction modality which means that multiple different modalities for input and/or for output were used. Of these, 2 conversational agents require a spoken input format [56,59], whereas 2 other agents allow for both spoken or written input formats [50,58]. One conversational agent uses a written or a visual input format [55], and 1 study employs spoken, written, visual as well as external content from a smartphone sensor as an input format [54]. Regarding the output formats of the multimodal agents, 2 agents use spoken and written output formats [50,56]. One conversational agent uses only a written output format [55], whereas 1 agent employs a written or a visual output format [59]. One agent uses a spoken, written, or a visual output format [54], while 1 study did not report upon the output format used [58]. The remaining 4 studies use a written format of interaction modality, which means that both input and output were in a written form [51-53,57].

Most of the conversational agents we identified were still in a prototype stage and were not publicly available [50,51,54,55,57-59]. Two conversational agents were commercially available [52,53], and 1 was available for free on Android app store [56].

Table 3. Overview and characteristics of the conversational agents reported in the included studies.

Study ID	Conversational agent name	Conversational agent goal	Interaction modality (input/output format)	Availability of conversational agent	AI[a] techniques	AI system development
Ferguson et al (2010)	Personal health management assistant	Data collection	Multimodal (s[b] or w[c]/s or w)	NA[d] (prototype)	Speech recognition, NLP[e]	Internal
Rhee et al (2014)	mASMAA (mobile phone-based asthma self-management aid for adolescents)	Support	Written	NA (prototype)	NLP	Internal
Griol and Callejas (2016)	NA (application, conversational agent)	Data collection	Multimodal (s, w, v[f], external sensors/s, w, v)	NA (prototype)	NN[g], ML[h], ASR[i], NLU[j], NLG[k], TTS[l]	External (Google API[m])
Ireland et al (2016)	Harlie (Human and Robot Language Interaction Experiment)	Now: data collection; future: education and support	Multimodal (s/s, w)	For free on Android app store	Speech recognition incl. STT[n] and TTS, NLP, AIML[o]	External (Google API)
Fitzpatrick et al (2017)	Woebot	Coaching	Written	Commercially available	Decision tree, NLP	External (Woebot Labs Inc)
Fulmer et al (2018)	Tess	Coaching	Written	Commercially available	Emotion algorithms, ML, NLP	External (X2AI Inc)
Easton et al (2019)	Avachat (=avatar & chat)/Ava	Support	Multimodal (s, w/NA)	NA (prototype)	Speech recognition	External (Kaldi toolkit)
Rose-Davis et al (2019)	JADE (Juvenile idiopathic Arthritis Dialogue-based Education)	Education	Written	NA (prototype)	NA	Internal
Roca et al (2020)	NA (Virtual Assistant)	Diagnosis	Multimodal (w, v/w)	NA (prototype)	AIML, NLP	NA
Rehman et al (2020)	MIRA (Medical Instructed Real-Time Assistant)	Diagnosis	Multimodal (s/w, v)	NA (prototype)	Speech recognition, NLP, NLU, NN, ML, DL[p]	Internal

[a]AI: artificial intelligence.

[b]s: spoken.

[c]w: written.

[d]NA: not available.

[e]NLP: natural language processing.

[f]v: visual.

[g]NN: neural network.

[h]ML: machine learning.

[i]ASR: automatic speech recognition.

[j]NLU: natural language understanding.

[k]NLG: natural language generation.

[l]TTS: text-to-speech.

[m]API: application programming interface.

[n]STT: speech-to-text.

[o]AIML: artificial intelligence markup language.

[p]DL: deep learning.

Artificial Intelligence Characteristics

Natural language processing represented the most used technique [50-53,55,56,59] before speech recognition (including speech-to-text and text-to-speech) [50,54,56,58,59], machine learning [53,54,59], natural language understanding [54,59], neural networks [54,59] and artificial intelligence markup language [56,57], as shown in Table 3. The following techniques were used in one study each: deep learning [59], natural language generation [54], emotion algorithms [53], and decision trees [52]. One study used AI-based argument theory for modeling its dialogue system [57]. Additional details regarding

the artificial intelligence architecture can be found in Multimedia Appendix 3.

A total of 4 studies developed the artificial intelligence system internally [50,51,57,59], and 5 studies relied on external sources [52-54,56,58]. Of the studies using external artificial intelligence systems for speech recognition (including text-to-speech and speech-to-text), 2 studies used an external Google application programming interface [54,56], and 1 study used the open-source Kaldi toolkit [58]. One study relied on the existing The Rochester Interactive Planning System natural dialogue system [51], and 1 study did not report upon the artificial intelligence system development [55].

Artificial intelligence categorization varied in its terminology across the studies. Four studies were classified as AI [53,56-58]. Other categorizations were natural interaction [50], state-of-the-art natural language understanding technology [51], fully automated [52], smart [55] and state of the art real-time assistant [59]. One study did not provide an explicit categorization [54].

Discussion

Principal Findings

Our systematic literature review identified 10 studies, of which 2 were RCTs and the majority were quasi-experimental studies. This is, to our knowledge, the only systematic literature review focusing specifically on AI-based conversational agents used in the context of health care for chronic diseases. Other recent reviews focused on conversational agents for either a specific health condition such as mental health [44], the general application of chatbots in health care [1], or specific features thereof such as personalization [45] or technical architectures [11].

A total of 80% of the papers that we identified were published relatively recently, from 2016 onward. Together with the small number of identified studies, this shows the immaturity of the field of AI-based conversational agents for chronic diseases. This finding is coherent with other recent reviews which found the general application of conversational agents in health care to be at a nascent but developing stage [1,11,45]. Most of the AI-based conversational agents we identified were still in a prototype stage and not publicly available. They are used for data collection, coaching, diagnosis, support, and education of patients suffering from chronic diseases.

Recent advances in AI software allow an increasing number of conversational agents to offer natural interactions between humans and their machine agent counterparts [11,12]. However, drawbacks such as biased and opaque decision-making leading to limited trust in the final outcomes still exist and are only partially solved [60]. Combined with the functional difficulty of needing large datasets for algorithmic training, this could explain the overall small number of existing applications [61].

The current chatbots operate on a variety of communication channels, out of which some are vendor specific such as tailored for Android devices. We advise future studies to keep track of such platform-dependent developments as it could point to a stronger influence of or dependence on technology providers regarding health care–related applications.

The identified research was not truly geographically diverse; 50% of studies were conducted in North America, only one each in Australia and an Asian country, and the remaining 30% in Europe. There was not a single study conducted in Africa. Additionally, 90% of these research locations are embedded in Western cultures, exerting a strong bias on the generalizability of their results. Given the worldwide prevalence of chronic conditions [37] and the need to apply health care system-specific solutions [62], future research should strive to include diverse geographies to ensure context-specific relevance. We advise to extend research foci beyond the Western socioeconomic cultural context and additionally include emerging economies such as India and China to increase variability and generalizability.

The majority of the identified studies aimed at fully designing, developing, or evaluating a conversational agent specific for only one chronic condition. This finding suggests that AI-based conversational agents evolve into providing tailored support for specific chronic conditions rather than general interventions applicable to a broad range of chronic diseases. Future research could investigate the effects of such specialization on treatment-related measures such as patient satisfaction or treatment adherence.

The evaluation measures of the identified AI-based conversational agents and their effects on the targeted chronic conditions were broad and not unified. The most commonly reported measurements were user experience and chatbot engagement, which are generalistic usability measurements for technical systems [63]. Only 2 studies assessed the technical performance of the conversational agents and 2 other studies reported on the health-related outcomes. Generally, however, the measured and reported results were positive and indicated both high overall performance and satisfactory user experience, high engagement, and positive health-related outcomes. Future research could enforce following standard guidelines for research in the health care area such as the Consolidated Standards of Reporting Trials of electronic and mobile health apps and online telehealth (CONSORT-EHEALTH) [64], the mobile health evidence reporting and assessment (mERA) checklist [65], or the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement [66] to increase quality and comparability of studies. The primarily quasi-experimental nature and subsequent inconsistency of evaluated measures of the found literature could explain the lack of use of such reporting guidelines at present.

Our review shows that current AI-based conversational agents address a broad variety of chronic diseases, categorized as chronic respiratory, cardiovascular, nervous system-related, rheumatic, autoimmune-related, eye-related, and psychological conditions. While it is informative to have such a wide investigation of different disease types, this variation complicates the comparability within and between conditions. Future research could aim at first developing and evaluating within-chronic disease-related differences of AI-based conversational agents (eg, individual chatbots for asthma, COPD, and sleep apnea as examples of chronic respiratory

diseases) before extending their scope of research to between-chronic disease-related comparisons (eg, respiratory vs cardiovascular chronic conditions).

Following such a research agenda could lead to the development of more consistent studies with higher standards and increased validity of reported findings. Similar considerations concern the large variety of reported health goals; while self-care management is the main health goal of 30% of existing AI-based conversational agents for chronic conditions before offering assistance of disease monitoring, the remaining 70% address intervention goals such as general conversation, therapy, education, and diagnosis. This inconsistency presents another complication of the comparability of the existing chatbots.

Of the studies investigated, 70% were quasi-experimental, 20% RCTs, and the remaining 10% proof-of-concept. Such quasi-experimental studies are typically cross-sectional, nonrandomized, and describe the first impression of a single instant [67]. For a better understanding of the real-world effects of AI-based conversational agents on health care for chronic diseases, future research should aim at conducting field experiments, which in the best case are designed as longitudinal experimentations in order to investigate long-term effects. This is especially important when considering the time span of chronic diseases; they typically affect patients for at least 12 months but can prevail for a significantly longer period of a patient's life span [39].

It is further noteworthy to point out that the only 2 RCTs of this review mentioned a commercial interest in the investigated conversational agent by at least one of the authors. We would encourage future research to assess commercially available conversational agents without similar business connections in order to enrich the chatbots' evaluation by a purely external point of view.

While it is not unexpected to find that patients were the majority of targeted intervention partners, it is somewhat surprising to see that only 2 conversational agents further included additional social contacts of patients, here the patients' parents. We want to highlight that chronic diseases often heavily affect the immediate and wider social context of the affected patient [61]. Future interventions could consider additional human involvement in order to better recognize the social effect of chronic diseases. This could further maximize treatment adherence and health outcomes, two important treatment goals [68].

Natural language processing technology is the most widely applied AI technique and outnumbers related further used techniques such as speech recognition, text-to-speech, and speech-to-text, natural language understanding, and natural language generation. Other prominent AI techniques such as deep learning, machine learning, neural networks, and decision trees are also used, but to a much smaller extent. This finding might be explained through the already mentioned prevalence of multimodal interaction approaches of the reported conversational agents, giving supremacy to the development and evaluation of communication-focused AI techniques. Currently, ongoing developments in the area of natural communication between conversational agents and humans increasingly address natural language generation and emotion recognition [69,70]. These advancements are expected to lead to AI-based conversational agents that converse even more naturally with patients than currently possible. This could have a plethora of effects on the relationship between patients and chatbots as well as on treatment-related outcomes and thus presents a relevant area for future research.

One potential danger of such presumably naturally conversing chatbots is harm or even death of the patient in case the chatbot's recommendations are inaccurate or wrong, especially when the advice concerns critical decisions such as changes or mix of medication [71]. Patients, who are often laypeople when it comes to assessing any technical or medical capabilities of AI-based conversational agents, might follow a chatbot's advice without additional medical clarification [71]. Future chatbot development and corresponding research should put an increased focus on addressing such shortcomings and threats in order to maximally ensure patient safety.

Except for the 2 studies developing and evaluating conversational agent architectures, the heterogeneity and general lack of depth of reported AI techniques and systems is a relevant point to consider. Even though all 10 studies explicitly state to apply AI-based systems, the lack of technical information critically hinders replicability and poses questions about the quality of reported findings. Such dearth of detail reinforces the application roadblocks of AI-based systems—opaque and biased decision-making processes and resulting lack of trust [60]. In addition, it hinders the development of a generic system architecture, which could be used as an informative framework for the development and structure of AI-based chatbots in the context of health care for chronic diseases. We strongly advise future researchers to report all necessary technical features required to replicate study results and further (partially or exemplarily) allow access to the developed AI-based conversational systems. In addition to the above-mentioned standardized guidelines for research in health care, future research should make use of already existing guidelines for reporting the technical part of AI-based conversational agents used in health care and medicine [72,73]. More generalized checklists aimed at assessing the overall structure of AI-related medical research such as the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) could be also consulted; they offer guidance on which specific information should be reported on the chosen AI model and its subsequent training, evaluation, and performance [74]. We further recommend future research to synthesize a generic system architecture and derive a framework for AI-based chatbots in the context of health care for chronic diseases once the field has progressed and more standardized data are available.

Half of the studies in our review made use of external systems for the development of (parts of) their AI architecture, which could indicate a trend of external and open access–based software development for AI-based health care conversational agents. Future research should pay attention to this in order to further shed light on this approach.

A final point to consider is the inconsistent taxonomy of AI-based software; while 4 studies clearly labeled their software

as AI, there was a broad variety of otherwise used terms such as natural interaction, state-of-the-art, smart, or fully automated. The inconsistent use of terms aggravates the use of a common terminology. We see value in the development and use of clear terms for the sake of clarity and comparability of future research.

Strengths and Limitations

This systematic literature review has several strengths as well as some limitations. It was conducted and reported according to the standardized PRISMA guidelines [46]. We conducted an extensive literature search by accessing 7 databases and deploying a thorough and comprehensive search strategy. In addition, we reviewed reference lists of relevant studies and used several Google alerts containing combinations of the search terms from November 2019 until April 2020 for identifying further papers not identified through the initial database searches.

We prioritized sensitivity over specificity with our search strategy in order to avoid missing important studies and construct a holistic view of AI-based conversational agents for health care for chronic diseases. We objectively defined the study eligibility criteria. Given the novelty of the search field, however, many search results were published conference abstracts that had to be omitted given the study eligibility criteria.

Study selection, title and abstract screening, full text screening, and data extraction were done independently by two reviewers. We checked for interrater reliability at several steps in the selection process and Cohen kappa showed substantial agreement per step.

We applied a narrative approach for reviewing the included studies. Intense team discussions concerned the classification of reported AI architectures. We decided in consensus to follow the proposed taxonomy of Montenegro et al [11]. However, the final study selection might still omit relevant AI-based conversational agents if a different taxonomy for study selection were applied.

Key limitations of this review are the heterogeneity and relatively small number of the included studies as well as the prevalence of quasi-experimental studies. This underlines the complexity and novelty of the searched field, and we thus did not conduct a meta-analysis.

Finally, risk of bias varied extensively between the included studies, reducing the reliability of findings in studies with high risk of bias. This reduced the trust we could place in the reported findings of studies with high risk of bias.

Conclusions

Technological advances facilitate the increasing use of AI-based conversational agents in health care settings. So far, this evolving field of research has a limited number of applications tailored for chronic conditions, despite their medical prevalence and economic burden to the health care systems of the 21st century. Existing applications reported in literature lack evidence-based evaluation and comparison within as well as between different chronic health conditions. Future research should focus on adhering to evaluation and reporting guidelines for technical aspects such as the underlying AI architecture as well as overall solution assessment.

Acknowledgments

We are grateful to Mr Julian Ventouris for his assistance with the study search process and Ms Grace Xiao for proofreading the document. This study is supported by the National Research Foundation, Prime Minister’s Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

Authors' Contributions

TS was responsible for the study design; search strategy; screening; data extraction and analysis; and first draft, revisions, and final draft of the manuscript. RK was responsible for screening, data extraction, and first draft of the manuscript. FW was responsible for the critical revision of the first draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study protocol.
[PDF File (Adobe PDF File), 158 KB - jmir_v22i9e20701_app1.pdf]

Multimedia Appendix 2
Search terms per construct.
[PDF File (Adobe PDF File), 13 KB - jmir_v22i9e20701_app2.pdf]

Multimedia Appendix 3
Overview and characteristics of included studies and conversational agents.
[PDF File (Adobe PDF File), 194 KB - jmir_v22i9e20701_app3.pdf]

Multimedia Appendix 4
The risk of bias tool (based upon the Consolidated Standards of Reporting Trials checklist and adapted from Maher et al [2014]).
[PDF File (Adobe PDF File), 143 KB - jmir_v22i9e20701_app4.pdf]

References

1. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
2. Tsai C, Wu J. Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Syst Appl Pergamon 2008 May;34(4):2639-2649. [doi: 10.1016/j.eswa.2007.05.019]
3. Davenport T, Guha A, Grewal D, Bressgott T. How artificial intelligence will change the future of marketing. J Acad Mark Sci 2019 Oct 10;48(1):24-42. [doi: 10.1007/s11747-019-00696-0]
4. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Dec;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
5. Russell S, Peter N. Artificial Intelligence: A Modern Approach. Prentice Hall 1995. [doi: 10.1016/0925-2312(95)90020-9]
6. Kaplan A, Haenlein M. Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 2019 Jan;62(1):15-25 [FREE Full text] [doi: 10.1016/j.bushor.2018.08.004]
7. De Bruyn A, Viswanathan V, Beh Y, Brock J, von Wangenheim F. Artificial intelligence and marketing: pitfalls and oppportunities. J Interact Mark 2020 Aug;51:91-105 [FREE Full text] [doi: 10.1016/j.intmar.2020.04.007]
8. Searle JR. Minds, brains, and programs. Behav Brain Sci 2010 Feb 04;3(3):417-424. [doi: 10.1017/s0140525x00005756]
9. Arkoudas K, Bringsjord S. Philosophical foundations. In: Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press; 2014:34-63.
10. Bostrom N, Yudkowsky E. The ethics of artificial intelligence. In: Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press; 2014:316-334.
11. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
12. Suta P, Lan X, Wu B, Mongkolnam P, Chan J. An overview of machine learning in chatbots. Int J Mech Engineer Robotics Res 2020;9(4):502-510. [doi: 10.18178/ijmerr.9.4.502-510]
13. Gentsch P. A bluffer's guide to AI, algorithmics and big data. In: AI in Marketing, Sales, and Service. Cham: Palgrave Macmillan; 2019:11-24.
14. McTear MF. Spoken dialogue technology: enabling the conversational user interface. ACM Comput Surv 2002;34(1):90-169. [doi: 10.1145/505282.505285]
15. Radziwill N, Benton M. Evaluating quality of chatbots and intelligent conversational agents. 2017. URL: https://arxiv.org/pdf/1704.04579 [accessed 2020-08-30]
16. Masche J, Le N. A review of technologies for conversational systems. In: Le NT, van Do T, Nguyen N, Thi H, editors. Advanced Computational Methods for Knowledge Engineering. Cham: Springer; 2017:212-225.
17. Cui L, Huang S, Wei F, Tan C, Duan C, Zhou M. SuperAgent: a customer service chatbot for e-commerce websites. URL: https://www.aclweb.org/anthology/P17-4017.pdf [accessed 2020-08-30]
18. Ivanov S, Webster C. Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies? A cost-benefit analysis. 2017 Presented at: Prepared for the International Scientific Conference "Contemporary Tourism: Traditions and Innovations"; 2017; Sofia. [doi: 10.1108/978-1-78756-687-320191002]
19. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
20. Weizenbaum J. ELIZA: a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36-45. [doi: 10.1145/365153.365168]
21. Colby KM. Artificial Paranoia: A Computer Simulation of Paranoid Processes. Oxford: Pergamon Press; Jan 1976.
22. Saygin A, Cicekli I, Akman V. Turing test: 50 years later. Minds Mach 2000;10(4):463-518. [doi: 10.1007/978-94-010-0105-2_2]
23. Epstein J, Klinkenberg W. From Eliza to Internet: a brief history of computerized assessment. Comput Human Behav 2001 May;17(3):295-314. [doi: 10.1016/s0747-5632(01)00004-8]
24. Overby MA. Psyxpert: an expert system prototype for aiding psychiatrists in the diagnosis of psychotic disorders. Comput Biol Med 1987;17(6):383-393. [doi: 10.1016/0010-4825(87)90056-4] [Medline: 3319379]
25. Levy M, Ferrand P, Chirat V. SESAM-DIABETE, an expert system for insulin-requiring diabetic patient education. Comput Biomed Res 1989 Oct;22(5):442-453. [doi: 10.1016/0010-4809(89)90037-2] [Medline: 2776447]
26. von Krogh G. Artificial intelligence in organizations: new opportunities for phenomenon-based theorizing. Acad Manag Discov 2018;4(4):404-409. [doi: 10.5465/amd.2018.0084]
27. Fadhil A. A conversational interface to improve medication adherence: towards AI support in patient's treatment. 2018. URL: https://arxiv.org/pdf/1803.09844 [accessed 2020-08-30]

28. Fadhil A. Beyond patient monitoring: conversational agents role in telemedicine & healthcare support for home-living elderly individuals. 2018. URL: https://arxiv.org/pdf/1803.06000 [accessed 2020-08-30]
29. Pereira J, Díaz O. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
30. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst 2019 Jun 25;43(8):246. [doi: 10.1007/s10916-019-1387-1] [Medline: 31240494]
31. Bickmore T, Pusateri A, Kimani E, Paasche-Orlow M, Trinh H, Magnani J. Managing chronic conditions with a smartphone-based conversational virtual agent. Proc 18th Int Conf Intell Virtual Agents 2018:119-124. [doi: 10.1145/3267851.3267908]
32. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
33. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
34. Denecke K, Hochreutener SL, Pöpel A, May R. Self-anamnesis with a conversational user interface: concept and usability study. Methods Inf Med 2018 Nov;57(5-06):243-252. [doi: 10.1055/s-0038-1675822] [Medline: 30875703]
35. Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. Methods Inf Med 2019 Jun;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
36. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
37. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar 10;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
38. Yach D, Hawkes C, Gould CL, Hofman KJ. The global burden of chronic diseases: overcoming impediments to prevention and control. JAMA 2004 Jun 02;291(21):2616-2622. [doi: 10.1001/jama.291.21.2616] [Medline: 15173153]
39. Hvidberg MF, Johnsen SP, Glümer C, Petersen KD, Olesen AV, Ehlers L. Catalog of 199 register-based definitions of chronic conditions. Scand J Public Health 2016 Jul;44(5):462-479 [FREE Full text] [doi: 10.1177/1403494816641553] [Medline: 27098981]
40. Paez KA, Zhao L, Hwang W. Rising out-of-pocket spending for chronic conditions: a ten-year trend. Health Aff (Millwood) 2009;28(1):15-25. [doi: 10.1377/hlthaff.28.1.15] [Medline: 19124848]
41. Stein RE, Bauman LJ, Westbrook LE, Coupey SM, Ireys HT. Framework for identifying children who have chronic conditions: the case for a new definition. J Pediatr 1993 Mar;122(3):342-347. [doi: 10.1016/s0022-3476(05)83414-6] [Medline: 8441085]
42. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [Medline: 12435261]
43. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 04;8:CD011682. [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
44. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
45. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
46. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;349:g7647 [FREE Full text] [Medline: 25555855]
47. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
48. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
49. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 [FREE Full text] [Medline: 20332509]
50. Ferguson G, Quinn J, Horwitz C, Swift M, Allen J, Galescu L. Towards a personal health management assistant. J Biomed Inform 2010 Oct;43(5 Suppl):S13-S16 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.014] [Medline: 20937478]

51. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient Prefer Adherence 2014;8:63-72 [FREE Full text] [doi: 10.2147/PPA.S53504] [Medline: 24470755]
52. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
53. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
54. Griol D, Callejas Z. Mobile conversational agents for context-aware care applications. Cogn Comput 2015 Aug 21;8(2):336-356. [doi: 10.1007/s12559-015-9352-x]
55. Roca S, Sancho J, García J, Alesanco A. Microservice chatbot architecture for chronic patient support. J Biomed Inform 2020 Feb;102:103305. [doi: 10.1016/j.jbi.2019.103305] [Medline: 31622802]
56. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello Harlie: enabling speech monitoring through chat-bot conversations. Stud Health Technol Inform 2016;227:55-60. [Medline: 27440289]
57. Rose-Davis B, Van Woensel W, Stringer E, Abidi S, Abidi SSR. Using an artificial intelligence-based argument theory to generate automated patient education dialogues for families of children with juvenile idiopathic arthritis. Stud Health Technol Inform 2019 Aug 21;264:1337-1341. [doi: 10.3233/SHTI190444] [Medline: 31438143]
58. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. J Med Internet Res 2019 May 30;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
59. Rehman UU, Chang DJ, Jung Y, Akhtar U, Razzaq MA, Lee S. Medical instructed real-time assistant for patient with glaucoma and diabetic conditions. Appl Sci 2020 Mar 25;10(7):2216. [doi: 10.3390/app10072216]
60. Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational decision-making structures in the age of artificial intelligence. Calif Manag Rev 2019 Jul 13;61(4):66-83. [doi: 10.1177/0008125619862257]
61. Faraj S, Pachidi S, Sayegh K. Working and organizing in the age of the learning algorithm. Inf Organ 2018 Mar;28(1):62-70. [doi: 10.1016/j.infoandorg.2018.02.005]
62. Amrita A, Biswas D. Health care social media: expectations of users in a developing country. Med 2.0 2013;2(2):e4 [FREE Full text] [doi: 10.2196/med20.2720] [Medline: 25075239]
63. Tullis T, Albert B. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Oxford: Newnes; 2013.
64. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
65. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]
66. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [Medline: 14998794]
67. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23 [FREE Full text] [doi: 10.1197/jamia.M1749] [Medline: 16221933]
68. Klok T, Kaptein AA, Brand PLP. Non-adherence in children with asthma reviewed: the need for improvement of asthma care and medical education. Pediatr Allergy Immunol 2015 May;26(3):197-205. [doi: 10.1111/pai.12362] [Medline: 25704083]
69. Oh K, Lee D, Ko B, Choi H. A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysissentence generation. 2017 Presented at: 18th IEEE International Conference on Mobile Data Management (MDM); 2017; Daejeon p. 371-375. [doi: 10.1109/mdm.2017.64]
70. Montenegro J, Da Costa C, Righi R, Roehrs A, Farias E. A proposal for postpartum support based on natural language generation model. 2018 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); 2018; Las Vegas p. 756-756. [doi: 10.1109/csci46756.2018.00151]
71. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Dec 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
72. Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. AJR Am J Roentgenol 2019 Jan;212(1):38-43. [doi: 10.2214/AJR.18.20224] [Medline: 30332290]
73. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 2018 Dec;286(3):800-809. [doi: 10.1148/radiol.2017171920] [Medline: 29309734]

74. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020 Mar 01;2(2):e200029. [doi: 10.1148/ryai.2020200029]

Abbreviations

AI: artificial intelligence
AGI: artificial general intelligence
CLAIM: Checklist for Artificial Intelligence in Medical Imaging
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of electronic and mobile health apps and online telehealth
COPD: chronic obstructive pulmonary disease
JIA: juvenile idiopathic arthritis
mERA: mobile health evidence reporting and assessment
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs

Edited by G Eysenbach; submitted 26.05.20; peer-reviewed by W Zhang, KL Ong; comments to author 14.07.20; revised version received 15.07.20; accepted 26.07.20; published 14.09.20.

Please cite as:
Schachner T, Keller R, v Wangenheim F
Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review
J Med Internet Res 2020;22(9):e20701
URL: http://www.jmir.org/2020/9/e20701/
doi:10.2196/20701
PMID:32924957

Review

Effectiveness of Virtual Reality in Nursing Education: Meta-Analysis

Feng-Qin Chen[1*], MA; Yu-Fei Leng[1*], MA; Jian-Feng Ge[1], MA; Dan-Wen Wang[1], PhD; Cheng Li[1], MA; Bin Chen[1], MA; Zhi-Ling Sun[1], PhD

School of Nursing, Nanjing University of Chinese Medicine, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Zhi-Ling Sun, PhD
School of Nursing
Nanjing University of Chinese Medicine
138 Xian Lin Road
Nanjing
China
Phone: 86 13813892093
Email: szl@njucm.edu.cn

Abstract

Background: Virtual reality (VR) is the use of computer technology to create an interactive three-dimensional (3D) world, which gives users a sense of spatial presence. In nursing education, VR has been used to help optimize teaching and learning processes.

Objective: The purpose of this study was to evaluate the effectiveness of VR in nursing education in the areas of knowledge, skills, satisfaction, confidence, and performance time.

Methods: We conducted a meta-analysis of the effectiveness of VR in nursing education based on the Cochrane methodology. An electronic literature search using the Cochrane Library, Web of Science, PubMed, Embase, and CINAHL (Cumulative Index to Nursing and Allied Health Literature), up to December 2019 was conducted to identify studies that reported the effectiveness of VR on knowledge, skills, satisfaction, confidence, and performance time. The study selection and data extraction were carried out by two independent reviewers. The methodological quality of the selected studies was determined using the Cochrane criteria for risk-of-bias assessment.

Results: A total of 12 studies, including 821 participants, were selected for the final analysis. We found that VR was more effective than the control conditions in improving knowledge (standard mean difference [SMD]=0.58, 95% CI 0.41-0.75, P<.001, I^2=47%). However, there was no difference between VR and the control conditions in skills (SMD=0.01, 95% CI –0.24 to 0.26, P=.93, I^2=37%), satisfaction (SMD=0.01, 95% CI –0.79 to 0.80, P=.99, I^2=86%), confidence (SMD=0.00, 95% CI –0.28 to 0.27, P=.99, I^2=0%), and performance time (SMD=–0.55, 95% CI –2.04 to 0.94, P=.47, I^2=97%).

Conclusions: The results of this study suggest that VR can effectively improve knowledge in nursing education, but it was not more effective than other education methods in areas of skills, satisfaction, confidence, and performance time. Further rigorous studies with a larger sample size are warranted to confirm these results.

(J Med Internet Res 2020;22(9):e18290) doi:10.2196/18290

KEYWORDS

virtual reality; nursing education; meta-analysis

Introduction

With the rapid development of information technology and shortages of nurse workforce, a transformation of nursing education is needed to prepare nursing students for evolving and complex health care environments [1-3]. In US nursing schools, 75,029 qualified applicants for bachelor's degrees and nursing postgraduate courses were rejected in 2018 due to an insufficient number of faculty, clinical sites, classroom space, clinical preceptors, and budget constraints [4].

The ultimate goal of nursing education is to promote the application of theoretical knowledge in clinical practice [5]. However, limited clinical practice time affects the opportunity for students of having clinical experience with real patients [6]. This lack of clinical practice, which prepares students for the real clinical environment, can contribute to nursing procedure errors that compromise the safety of patients [7]. Narrowing the gap between theory and practice during the educational process is necessary, but poses several challenges to nursing educators [8]. In this scenario, to guarantee the quality and safety of nursing education, educators have adopted various teaching strategies including simulation experience for students [9].

Simulation has been shown to be a valuable teaching-learning strategy to support the changing world of nursing education and to help optimize the teaching process [10-12]. As the National Council of State Boards of Nursing stated, simulation is a key component of nursing education [13]. The use of simulation as a nursing education tool is becoming increasingly common, providing students with realistic opportunities to practice skills learned in theory [14]. Through simulation, students have a variety of practical opportunities to repeat clinical scenarios and make immediate decisions and reflections [15].

With the development of simulation technology, the virtual world was discovered—initially used in military and medical science and later, in medical education [16,17]. Virtual reality (VR) is the use of computer technology to create an interactive three-dimensional (3D) world in which users have a sense of spatial presence [18]. It provides a first-person active learning experience through different degrees of immersion, or, in other words, the real perception of the digital world and the ability to interact with objects and/or perform a series of actions in this digital world [19,20]. VR is highly conducive to clinical and procedure-focused training by enabling simulation [21]. VR simulation refers to the use of a variety of immersive, highly visual, 3D characteristics to replicate real-life situations and health care procedures, incorporating physical or other interfaces such as a computer keyboard, a mouse, speech/voice recognition, motion sensors, or haptic devices [22]. Virtual simulation refers to the involvement of real people operating simulated systems via a computer screen (virtual, that is, as the situation is not physical or in real time), and may include surgical simulators used for on-screen procedural training, usually integrated with haptic devices to interact with the system [18]. In general, VR can make simulation become an effective supplemental tool for teaching [22,23].

As VR technology advances and becomes increasingly affordable, nursing education is being transformed [24]. VR has gained increasing attention in the field of nursing education and been used to teach many nursing concepts including leadership, communication, decision-making, critical thinking, inclusivity, health appraisal, and disaster triage [25,26]. The use of VR in simulations allows repetitive, hands-on training to develop cognitive and skill mastery among nursing students, which are usually defined as the measure of participants' understanding of concepts and the ability of a participant to demonstrate a procedure or technique, respectively [8,27]. Additionally, VR simulations can give nursing students the opportunity to practice skills in a safe environment without risk to patients [28]. In a study, 98% of the participating students recommended virtual simulation for future use in nursing education [29].

Although the use of VR has many advantages, some researchers have reported that VR is not more effective than other traditional methods on some outcomes such as knowledge and performance scores [30,31]. There are still some inconsistencies on the effectiveness of VR among studies. Up to date, meta-analyses on the effectiveness of VR have been conducted in some areas of medicine and education [32,33]. However, to the best of our knowledge, there is no meta-analysis evaluating the effectiveness of VR in nursing education. Only one article systematically reviewed and evaluated the effectiveness of VR without meta-analysis, focusing on the effectiveness of VR simulation compared to other simulated methods on clinical psychomotor skills for pre-registration nursing students [34]. Therefore, there is a need to determine the effectiveness of VR in nursing education. The aim of this study was to perform a meta-analysis of the effectiveness of VR use on knowledge (participants' understanding of concepts), skills (ability of participants to demonstrate a procedure or technique), satisfaction (participants' perception with VR learning intervention), confidence (self-confidence in learning content and process), and performance time (time taken on the test task) in nursing education.

Methods

This meta-analysis was conducted according to the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) guidelines [35].

Search Strategy

An electronic literature search was carried out in the Cochrane Library, Web of Science, PubMed, Embase, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) from their inception to December 2019. The search strategies used in PubMed, Embase, and the Cochrane Library are listed in the Multimedia Appendix 1. Slightly modified search strategies were used in the other databases. Additionally, we manually examined reference lists of the selected articles to retrieve other relevant publications. Two investigators conducted searches independently, and EndNote software was used to import and manage selected documents.

Inclusion Criteria

This study included randomized controlled trials (RCTs) or trials employing quasi-experimental randomized design, including those in the form of dissertations and conference papers, based on the PICO (Population–Intervention–Comparison–Outcome) method. In this study, the PICO elements were as follows:

- Population: pre-/post-registration nursing students or nursing staff
- Intervention: all kinds of VR education methods
- Comparison: traditional education methods (including presentations, classes, written instructions, etc) and non-VR simulation methods (including high/low fidelity simulation, mannequin-based simulation, etc)

- Outcomes: knowledge, skills, satisfaction, confidence, and performance time

Data Extraction

Two reviewers (FQC and YFL) independently extracted information based on preset standards, including authors, publication date, nation, sample size, participants type, research project, intervention regimens, and outcomes.

Risk-of-Bias Assessment

Two reviewers (FQC and YFL) assessed the studies' quality independently by referring to the Cochrane Handbook for Systematic Reviews of Interventions [36], which includes 7 domains corresponding to a specific type of bias: (1) random sequence generation (selection bias); (2) allocation concealment (selection bias); (3) blinding of participants and personnel (performance bias); (4) blinding of outcome assessment (detection bias); (5) incomplete outcome data (attrition bias); (6) selective reporting (reporting bias); and (7) other biases. A judgement of "low risk," "high risk," or "unclear risk" of bias was assigned to each domain. When disagreements between reviewers could not be resolved through discussion, two additional reviewers (ZLS and JFG) made the final decision.

Data Synthesis and Analysis

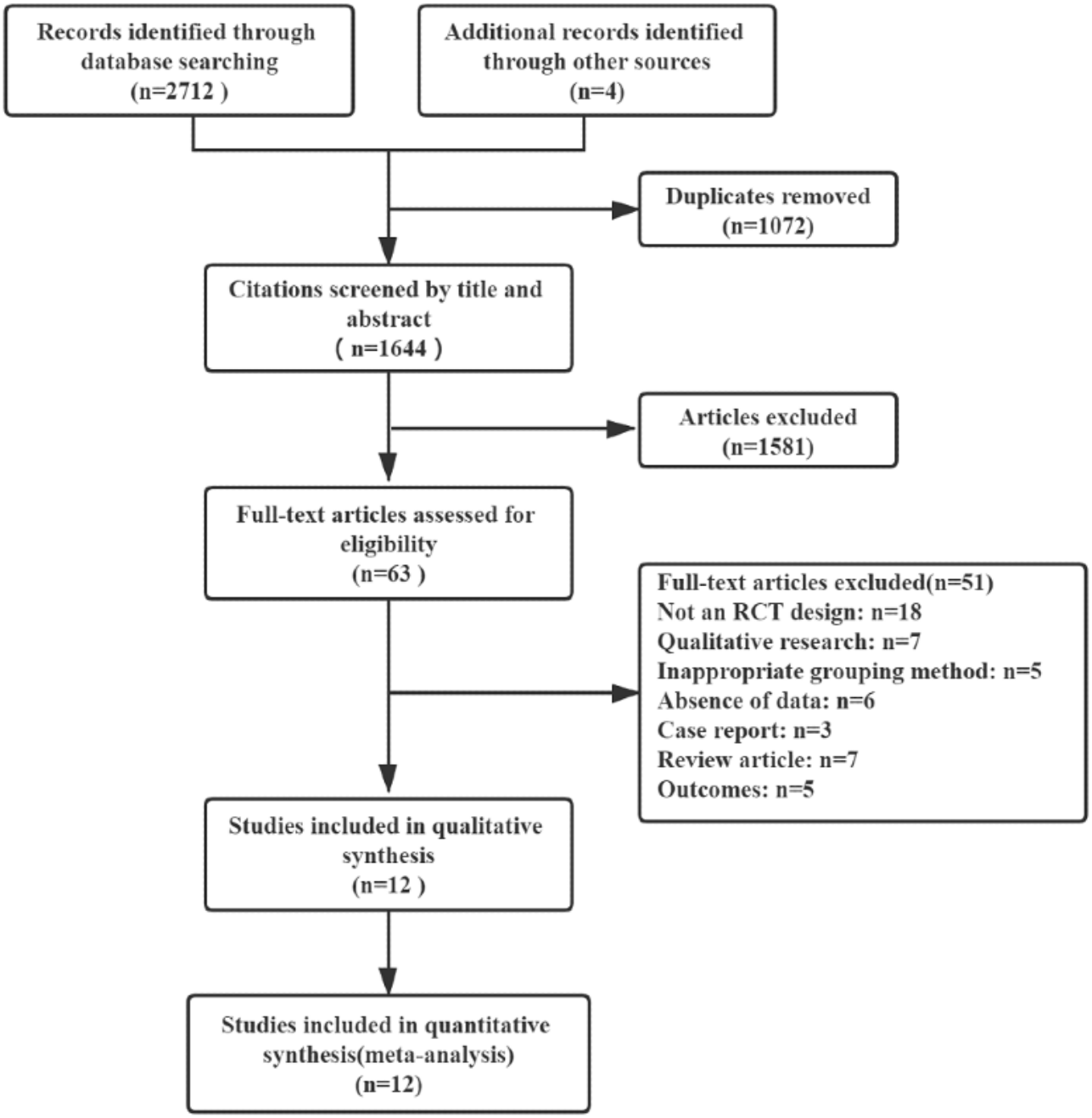
The meta-analysis was conducted using RevMan 5.3 [37], a desktop version of Review Manager software used for Cochrane intervention and flexible reviews. For continuous data, we reported standard mean difference (SMD) with 95% confidence intervals. In each analysis, I^2 was used to measure the statistical heterogeneity among studies. According to the values of P and I^2, the fixed-effect model ($P>.1$, $I^2<50\%$) or random-effects model ($0<P<.1$, $I^2\geq50\%$) were selected [38].

Results

Results of the Literature Search

A total of 2716 potential studies were identified from 5 databases (n=2712) and relevant references (n=4). After removing 1072 duplicates, the remaining articles were reviewed and those that did not meet the inclusion criteria were excluded. A total of 1644 articles were screened by title and abstract, of which 1581 articles were excluded. A total of 63 full-text articles were downloaded and assessed, from which 51 were excluded. Finally, 12 studies, including 821 participants, were selected for this study. A flow chart of the study selection process is presented in Figure 1.

Figure 1. Flowchart of the study selection process. RCT: randomized controlled trial.



Study Characteristics

Studies included trials conducted in 7 countries: United States [31,39-42], Turkey [43], Canada [44], Korea [45], Singapore [46], Portugal [47], and China [48,49]. Two trials adopted a 3-arm group design [42,45], while 10 trials used a 2-arm group design. Sample sizes ranged from 20 to 172 participants. In all trials, participants were nursing students, except for one study in which participants were nursing staff [49]. Six of 12 trials compared VR education with traditional education [31,39,41,42,48,49], while the remaining trials compared VR education with other simulation types including fidelity manikin [44,47], mannequin-based simulation [40,45,46], and plastic model [43]. The characteristics of the participants, intervention details, and outcome measures are presented in Table 1. Supplementary information of intervention in experimental and control conditions is shown in Multimedia Appendix 2.

Table 1. Characteristics of the 12 included studies.

Author (year), country	Type of participant	Research project	Number of participants			Outcomes
			Total (number of groups)	Experimental group (VR[a])	Control group (condition)	
Bryant et al (2015) [31], USA	Nurse practitioner students	Advanced health assessment	60 (2)	22	38 (traditional education)	Satisfaction, confidence
Butt et al (2018) [39], USA	Junior level nursing students	Urinary catheterization	20 (2)	10	10 (traditional education)	Performance time
Cobbett and Snelgrove-Clarke (2016) [44], Canada	Third-year nursing students	Maternal -newborn nursing	56 (2)	27	28 (non-VR simulation)	Self-confidence
Haerling (2018) [40], USA	Fifth- and sixth-quarter associate degree in nursing students	Nursing care of patients with chronic obstructive pulmonary disease	28 (2)	13	15 (non-VR simulation)	Knowledge assessment, performance scores, satisfaction, and self-confidence
Ismailoglu and Zaybak (2018) [43], Turkey	Second-year nursing students	Intravenous catheter insertion	65 (2)	33	32 (non-VR simulation)	Knowledge assessment, skill scores, self-confidence scores
Jung et al (2012) [45], Korea	First-year nursing students	Intravenous injection	114 (3)	38	38 (non-VR simulation) and 38 (VR plus non-VR simulation)	Procedure score, satisfaction, performance time
Leflore et al (2012) [41], USA	Senior nursing students	Care of pneumonia and cystic fibrosis exacerbation	93 (2)	46	47 (traditional education)	Knowledge assessment
Liaw et al (2014) [46], Singapore	Senior nursing students	Assessing and managing deterioration	61 (2)	31	30 (non-VR simulation)	Performance scores
Padilha et al (2019) [47], Portugal	Second-year nursing students	Respiratory process in relation to ineffective airway clearance and hypoxia	42 (2)	21	21 (non-VR simulation)	Knowledge assessment, satisfaction
Smith et al (2018) [42], USA	Senior nursing students	Decontamination training	172 (3)	59 (immersive VR) 58 (desktop VR)	55 (traditional education)	Knowledge assessment
Tsai et al (2008) [49], China	Novice nurses	Port-A cath injection	82 (2)	42	40 (traditional education)	Knowledge assessment
Gu et al (2017) [48], China	Second-year students	Course of fundamental of nursing	28 (2)	14	14 (traditional education)	Knowledge assessment

[a]VR: virtual reality.

Risk of Bias

Based on the Cochrane criteria, a risk-of-bias assessment is presented in Figures 2 and 3. Four of 12 studies reported randomized methods in detail [41,43,46,47], while the remaining 8 trials did not provide the methods of sequence generation. None of the trials provided concealment methods, except for one that reported the use of anonymization [47]. In all trials, no blind method was used on participants due to the particularity of the intervention methods. Two trials reported employing blinding of assessors [39,43]. Additionally, 2 studies mentioned dropouts without detail on handling information [46,49].

Figure 2. Risk of bias analysis of each included study.

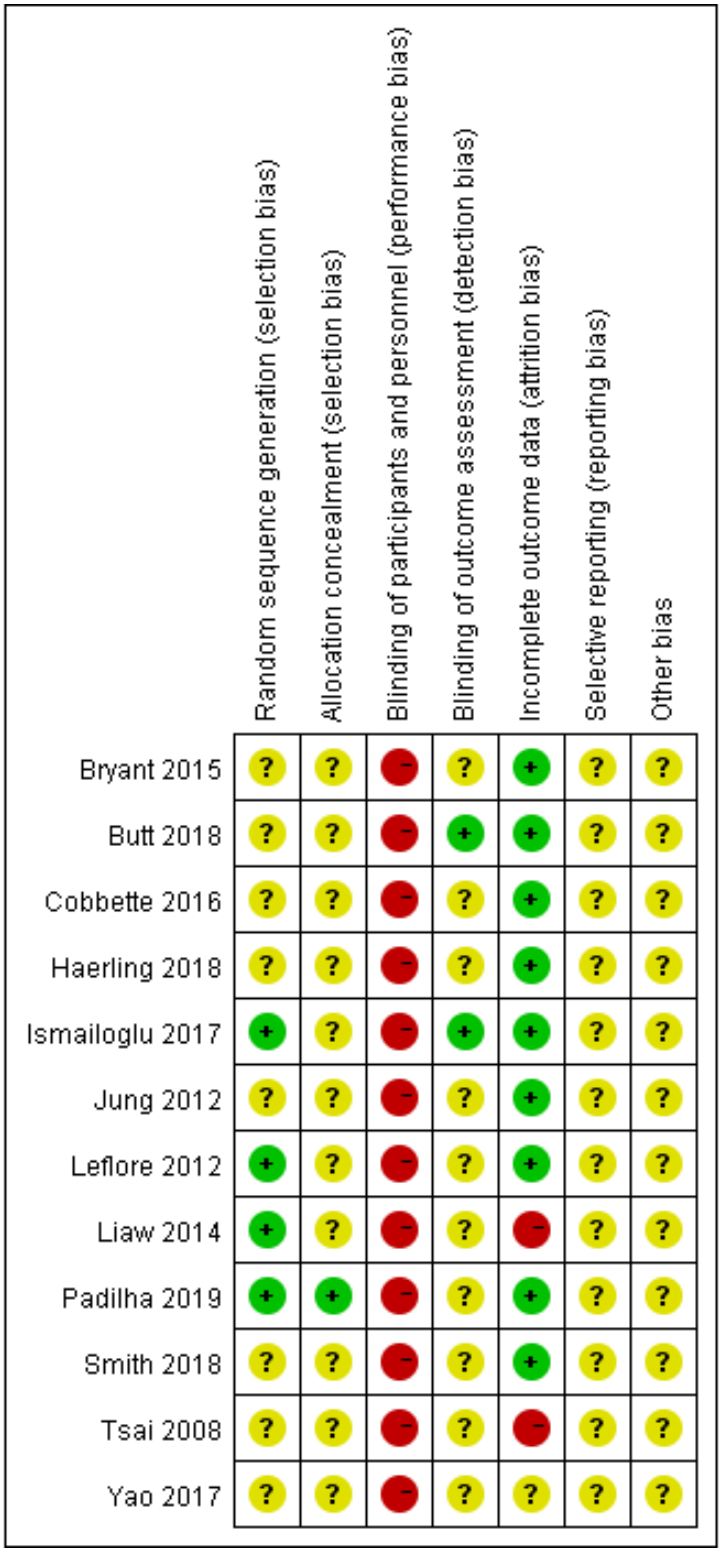
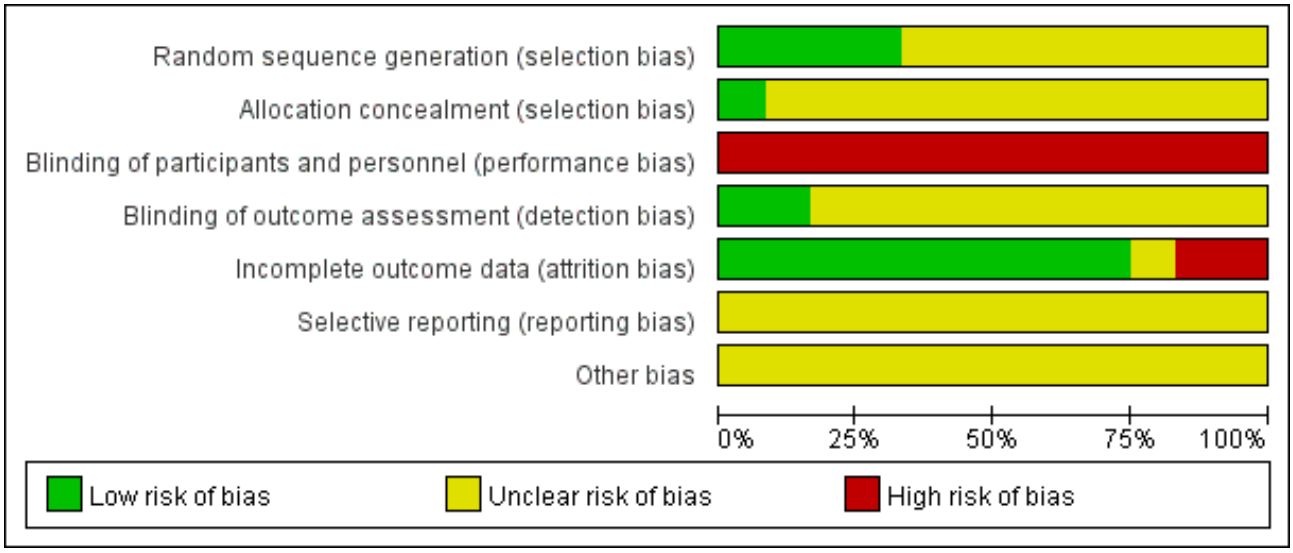


Figure 3. Overall risk of bias analysis of included studies.

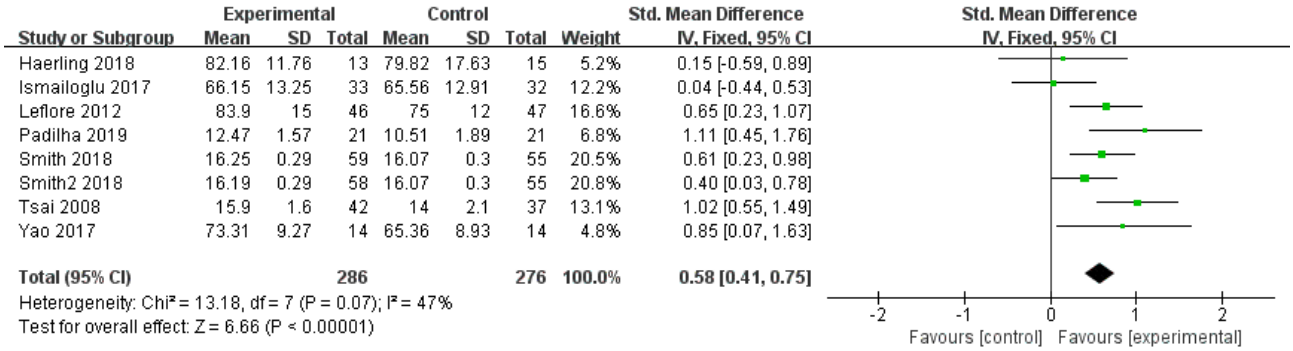


Results of the Meta-analysis

Knowledge

A total of 7 studies reported knowledge scores as the outcome [40-43,47-49]. The results indicated that VR education can improve knowledge of participants more effectively than the control conditions (SMD=0.58, 95% CI 0.41-0.75, P<.001, I^2=47%, Figure 4).

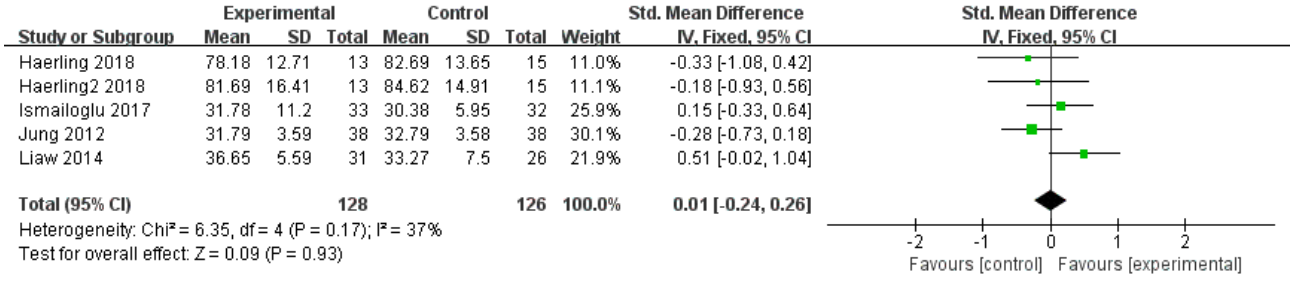
Figure 4. Forest plot of virtual reality on knowledge.



Skills

A total of 5 trials used skills as the outcome measure [40,42,43,45,46]. The results indicated that there was no significant difference between VR education and other education methods on skills enhancement (SMD=0.01, 95% CI –0.24 to 0.26), P=.93, I^2=37%; Figure 5).

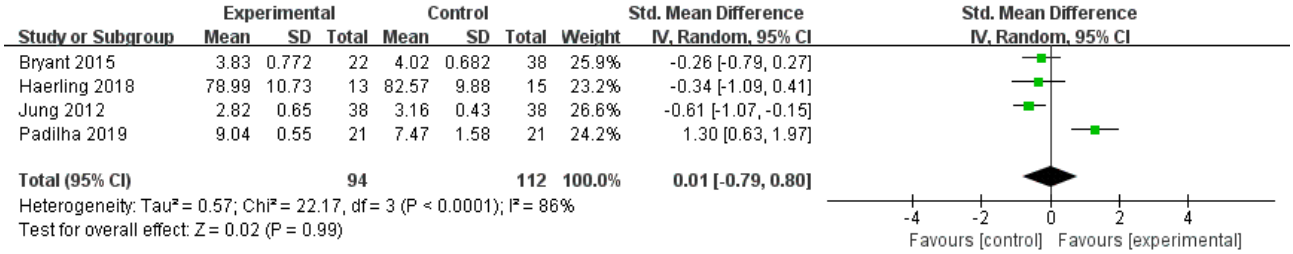
Figure 5. Forest plot of virtual reality on skills.



Satisfaction

A total of 4 articles reported participants' satisfaction scores [31,40,45,47]. Participants in VR groups showed no difference when compared to control groups (SMD=0.01, 95% CI –0.79 to 0.80, P=.99, I^2=86%). High heterogeneity was found. The leave-one-out method was used to carry out sensitivity analysis, and the random-effects model was adopted. One trial [47] caused significant heterogeneity, showing VR education is more satisfactory to participants than the control conditions (SMD=1.30, 95% CI 0.63-1.97, P=.001; Figure 6).

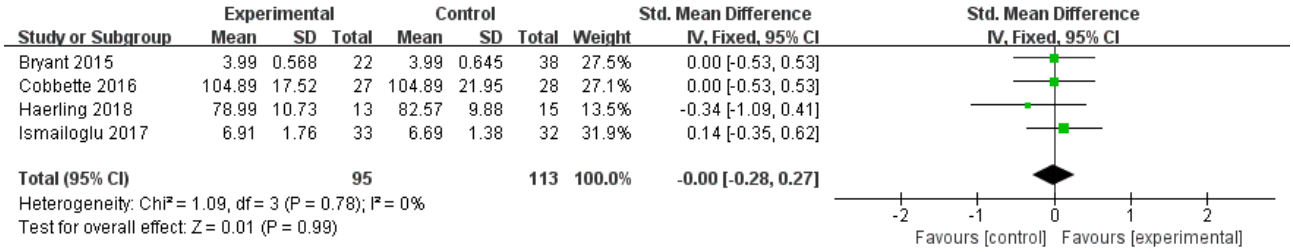
Figure 6. Forest plot of virtual reality on satisfaction.



Confidence

A total of 4 studies reported confidence results [31,40,43,44] and showed no statistical difference between VR education and other education methods (SMD=0.00, 95% CI –0.28 to 0.27, P=.99, I^2=0%; Figure 7).

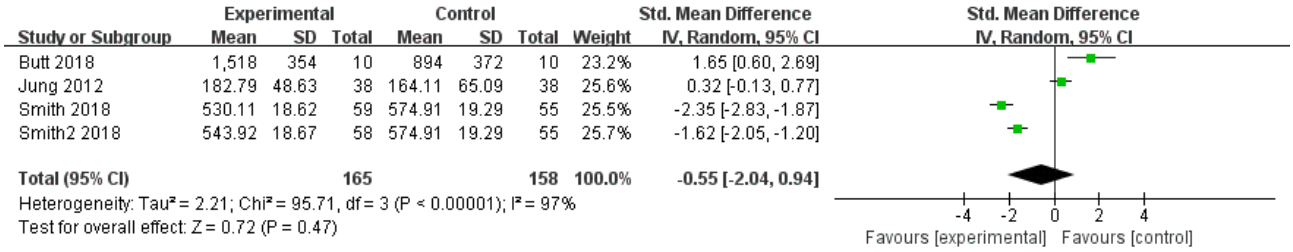
Figure 7. Forest plot of virtual reality on confidence.



Performance time

Performance time was employed as an outcome measure in 3 trials [39,42,45]. There was no significant difference between the experimental and control groups (SMD=–0.55, 95% CI –2.04 to 0.94, P=.47, I^2=97%]. Heterogeneity in this outcome was high. Therefore, the random-effects model was used, and the sensitivity analysis was carried out by using the leave-one-out method. Nevertheless, heterogeneity remained significant even when removing one study at a time (Figure 8).

Figure 8. Forest plot of virtual reality on performance time.



Discussion

This meta-analysis assessed the effectiveness of VR simulation methods in nursing education. We found that VR education methods can improve the knowledge of nursing students. However, there was no difference between VR and other education methods on the outcomes of skills, satisfaction, confidence, and performance time.

A total of 12 trials with 821 participants were included in the meta-analysis. All studies used VR education as the interventions in experimental groups, and education methods in control groups including traditional education, high/low fidelity manikin, mannequin-based simulation, and plastic model. Among the 12 studies, 4 trials reported random sequence generation. Only 1 study described the allocation concealment; 2 studies reported the blindness of outcome assessment. In addition, blinded interventions of students and educators were not possible because of the particularity of the VR education method. In general, the overall risk of bias of the included studies was judged to be unclear due to lack of information.

Knowledge

For the outcome of knowledge, VR education showed more effectiveness on nursing education than traditional education or other simulation education methods. A qualitative study on VR use in nursing education also concluded that, through the concrete experience of the virtual patient simulation and the reflection tool, students could understand what they were taught and how to utilize the new knowledge [50]. Additionally, a previous study, which focused on virtual reality for health professions education, indicated that VR with higher interactivity showed more effectiveness for knowledge [21]. These studies support the fact that an interactive learning environment encourages students to establish connections between concepts [51]. Most of the studies included in our meta-analysis used interactive VR education methods, which could explain the results.

Skills

Our results found no significant difference between VR education and other education methods for the outcome of skills, which seems to be in line with a previous systematic review [34]. The review concluded that virtual reality groups performed comparably to simulation groups in skill performance scores and skill success rate [34]. In our study, all the included trials that reported skills employed other simulation education methods in control groups. Similarly, we concluded that VR was not more effective in improving skills than other simulation methods in nursing education. A possible reason for these results is that there is a gap between completing virtual cases and real practice. Nursing skills learned on a virtual platform may not be transferable to real situations effectively because of the immaturity of VR technology [48].

Satisfaction

There was no significant difference on participants' satisfaction between VR education and education methods in control groups. High heterogeneity was found. Through sensitive analysis, we found that 1 of the 4 included studies showed that VR was more satisfactory [47]. In one trial conducted in 2012, some participants pointed out the immaturity of VR technology affecting users' satisfaction [45]. In contrast, 2 studies in recent years showed no difference between the 2 groups [31,40]. Thus, we consider that participants' satisfaction with VR education may vary according to technical conditions. Although in the 21st century nursing students had already shown high levels of usefulness, ease, and intention to use clinical VR simulation, VR is not widely used in nursing education [52]. With the progress of technology, VR can better satisfy the users. However, further research is needed to confirm our results.

Confidence

The results in confidence indicated no difference between experimental and control conditions. VR could not enhance the confidence of participants more effectively than control conditions, which was consistent with a previous study from Korea [53]. When VR was used for operation exercises, it was often necessary to use a mouse at the same time [53]. Thus, the operation method is more difficult in VR when compared with other simulations such as the manikin.

Performance Time

We also conducted a meta-analysis of performance time. The results suggested that VR was not more effective on reducing performance time than other educational methods. We found large heterogeneity among studies, even when a sensitivity analysis was conducted by using the leave-one-out method. The observed heterogeneity may be due to the different research designs of the selected studies, such as operation projects, VR devices, and education methods in control groups. One study on the effectiveness of VR endoscopy simulation training analyzed performance time with sufficient data and found no difference between VR and control groups; however, the quality of the evidence was very low [54]. In contrast, a study conducted in clinical medicine found that VR can help operators shorten performance time [55]. Therefore, more experiments are needed in the future to study the effectiveness of VR on performance time in nursing education.

Strengths and Limitations

Our study has the following strengths. First, our study is the first meta-analysis assessing the impact of VR on nursing education. Second, to assess the effectiveness of VR education, we evaluated 5 outcome measurements—knowledge, skills, satisfaction, confidence, and performance time—which can probably provide reference for nursing education.

There are also some limitations in our study. First, we only included articles published in English, which may affect the results of meta-analysis. Second, some of the included studies failed to provide the details of sequence generation, allocation concealment, and blinding methods. Third, we included 12 studies that have different interventions in control groups, which may cause significant heterogeneity among the studies.

Conclusions

This meta-analysis provides a comprehensive evaluation of the use of VR on nursing education. We found that VR education methods can improve nursing students' knowledge. However, for the outcomes of skills, satisfaction, confidence, and performance time, there seems to be no difference between VR and other education methods. In general, the use of VR should be considered to enhance knowledge and as a complement of other simulation strategies to improve the quality and safety of clinical practice. However, the heterogeneity and risk of bias among the included studies should be taken into consideration. Rigorously designed large-scale studies are required to further confirm the results in this review.

Acknowledgments

ZLS was supported by Key project of Jiangsu Province Education Science 13th Five-Year Plan 2016 projects (B-a/2016/01/18), Nanjing University of Chinese Medicine 2018 Flipped Classroom Course & Open Online Course - Health Assessment, and Opening project of innovative practical teaching team of Qing Lan Project in Jiangsu Province (NZYHLXPPQL2019-25).

Authors' Contributions

FQC and YFL searched the medical database and collected and extracted the data. FQC, YFL, and JFG discussed and analyzed data together and wrote papers. ZLS, DWW, CL, and BC provided suggestions for writing preparation and process. The final version of the article was reviewed by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies of PubMed, Embase and the Cochrane Library.
[DOCX File , 14 KB - jmir_v22i9e18290_app1.docx]

Multimedia Appendix 2
Supplementary information of intervention in experimental and control groups.
[DOC File , 45 KB - jmir_v22i9e18290_app2.doc]

References

1. Pepin J, Goudreau J, Lavoie P, Bélisle M, Blanchet Garneau A, Boyer L, et al. A nursing education research framework for transformative learning and interdependence of academia and practice. Nurse Educ Today 2017 May;52:50-52. [doi: 10.1016/j.nedt.2017.02.001] [Medline: 28249197]
2. Nardi DA, Gyurko CC. The global nursing faculty shortage: status and solutions for change. J Nurs Scholarsh 2013 Sep;45(3):317-326. [doi: 10.1111/jnu.12030] [Medline: 23895289]
3. Juraschek SP, Zhang X, Ranganathan V, Lin VW. United States registered nurse workforce report card and shortage forecast. Am J Med Qual 2019;34(5):473-481. [doi: 10.1177/1062860619873217] [Medline: 31479295]
4. Fact sheet: Nursing faculty shortage. American Association of Colleges of Nursing. 2019. URL: https://www.aacnnursing.org/Portals/42/News/Factsheets/Faculty-Shortage-Factsheet.pdf [accessed 2020-06-26]
5. Shin S, Park J, Kim J. Effectiveness of patient simulation in nursing education: meta-analysis. Nurse Educ Today 2015 Jan;35(1):176-182. [doi: 10.1016/j.nedt.2014.09.009] [Medline: 25459172]
6. Yuan HB, Williams BA, Fang JB, Ye QH. A systematic review of selected evidence on improving knowledge and skills through high-fidelity simulation. Nurse Educ Today 2012 Apr;32(3):294-298. [doi: 10.1016/j.nedt.2011.07.010] [Medline: 21862186]
7. Uysal N. Improvement of nursing students' learning outcomes through scenario-based skills training. Rev Lat Am Enfermagem 2016 Aug 08;24:e2790 [FREE Full text] [doi: 10.1590/1518-8345.1310.2790] [Medline: 27508922]
8. Dubovi I, Levy ST, Dagan E. Now I know how! The learning process of medication administration among nursing students with non-immersive desktop virtual reality simulation. Comput Educ 2017 Oct;113:16-27. [doi: 10.1016/j.compedu.2017.05.009]
9. Dalton L, Gee T, Levett-Jones T. Using clinical reasoning and simulation-based education to 'flip' the Enrolled Nurse curriculum. Aust J Adv Nurs 2015;33(2):28-34 [FREE Full text]
10. Ricketts B. The role of simulation for learning within pre-registration nursing education - a literature review. Nurse Educ Today 2011 Oct;31(7):650-654. [doi: 10.1016/j.nedt.2010.10.029] [Medline: 21074297]
11. Robinson BK, Dearmon V. Evidence-based nursing education: effective use of instructional design and simulated learning environments to enhance knowledge transfer in undergraduate nursing students. J Prof Nurs 2013;29(4):203-209. [doi: 10.1016/j.profnurs.2012.04.022] [Medline: 23910921]
12. Cant RP, Cooper SJ. Use of simulation-based learning in undergraduate nurse education: An umbrella systematic review. Nurse Educ Today 2017 Feb;49:63-71. [doi: 10.1016/j.nedt.2016.11.015] [Medline: 27902949]
13. Alexander M, Durham CF, Hooper JI, Jeffries PR, Goldman N, Kardong-Edgren S, et al. NCSBN Simulation Guidelines for Prelicensure Nursing Programs. Journal of Nursing Regulation 2015 Oct;6(3):39-42. [doi: 10.1016/s2155-8256(15)30783-3]
14. Brewer E. Successful techniques for using human patient simulation in nursing education. J Nurs Scholarsh 2011 Sep;43(3):311-317. [doi: 10.1111/j.1547-5069.2011.01405.x] [Medline: 21884377]
15. Flott EA, Linden L. The clinical learning environment in nursing education: a concept analysis. J Adv Nurs 2016 Mar 09;72(3):501-513. [doi: 10.1111/jan.12861] [Medline: 26648579]
16. Dev P, Youngblood P, Heinrichs WL, Kusumoto L. Virtual worlds and team training. Anesthesiol Clin 2007 Jun;25(2):321-336. [doi: 10.1016/j.anclin.2007.03.001] [Medline: 17574193]
17. Rosen KR. The history of medical simulation. J Crit Care 2008 Jun;23(2):157-166. [doi: 10.1016/j.jcrc.2007.12.004] [Medline: 18538206]
18. Agency for Healthcare Research and Quality. In: Lopreiato L, editor. Healthcare Simulation Dictionary. Rockville: Agency for Healthcare Research and Quality; 2016.
19. Cao C, Cerfolio RJ. Virtual or augmented reality to enhance surgical education and surgical planning. Thorac Surg Clin 2019 Aug;29(3):329-337. [doi: 10.1016/j.thorsurg.2019.03.010] [Medline: 31235302]
20. Sherman WR, Craig AB. Understanding virtual reality—Interface, application, and design. Presence: Teleoperators Virtual Environ 2003 Aug;12(4):441-442. [doi: 10.1162/105474603322391668]

21. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: Systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
22. Shin H, Rim D, Kim H, Park S, Shon S. Educational characteristics of virtual simulation in nursing: An integrative review. Clinical Simulation in Nursing 2019 Dec;37:18-28. [doi: 10.1016/j.ecns.2019.08.002]
23. Smith PC, Hamilton BK. The effects of virtual reality simulation as a teaching strategy for skills preparation in nursing students. Clinical Simulation in Nursing 2015 Jan;11(1):52-58. [doi: 10.1016/j.ecns.2014.10.001]
24. Foronda CL, Alfes CM, Dev P, Kleinheksel A, Nelson DA, O Donnell JM, et al. Virtually nursing: Emerging technologies in nursing education. Nurse Educ 2017;42(1):14-17. [doi: 10.1097/NNE.0000000000000295] [Medline: 27454054]
25. Fealy S, Jones D, Hutton A, Graham K, McNeill L, Sweet L, et al. The integration of immersive virtual reality in tertiary nursing and midwifery education: A scoping review. Nurse Educ Today 2019 Aug;79:14-19. [doi: 10.1016/j.nedt.2019.05.002] [Medline: 31078869]
26. Wright RR, Tinnon EA, Newton RH. Evaluation of vSim for nursing in an adult health nursing course: A multisite pilot study. Comput Inform Nurs 2018 Feb;36(2):84-89. [doi: 10.1097/CIN.0000000000000388] [Medline: 28952980]
27. Smith SJ, Farra S, Ulrich DL, Hodgson E, Nicely S, Matcham W. Learning and retention using virtual reality in a decontamination simulation. Nurs Educ Perspect 2016;37(4):210-214. [doi: 10.1097/01.NEP.0000000000000035] [Medline: 27740579]
28. Weiner E, Gordon J, Rudy S, McNew R. Expanding virtual reality to teach ultrasound skills to nurse practitioner students. Stud Health Technol Inform 2019 Aug 21;264:893-897. [doi: 10.3233/SHTI190352] [Medline: 31438053]
29. Foronda C, Gattamorta K, Snowden K, Bauman EB. Use of virtual clinical simulation to improve communication skills of baccalaureate nursing students: a pilot study. Nurse Educ Today 2014 Jun;34(6):e53-e57. [doi: 10.1016/j.nedt.2013.10.007] [Medline: 24231637]
30. Farra SL, Smith S, Gillespie GL, Nicely S, Ulrich DL, Hodgson E, et al. Decontamination training: with and without virtual reality simulation. Adv Emerg Nurs J 2015;37(2):125-133. [doi: 10.1097/TME.0000000000000059] [Medline: 25929223]
31. Bryant R, Miller CL, Henderson D. Virtual clinical simulations in an online advanced health appraisal course. Clin Simul Nurs 2015 Oct;11(10):437-444. [doi: 10.1016/j.ecns.2015.08.002]
32. Negu A, Matu S, Sava FA, David D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. Clin Neuropsychol 2016 Feb;30(2):165-184. [doi: 10.1080/13854046.2016.1144793] [Medline: 26923937]
33. Merchant Z, Goetz ET, Cifuentes L, Keeney-Kennicutt W, Davis TJ. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. Comput Educ 2014 Jan;70:29-40. [doi: 10.1016/j.compedu.2013.07.033]
34. Rourke S. How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review. Int J Nurs Stud 2020 Feb;102:103466. [doi: 10.1016/j.ijnurstu.2019.103466] [Medline: 31783192]
35. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
36. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, editors. Cochrane handbook for systematic reviews of interventions version 6.: Cochrane; 2019. URL: https://training.cochrane.org/handbook [accessed 2019-11-05]
37. RevMan 5. Cochrane training. URL: http://community.cochrane.org/tools/review-production-tools/revman-5 [accessed 2020-06-26]
38. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
39. Butt AL, Kardong-Edgren S, Ellertson A. Using game-based virtual reality with haptics for skill acquisition. Clin Simul Nurs 2018 Mar;16:25-32. [doi: 10.1016/j.ecns.2017.09.010]
40. Haerling KA. Cost-utility analysis of virtual and mannequin-based simulation. Simul Healthc 2018 Feb;13(1):33-40. [doi: 10.1097/SIH.0000000000000280] [Medline: 29373382]
41. LeFlore JL, Anderson M, Zielke MA, Nelson KA, Thomas PE, Hardee G, et al. Can a virtual patient trainer teach student nurses how to save lives--teaching nursing students about pediatric respiratory diseases. Simul Healthc 2012 Feb;7(1):10-17 [FREE Full text] [doi: 10.1097/SIH.0b013e31823652de] [Medline: 22228285]
42. Smith SJ, Farra SL, Ulrich DL, Hodgson E, Nicely S, Mickle A. Effectiveness of two varying levels of virtual reality simulation. Nurs Educ Perspect 2018;39(6):E10-E15. [doi: 10.1097/01.NEP.0000000000000369] [Medline: 30335708]
43. Günay İsmailoğlu E, Zaybak A. Comparison of the effectiveness of a virtual simulator with a plastic arm model in teaching intravenous catheter insertion skills. Comput Inform Nurs 2018 Feb;36(2):98-105. [doi: 10.1097/CIN.0000000000000405] [Medline: 29176359]
44. Cobbett S, Snelgrove-Clarke E. Virtual versus face-to-face clinical simulation in relation to student knowledge, anxiety, and self-confidence in maternal-newborn nursing: A randomized controlled trial. Nurse Educ Today 2016 Oct;45:179-184. [doi: 10.1016/j.nedt.2016.08.004] [Medline: 27537670]

45. Jung E, Park DK, Lee YH, Jo HS, Lim YS, Park RW. Evaluation of practical exercises using an intravenous simulator incorporating virtual reality and haptics device technologies. Nurse Educ Today 2012 May;32(4):458-463. [doi: 10.1016/j.nedt.2011.05.012] [Medline: 21664014]
46. Liaw SY, Chan SW, Chen F, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. J Med Internet Res 2014 Sep 17;16(9):e214 [FREE Full text] [doi: 10.2196/jmir.3322] [Medline: 25230684]
47. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical virtual simulation in nursing education: Randomized controlled trial. J Med Internet Res 2019 Mar 18;21(3):e11529 [FREE Full text] [doi: 10.2196/11529] [Medline: 30882355]
48. Gu Y, Zou Z, Chen X. The effects of vSIM for nursing™ as a teaching strategy on fundamentals of nursing education in undergraduates. Clin Simul Nurs 2017 Apr;13(4):194-197. [doi: 10.1016/j.ecns.2017.01.005]
49. Tsai S, Chai S, Hsieh L, Lin S, Taur F, Sung W, et al. The use of virtual reality computer simulation in learning Port-A cath injection. Adv Health Sci Educ Theory Pract 2008 Mar;13(1):71-87. [doi: 10.1007/s10459-006-9025-3] [Medline: 16944000]
50. Forsberg E, Ziegert K, Hult H, Fors U. Assessing progression of clinical reasoning through virtual patients: An exploratory study. Nurse Educ Pract 2016 Jan;16(1):97-103. [doi: 10.1016/j.nepr.2015.09.006] [Medline: 26482401]
51. Jeffries P. A framework for designing, implementing, and evaluating simulations used as teaching strategies in nursing. Nurs Educ Perspect 2005;26(2):96-103. [Medline: 15921126]
52. Padilha JM, Machado PP, Ribeiro AL, Ramos JL. Clinical Virtual Simulation in Nursing Education. Clin Simul Nurs 2018 Feb;15:13-18. [doi: 10.1016/j.ecns.2017.09.005]
53. Hwang J, Kim H. Comparison of training effectiveness for IV injections: Intravenous (IV) arm model versus computer simulator. J Korean Acad Fundam Nurs 2014 Aug 31;21(3):302-310 [FREE Full text] [doi: 10.7739/jkafn.2014.21.3.302]
54. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev 2018 Aug 17;8:CD008237 [FREE Full text] [doi: 10.1002/14651858.CD008237.pub3] [Medline: 30117156]
55. Mishra S, Kurien A, Patel R, Patil P, Ganpule A, Muthu V, et al. Validation of virtual reality simulation for percutaneous renal access training. J Endourol 2010 Apr;24(4):635-640. [doi: 10.1089/end.2009.0166] [Medline: 20218892]

Abbreviations

CINAHL: cumulative index to nursing and allied health literature
PICO: Population–Intervention–Comparison–Outcome
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
RCT: randomized controlled trials
VR: virtual reality

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by M Padilha, K Glover, T Hebda; comments to author 24.03.20; revised version received 17.04.20; accepted 26.07.20; published 15.09.20.

Please cite as:
Chen FQ, Leng YF, Ge JF, Wang DW, Li C, Chen B, Sun ZL
Effectiveness of Virtual Reality in Nursing Education: Meta-Analysis
J Med Internet Res 2020;22(9):e18290
URL: http://www.jmir.org/2020/9/e18290/
doi:10.2196/18290
PMID:32930664

Review

Patient Portal Functionalities and Patient Outcomes Among Patients With Diabetes: Systematic Review

Abrar Alturkistani[1], MPH; Ambar Qavi[2], MPH; Philip Emeka Anyanwu[1], MPH, PhD; Geva Greenfield[3], PhD; Felix Greaves[1], BM, BCh, MPH, MBA, PhD; Ceire Costelloe[1], MSc, PhD, CStat

[1]Global Digital Health Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[2]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[3]Public Health Policy Evaluation Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Abrar Alturkistani, MPH
Global Digital Health Unit
Department of Primary Care and Public Health
Imperial College London
St Dunstan's Rd, Hammersmith
London, W6 8RP
United Kingdom
Phone: 44 (0)20 7589 5111
Email: Abrar.alturkistani16@imperial.ac.uk

Abstract

Background: Patient portal use could help improve the care and health outcomes of patients with diabetes owing to functionalities, such as appointment booking, electronic messaging (e-messaging), and repeat prescription ordering, which enable patient-centered care and improve patient self-management of the disease.

Objective: This review aimed to summarize the evidence regarding patient portal use (portals that are connected to electronic health care records) or patient portal functionality use (eg, appointment booking and e-messaging) and their reported associations with health and health care quality outcomes among adult patients with diabetes.

Methods: We searched the MEDLINE, Embase, and Scopus databases and reported the review methodology using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Three independent reviewers screened titles and abstracts, and two reviewers assessed the full texts of relevant studies and performed data extraction and quality assessments of the included studies. We used the Cochrane Collaboration Risk of Bias Tool and the National Heart, Lung and Blood Institute (NHLBI) Study Quality Assessment Tool to assess the risk of bias of the included studies. Data were summarized through narrative synthesis.

Results: Twelve studies were included in this review. Five studies reported overall patient portal use and its association with diabetes health and health care quality outcomes. Six studies reported e-messaging or email use–associated outcomes, and two studies reported prescription refill–associated outcomes. The reported health outcomes included the associations of patient portal use with blood pressure, low-density lipoprotein cholesterol, and BMI. Few studies reported health care utilization outcomes such as office visits, emergency department visits, and hospitalizations. A limited number of studies reported overall quality of care for patients with diabetes who used patient portals.

Conclusions: The included studies mostly reported improved glycemic control outcomes for patients with diabetes who used patient portals. However, limitations of studying the effects of patient portals exist, which do not guarantee whether the outcomes reported are completely the result of patient portal use or if confounding factors exist. Randomized controlled trials and mixed-methods studies could help understand the mechanisms involved in health outcome improvements and patient portal use among patients with diabetes.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42019141131; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019141131.

International Registered Report Identifier (IRRID): RR2-10.2196/14975

(J Med Internet Res 2020;22(9):e18976) doi:10.2196/18976

KEYWORDS

personal health record; patient portal; electronic health records; online access; patient records; systematic review

Introduction

Background

Patient portals are online tools connected to health care systems' electronic health records (EHRs). The portals may improve patient health outcomes by improving communication with health care providers, enabling self-management of the disease, increasing patients' involvement in care, empowering patients, and improving their knowledge about the disease [1-6]. By offering access to information, such as visit summaries and health records, portals can help patients review information and remember doctors' instructions [7,8]. Asynchronous communication with health care providers through electronic messages (e-messages) or emails (a potential functionality of patient portals) can increase patients' interaction with the health care system and enable continuity of care [3,5,9,10]. Services, such as repeat prescription refills through the portal, could also improve efficiency and accelerate medication dispensing [3].

Rationale

The World Health Organization recommends a patient-centered approach when it comes to diabetes care and the use of technologies to engage patients [11]. The chronic care model, an evidence-based approach to manage chronic diseases, recommends "self-management support" to provide the best care for patients with chronic diseases [12]. Synthesizing and weighting the evidence about patient portals' effectiveness in improving diabetes health outcomes and quality of care could help inform health care professionals and policymakers about the potential benefits of patient portals.

There are several systematic reviews about patient portals used by patients with diabetes. However, published reviews are either outdated owing to new studies about patient portals being published in the last 2 to 3 years [13-16] or do not report patient outcomes [13]. Previous reviews that looked at diabetes health outcomes associated with portal use had a broader definition of portals and included portals that are co-delivered with other interventions, such as a diabetes management system, home visits [14], and coaching programs [16]. Another review only reported on the user characteristics of patients with diabetes, as well as facilitators and barriers of portal use [13], but did not report on the health outcomes associated with portal use. We hence aimed to close this gap by conducting a systematic review to summarize and evaluate the study findings that reported health and health care quality outcomes associated with the use of patient portals among adult patients with diabetes.

Objective

We aimed to summarize the evidence regarding the use of patient portals (portals that are connected to EHRs) and its reported association with health and health care quality outcomes among adult patients with diabetes. The review research questions were as follows: (1) What kind of health outcomes do patient portals contribute to in adult patients (18 years or older) with diabetes? (2) What kind of health care quality outcomes, including health care utilization outcomes, do patient portals contribute to in adult patients (18 years or older) with diabetes?

Methods

Guidelines and Study Registration

This review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [17] (Multimedia Appendix 1). The protocol of the review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42019141131) and was published in JMIR Research Protocols (RR1-10.2196/14975) [18].

Eligibility Criteria

The population included adult patients with diabetes aged 18 years or older. The initial intent in the review protocol [18] was to include all adult patient portal users, without focusing on patients with a specific disease. However, owing to the large number of studies reporting patient portal–related health outcomes and the diversity of the patient populations studied, we decided to focus on patients with diabetes only. The intervention only included tethered patient portals that are connected to a health care system's EHR. We excluded studies with additional interventions besides the patient portal, such as a wearable device and a portal with a mood monitoring tool [19], as we were unable to determine the outcomes associated with portal use only. Studies with comparators and no comparators were included. Outcomes of interest were health or health care quality outcomes. Qualitative and conference papers were excluded. Mixed-methods studies were only included if the quantitative results were of the outcomes of interest of the study. Finally, we excluded usability-only studies.

Information Sources

The MEDLINE, Embase, and Scopus databases were searched for relevant articles. The complete search strategy for each database has been provided in Multimedia Appendix 2. The search was performed up to September 2019, but there was no restriction on the start date of the search.

Study Selection and Data Extraction

Three reviewers independently performed title and abstract screening. Two reviewers (AA and AQ) independently assessed all full texts for eligibility, while a third reviewer (PEA) performed 25% (5 out of 20 articles) of the full-text reading. Data extraction was also performed independently by the two reviewers (AA and AQ). The extracted data included study design, population characteristics, patient portal characteristics, and study outcomes, and extraction was performed using the Cochrane primary screening and data extraction tool (Covidence) [20]. Any conflicts between the two reviewers were resolved through discussions with the third reviewer (PEA).

Risk of Bias

Studies were assessed for risk of bias using the Cochrane Collaboration Risk of Bias Tool [21] for randomized controlled trials (RCTs) and the National Heart, Lung and Blood Institute (NHLBI) Study Quality Assessment Tool for observational cohort and cross-sectional studies [22]. The NHLBI tool helps identify the "internal validity of studies" by guiding the users to identify methodological limitations [22]. Although studies are rated as good, fair, or poor, the tool does not assign numeric values or definite judgements of the quality of the studies. Thus, it is up to the authors' judgement to decide the severity of the risk of bias in studies using the guidance questions. In this review, we considered "good" as having a low risk of bias, "fair" as having a moderate risk of bias, and "poor" as having a high risk of bias. The risk of bias outcomes were considered when interpreting study findings in the discussion section.

Data Synthesis

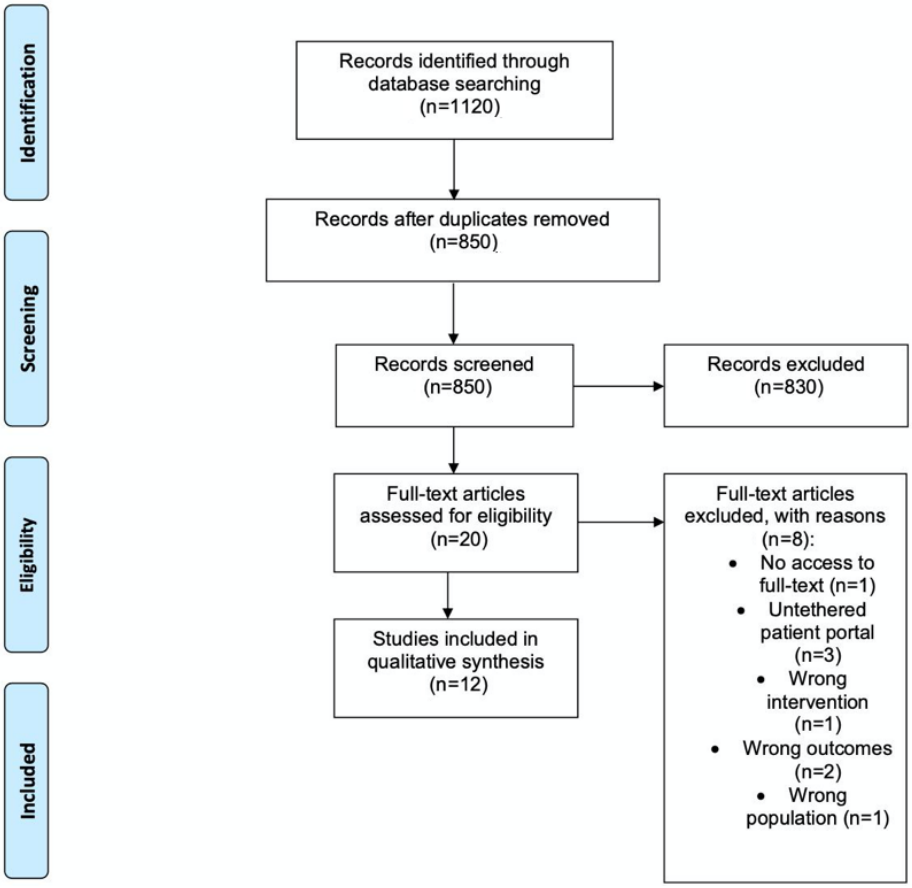
We were unable to carry out a meta-analysis owing to the variation in outcomes and methodologies used in the included studies. Therefore, we conducted a narrative synthesis of the results from the included studies based on the study designs and outcomes reported, paying attention to the relationship between the studies. We examined the relationship between the studies based on the patient portal functionalities. We decided to only report the outcomes of interest in relation to the patient portal functionality use because we found that there were similar patterns in the outcomes reported based on the functionalities. Although we collected information about the health care setting of each of the included studies, we did not find sufficient information to judge if the health care setting had a relevant influence on the outcomes. Thus, we did not compare study findings based on the health care setting. The narrative synthesis followed the methodologies proposed by the Cochrane Consumers and Communication Review Group data synthesis and analysis document [23] and the methodologies proposed as part of the UK Economic and Social Research Council Methods Program [24].

Results

Study Selection

A total of 1120 records were initially identified from the database searches (Figure 1). Among these, 830 studies were excluded after title and abstract screening (conference papers and irrelevant studies were excluded at this stage). Twenty full texts were assessed for eligibility, of which eight were excluded (Multimedia Appendix 3). Eventually, 12 studies were included in the final review, and a narrative synthesis was performed (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the search and study selection process.



Characteristics of the Included Studies

A summary table of the characteristics of the included studies is provided in Multimedia Appendix 4. Most of the studies were from the United States (n=11), with only one study from Canada [25]. Studies were performed in mixed settings including primary, secondary, and tertiary care (outpatient setting), and about half of the studies only included patients with type 2

diabetes. The primary care setting included patient portals offered by patients' primary care providers. The secondary care setting included patient portals provided in a hospital setting. The tertiary care setting included specialized outpatient care (eg, diabetologists). Integrated health care systems included primary care services provided in a hospital setting. None of the studies specified whether the patient portal was used in an in-patient setting. The retrospective cohort study design was the design of choice in most cases, and there was one RCT [26] and one cross-sectional study [27].

Table 1 lists the different patient portal functionalities used in the included studies. The functionalities varied between studies but mostly included viewing laboratory results (n=9), scheduling appointments (n=6), refilling medications (n=7), and sending messages or emails to health care providers (n=12) [25-36]. Other functionalities of the portals included requesting medical advice, updating demographic information by patients [28], entering flowsheet data, and offering a patient journal [25]. The intervention portal in the RCT [26] allowed patients to edit their medication lists, collected patient data on adverse effects of medication and response to therapy, and allowed patients to raise their health concerns through the portal. The control arm (patient portal) of the RCT included similar functionalities to the intervention portal and allowed patients to enter family medical history and review their patient records concerning nondiabetes-related health concerns such as cancer screening [26].

Table 1. Patient portal functionalities in the included studies.

Study	Patient portal functionalities						
	Lab results	Visit notes	Appointment booking	Repeat medication refill	Secure message/email	Patient education	Other
Randomized controlled trials							
Grant et al, 2008 [26]	Yes[a]	No[b]	Yes	Yes	Yes	No	Yes
Retrospective cohort studies							
Chung et al, 2017 [28]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Devkota et al, 2016 [29]	Yes	No	No	No	Yes	Yes	Yes
Lau et al, 2014 [25]	Yes	No	No	No	Yes	Yes	Yes
Lyles et al, 2016 [30]	Yes	Yes	Yes	Yes	Yes	No	No
McClellan et al, 2016 [31]	Yes	No	No	Yes	Yes	No	Yes
Petullo et al, 2016 [32]	No	No	No	No	Yes	No	Yes
Price-Haywood et al, 2017 [33]	No	No	Yes	Yes	Yes	No	Yes
Reed et al, 2019 [34]	Yes	Yes	Yes	Yes	Yes	No	No
Shimada et al, 2016 [35]	No	No	No	Yes	Yes	No	Yes
Tenforde et al, 2012 [36]	Yes	No	No	No	Yes	Yes	No
Cross-sectional studies							
Wade-Vuturo et al, 2013 [27]	Yes	No	Yes	No	Yes	No	Yes

[a]Yes indicates availability of the functionality.

[b]No indicates nonavailability or no mention of the functionality in the study.

Risk of Bias

The RCT included in this review [26] had a slightly high risk of bias (Multimedia Appendix 5) stemming from not being able to blind study participants or outcome assessors to the exposure status of participants. The RCT also had a high risk of bias owing to not reporting some of the study outcomes despite mentioning the outcomes in the methods section [26].

We rated most of the observational studies as having a low or moderate risk of bias (Multimedia Appendix 6). The studies that we rated as having a low or moderate risk of bias generally measured exposure before the outcomes [29,30,32-36], measured different levels of exposure (eg, compared portal functionality use by the number of days or number of times the functionality was used instead of having only one category for usage) [29-32,35,36], and controlled for key confounding variables [29-31,34-36]. Few studies that looked at the frequency or volume of e-messages, emails, or prescription refill use found that patients who used the functionality the most had better outcomes than patients who did not use the functionality or who used it less frequently. For example, one study reported that only patients who both read and wrote emails had much better glycemic control at follow up (odds ratio [OR] 1.43, 95% CI 1.11-1.83), which was not true for patients only reading emails or only using the patient portal [29]. Similarly, another study found that the odds of glycemic control was the highest among patients using the e-messaging functionality for 3 years or more (OR 1.28, 95% CI 1.13-1.44) compared with portal-only users (using the patient portal without the e-messaging functionality) [35]. One study found that patients sending four or more

messages per year were more likely to meet the glycemic control threshold compared with those sending one message only (OR 1.55, 95% CI 1.43-1.69) [28].

Health Outcomes

Table 2 summarizes the different outcomes reported to be associated with patient portal use, e-messaging or emailing, and medication refill through the portal. Overall, patient portal use was reported to be associated with glycemic control, reduced glycated hemoglobin A_{1c} (HbA_{1c}%) at follow up, reduced blood pressure, increased office visits, reduced hospitalizations, medication adherence, and medication adjustment. One study did not find a significant improvement in glycemic control as a result of using a patient portal (*P*=.62); however, the study offered both the control and experimental groups access to a patient portal [26]. E-message or email use was reported to be associated with glycemic control, reduced HbA_{1c}% at follow up, reduced low-density lipoprotein cholesterol, and increased office visits. Only one study examined the difference in BMI among portal users and nonusers and found no relevant difference [36]. Refilling medications through the patient portal was reported to be associated with glycemic control, blood pressure control, and medication adherence [35]. Only one study reported an association between refilling medications exclusively through the patient portal and improved statin adherence [30]. Although not listed in Table 2, one study found no correlation regarding reviewing laboratory results; viewing medical records; accessing billing information, the telephone directory, maps/directions, and insurance information; finding a doctor; and paying medical bills through the portal [27].

Table 2. Patient portal or patient portal functionality use and the reported associations with diabetes health and health care outcomes in the included studies.

Outcome	Overall patient portal use (n=5)	Electronic messaging or email use (n=6)	Prescription refill use (n=2)
Glycemic control	Positive association (n=1) [25], no association (n=1) [26]	Positive association (n=3) [28,29,35], weak correlation (n=1) [27], no association (n=1) [31]	Positive association (n=1) [35]
Hemoglobin A_{1c} (HbA_{1c}%) at follow up	Inverse association (n=3) [25,33,36]	Inverse association (n=3) [29,31,32]	—[a]
Low-density lipoprotein cholesterol	No association (n=2) [25,36]	Inverse association (n=1) [35], no association (n=1) [31]	Inverse association (n=1) [35]
Blood pressure	Inverse association (n=1) [36], no association (n=1) [25]	No association (n=2) [31,35]	Inverse association (n=1) [35]
BMI	No association (n=1) [36]	—	—
Office visits	Positive association (n=1) [34]	Positive association (n=1) [28]	—
Emergency visits	Inverse association (n=1) [34]	No association (n=1) [32]	—
Hospitalization	Inverse association (n=1) [34]	No association (n=1) [32]	—
Medication adherence	—	—	Positive association (n=1) [30]
Medication adjustment	Positive association (n=1) [26]	—	—

[a]There were no studies reporting an association between functionality and outcome.

Diabetes Care Quality Outcomes

Three studies reported that patient portal or e-message users were more likely to meet most of the diabetes care standards, such as the Diabetes Healthcare Effectiveness Data and Information Set (HEDIS) quality measures [28,31], or the diabetes standards by the Better Health Partnership: Diabetes Standards [36].

Discussion

Review of the Findings

Our review found a limited number of studies examining the association between patient portal use and diabetes health and health care quality outcomes. Nevertheless, among the studies included, patient portal use or patient portal functionality use was reported to be associated with improvements in health outcomes, such as glycemic control. Secure messaging or emailing, or repeat prescription ordering through the patient portal was reported to be associated with improved glycemic control, and outcomes appeared to improve with increased use. It was also reported that patient portal use may be associated with improved low-density lipoprotein cholesterol outcomes or blood pressure control. Patient portal use or patient portal functionality use might affect health care utilization and may be associated with increased office visits and decreased emergency department visits. Finally, some of the included studies suggested that patient portal use might be associated with improved quality of care for patients with diabetes.

The majority of studies we reviewed were determined to have low to moderate risk of bias. However, some factors may not be measured through standard risk of bias tools, which might have affected the results reported by the studies. For instance, it is challenging to separate outcomes that result exclusively from portal use owing to the possibility of the presence of other factors that might confound the association. It is also challenging to conclude which functionality of the patient portal contributes the most to improving health outcomes, as some studies only

report overall portal use. It was previously reported that secure messaging improves health outcomes for patients with diabetes [14]. Increased contact between patients with diabetes and health care professionals was one of the functionalities most associated with reduced HbA_{1c} in diabetes disease-management programs as reported in a meta-analysis of studies [37]. Our review also included studies that suggest glycemic control is improved in patients who use secure messaging owing to improved communication and increased access to care [29], resulting in "better diabetes management" [28].

Outcomes related to the quality of care and health care utilization were mixed. While some studies found reductions in emergency visits, others did not. A previous study in a diverse patient population found that there was no association between patient portal use, hospital admission, and 30-day readmission, suggesting that patient portals could be more effective in managing chronic care than acute care [38]. Alternatively, one of the included studies in this review suggested that actions related to portal use, such as checking a test result, can increase office visits, while actions, such as repeat prescription ordering during the after-hours period, might reduce hospitalizations [34]. A recent survey study of a patient portal with access to health care records, test results, e-messaging, and appointment booking reported that patients believed portal use helped them "avoid a clinic visit" [39]. Few studies in the literature also examined the association between patient portal use and missing medical appointments [39,40], which was not examined in any of the studies included in this review. The health care utilization outcomes associated with patient portal use may need further investigation as the number of studies examining these associations is limited.

Knowledge Gap

The outcomes of this review indicate that there remain persistent gaps in the literature about patient portals used by patients with diabetes. First, there is some evidence that increased frequency of patient portal or patient portal functionality use could be associated with increased benefits, suggesting a dose-response relationship. Patient portal adoption does not indicate continuous use [40]. Since differences in the frequency of use may lead to inconsistencies in benefits acquired from the patient portal, studies need to account for the frequency of patient portal use as much as possible. Additionally, there is a need for mixed-methods studies to evaluate the mechanisms through which portal use might impact the outcomes reported. Further examination of health care utilization outcomes could help understand if patient portals can play a role in improving health care utilization patterns among patients with diabetes.

Limitations

The results reported by the studies in this review could be biased owing to factors that may not have been controlled. For example, patient portal users can be generally more motivated to be involved in their care and to improve their health outcomes [29]. A cross-sectional study reported that patients who preferred using portals had higher "self-determination" to manage their health conditions [41]. RCTs could help explore causal relationships between portal use and outcomes in patients with diabetes. Additionally, qualitative studies or mixed-methods studies can help explain if portal use or patient portal functionality use is responsible for improving health and quality of care outcomes among patients with diabetes. Qualitative studies could help explore patients' motives and patterns in self-management to further help understand the mechanisms involved in improving health outcomes through patient portal use. There continues to be a need for studies to report outcomes based on functionality whenever possible [16].

Although this review tried to report portal functionality–related outcomes along with overall portal-related outcomes, most included studies did not sufficiently report outcomes by functionality. Another limitation of this review is that all studies, except one, were from the United States, which has a diverse health care system involving private health care organizations, nonprofit organizations, and government-owned organizations. The way that the health care organization is organized may limit the application of the findings of this review to other health care settings and systems.

Another limitation of this review is the small number of studies included. The limited number of studies reduced the generalizability of the review findings. However, the review only attempted to identify associations as reported by the included studies, which warrants further appropriately designed studies in order to assess causal associations.

Conclusion

Most of the included studies reported improved glycemic control outcomes for patients with diabetes who used patient portals. However, limitations of studying the effects of patient portals exist, which do not guarantee whether the outcomes reported are completely the result of patient portal use or if confounding factors exist. RCTs and mixed-methods studies could help understand the mechanisms involved in health outcome improvements and patient portal use among patients with diabetes.

Acknowledgments

PA is funded by the Economic and Social Research Council (ESRC) through the Antimicrobial Resistance Cross Council Initiative supported by the seven research councils in partnership with other funders (grant reference: ES/P008232/1). CC is supported through a personal NIHR Career development fellowship (reference: NIHR-CDF-016-015) and by NW London NIHR Applied Research Collaboration (reference NIHR-200180). This report is independent research supported by the National Institute for Health Research Applied Research Collaboration Northwest London (ARC Grant No.NIHR200180). The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research or the Department of

Health and Social Care. Imperial College London is grateful for support from the NW London NIHR Applied Research Collaboration. The support of the ESRC is gratefully acknowledged.

Authors' Contributions

AA, AQ, and PEA performed screening of the titles and abstracts, and full-text reading. AA and AQ completed data extraction and quality assessment of the included studies. GG, FG, and CC contributed to the development of the protocol and the review methodology. All authors reviewed and approved the final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2009 checklist.
[DOCX File , 19 KB - jmir_v22i9e18976_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 14 KB - jmir_v22i9e18976_app2.docx]

Multimedia Appendix 3
List of excluded studies with reasons for exclusion.
[DOCX File , 17 KB - jmir_v22i9e18976_app3.docx]

Multimedia Appendix 4
Characteristics of the included studies.
[DOCX File , 17 KB - jmir_v22i9e18976_app4.docx]

Multimedia Appendix 5
Risk of bias assessment results on applying the Cochrane Collaboration Risk of Bias Tool for randomized controlled trials by Grant et al [26].
[DOCX File , 14 KB - jmir_v22i9e18976_app5.docx]

Multimedia Appendix 6
Risk of bias assessment results on applying the National Heart, Lung and Blood Institute quality assessment tool for observational cohort and cross-sectional studies.
[DOCX File , 14 KB - jmir_v22i9e18976_app6.docx]

References

1. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
2. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
3. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
4. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct Release of Test Results to Patients Increases Patient Engagement and Utilization of Care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
5. Robinson SA, Zocchi MS, Netherton D, Ash A, Purington CM, Connolly SL, et al. Secure Messaging, Diabetes Self-management, and the Importance of Patient Autonomy: a Mixed Methods Study. J Gen Intern Med 2020 May 21. [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]
6. Nahm E, Zhu S, Bellantoni M, Keldsen L, Charters K, Russomanno V, et al. Patient Portal Use Among Older Adults: What Is Really Happening Nationwide? J Appl Gerontol 2020 Apr;39(4):442-450. [doi: 10.1177/0733464818776125] [Medline: 29779422]

7. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
8. Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record--a quantitative study. Fam Pract 2011 Apr;28(2):188-194. [doi: 10.1093/fampra/cmq092] [Medline: 21084568]
9. Bavafa H, Hitt LM, Terwiesch C. The Impact of E-Visits on Visit Frequencies and Patient Health: Evidence from Primary Care. Management Science 2018 Dec;64(12):5461-5480. [doi: 10.1287/mnsc.2017.2900]
10. Hefner JL, MacEwan SR, Biltz A, Sieck CJ. Patient portal messaging for care coordination: a qualitative study of perspectives of experienced users with chronic conditions. BMC Fam Pract 2019 May 03;20(1):57 [FREE Full text] [doi: 10.1186/s12875-019-0948-1] [Medline: 31053063]
11. Global Report on Diabetes. Geneva, Switzerland: World Health Organization; 2016.
12. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. Prev Chronic Dis 2013;10:E26 [FREE Full text] [doi: 10.5888/pcd10.120180] [Medline: 23428085]
13. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
14. Coughlin SS, Williams LB, Hatzigeorgiou C. A systematic review of studies of web portals for patients with diabetes mellitus. Mhealth 2017;3:23 [FREE Full text] [doi: 10.21037/mhealth.2017.05.05] [Medline: 28736732]
15. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]
16. Sun R, Korytkowski MT, Sereika SM, Saul MI, Li D, Burke LE. Patient Portal Use in Diabetes Management: Literature Review. JMIR Diabetes 2018 Nov 06;3(4):e11199 [FREE Full text] [doi: 10.2196/11199] [Medline: 30401665]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
18. Alturkistani A, Greenfield G, Greaves F, Aliabadi S, Jenkins RH, Costelloe C. Patient Portal Functionalities and Uptake: Systematic Review Protocol. JMIR Res Protoc 2020 Jul 31;9(7):e14975 [FREE Full text] [doi: 10.2196/14975] [Medline: 32734928]
19. van den Heuvel SC, Meije D, Regeer EJ, Sinnema H, Riemersma RF, Kupka RW. The user experiences and clinical outcomes of an online personal health record to support self-management of bipolar disorder: A pretest-posttest pilot study. J Affect Disord 2018 Oct 01;238:261-268. [doi: 10.1016/j.jad.2018.05.069] [Medline: 29894931]
20. Covidence. URL: https://www.covidence.org/ [accessed 2019-09-29]
21. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
22. Study Quality Assessment Tools. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2019-09-29]
23. Cochrane Consumers and Communication. URL: http://cccrg.cochrane.org [accessed 2019-09-28]
24. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme. 2006. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.3100&rep=rep1&type=pdf [accessed 2020-06-29]
25. Lau M, Campbell H, Tang T, Thompson DJ, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
26. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 08;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]
27. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
28. Chung S, Panattoni L, Chi J, Palaniappan L. Can Secure Patient-Provider Messaging Improve Diabetes Care? Diabetes Care 2017 Oct;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
29. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an Online Patient Portal and Glucose Control in Primary Care Patients with Diabetes. Popul Health Manag 2016 Apr;19(2):125-131. [doi: 10.1089/pop.2015.0034] [Medline: 26237200]
30. Lyles CR, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
31. McClellan SR, Panattoni L, Chan AS, Tai-Seale M. Patient-initiated Electronic Messages and Quality of Care for Patients With Diabetes and Hypertension in a Large Fee-for-Service Medical Group. Medical Care 2016;54(3):287-295. [doi: 10.1097/mlr.0000000000000483]

32. Petullo B, Noble B, Dungan KM. Effect of Electronic Messaging on Glucose Control and Hospital Admissions Among Patients with Diabetes. Diabetes Technol Ther 2016 Sep;18(9):555-560 [FREE Full text] [doi: 10.1089/dia.2016.0105] [Medline: 27398824]
33. Price-Haywood EG, Luo Q. Primary Care Practice Reengineering and Associations With Patient Portal Use, Service Utilization, and Disease Control Among Patients With Hypertension and/or Diabetes. Ochsner J 2017;17(1):103-111 [FREE Full text] [Medline: 28331456]
34. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, Hsu J, et al. Patients with complex chronic conditions: Health care use and clinical events associated with access to a patient portal. PLoS One 2019;14(6):e0217636 [FREE Full text] [doi: 10.1371/journal.pone.0217636] [Medline: 31216295]
35. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. J Med Internet Res 2016 Jul 01;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
36. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
37. Pimouguet C, Le Goff M, Thiébaut R, Dartigues JF, Helmer C. Effectiveness of disease-management programs for improving diabetes care: a meta-analysis. CMAJ 2011 Feb 08;183(2):E115-E127 [FREE Full text] [doi: 10.1503/cmaj.091786] [Medline: 21149524]
38. Dumitrascu AG, Burton MC, Dawson NL, Thomas CS, Nordan LM, Greig HE, et al. Patient portal use and hospital outcomes. J Am Med Inform Assoc 2018 Apr 01;25(4):447-453. [doi: 10.1093/jamia/ocx149] [Medline: 29300961]
39. Graham T, Ali S, Avdagovska M, Ballermann M. Effects of a Web-Based Patient Portal on Patient Satisfaction and Missed Appointment Rates: Survey Study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
40. Zhong X, Park J, Liang M, Shi F, Budd PR, Sprague JL, et al. Characteristics of Patients Using Different Patient Portal Functions and the Impact on Primary Care Service Utilization and Appointment Adherence: Retrospective Observational Study. J Med Internet Res 2020 Feb 25;22(2):e14410 [FREE Full text] [doi: 10.2196/14410] [Medline: 32130124]
41. Assadi V, Hassanein K. Consumer Adoption of Personal Health Record Systems: A Self-Determination Theory Perspective. J Med Internet Res 2017 Jul 27;19(7):e270 [FREE Full text] [doi: 10.2196/jmir.7721] [Medline: 28751301]

Abbreviations

EHR: electronic health record
OR: odds ratio
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 30.03.20; peer-reviewed by R Sun, C Kruse, S Shimada; comments to author 12.06.20; revised version received 16.07.20; accepted 11.08.20; published 22.09.20.

Please cite as:
Alturkistani A, Qavi A, Anyanwu PE, Greenfield G, Greaves F, Costelloe C
Patient Portal Functionalities and Patient Outcomes Among Patients With Diabetes: Systematic Review
J Med Internet Res 2020;22(9):e18976
URL: http://www.jmir.org/2020/9/e18976/
doi:10.2196/18976
PMID:32960182

Review

Digital Health Coaching Programs Among Older Employees in Transition to Retirement: Systematic Literature Review

Vera Stara[1*], PhD; Sara Santini[2*], MSc; Johannes Kropf[3], MSc; Barbara D'Amen[2*], PhD

[1]Model of Care and New Technologies, IRCCS INRCA-National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy

[2]Centre for Socio-Economic Research on Aging, IRCCS INRCA-National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy

[3]Health and Environment, Austrian Institute of Technology, Vienna, Austria

[*]these authors contributed equally

Corresponding Author:
Sara Santini, MSc
Centre for Socio-Economic Research on Aging
IRCCS INRCA-National Institute of Health and Science on Aging
Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani
via S. Margherita 5
Ancona, 60124
Italy
Phone: 39 3286191298
Email: s.santini2@inrca.it

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/12/e25065/

Abstract

Background: The rapid increase of the aging population is pushing many national governments to reshape retirement legislation in order to extend older adults' working life. Once retired, older adults can be invaluable resources for the community as family carers, as volunteers, or by returning to work. Healthy aging is one of the main conditions for being able to work longer and being active after retirement. The latter, indeed, represents a very sensitive life transition, which can entail psychological and social difficulties. Interventions for promoting older workers' health and well-being and supporting the transition to retirement are on the top of the policy agenda of most European countries. Recently, computer-based and digital health interventions have been seen as promising means to reach this purpose.

Objective: This systematic literature review aimed to explore studies on digital health coaching programs for older workers that followed a user-centered design approach and evaluated their effectiveness in providing older adults with guidance for adopting a healthy lifestyle and being active in the community.

Methods: The search identified 1931 papers, and 2 relevant articles were selected by applying specific eligibility criteria.

Results: To our knowledge, only few digital health coaching programs have targeted the population of older workers to date; there is an insufficient number of studies on the efficacy of such programs. The results show the difficulties of assessing the efficacy of digital coaching itself and with respect to older employees. The 2 studies suggest that digital health programs for workplaces can improve various aspects of older employees' well-being; however, they considered health mainly from a physical perspective and neglected contextual, social, psychological, and cultural factors that can influence older workers' health and general well-being. Future digital health coaching programs should adopt the healthy aging paradigm as a multidimensional lens for interpreting the impact of eHealth technology on aging and retirement. The literature around this issue remains at an embryonic state, and this gap needs to be filled by further investigations that apply a user-centered approach for designing the technology, test innovative research methodologies, and adopt new technical solutions for high-quality interaction design.

Conclusions: Further digital health coaching programs aimed at supporting healthy and active living for older workers and retirees are necessary. The user-centered design approach is recommended in order to fully address the users' health needs and the technological requirements throughout development. Moreover, the healthy aging perspective allows inclusion of physical,

social, and psychological factors influencing the transition from work to retirement, as well as the experiences and interactions of individuals with the technology.

(J Med Internet Res 2020;22(9):e17809) doi:10.2196/17809

KEYWORDS

older workers; retirees; transition to retirement; healthy aging; active aging; digital coach; virtual coach; user-centered design; virtual agent; avatar, virtual personal assistant

Introduction

The population is aging rapidly worldwide and by 2050, or soon thereafter, the number of persons aged 65 or older will outnumber those aged 25 or under in eastern and southeastern Asia, Latin America and the Caribbean, Europe, and North America [1], thereby determining the aging of part of the workforce in most countries. In fact, between 2002 and 2018, the employment rate of persons aged 55-64, called older workers (despite the lack of general consensus on the age range among researchers and policy makers) [2], increased from 38.4% to 58.7% throughout Europe [3] and from 58.9% to 62.5% [4] in the United States.

The aging and retiring workforce can result in a loss of know-how and expertise for companies, which may negatively affect the economy and the sustainability of social security systems [5].

In order to mitigate the consequences of an aging workforce, in the last decade, many national governments of high-income countries have been changing retirement legislation to postpone the retirement age and extend older adults' working life [6]. Continuing work in later life can indeed be considered as an aspect of *successful aging* [7], a topic on which, not surprisingly, the scientific community and the labor market has been lavishing more and more attention during the last years. An exact definition of successful aging is still a mooted point; nevertheless, it can be considered as a multidimensional research construct consisting of distinct but interrelated facets for the identification of determinants and predictors related to a favorable aging trajectory as opposed to a pathological one [8,9]. This paradigm puts emphasis on the responsibility of older people to continue to contribute to the society by working and being active [10].

Since people can be active and productive when they are in good health, it seems that a precondition of successful aging is healthy aging. The latter is not just the absence of illness, but also a "process of optimizing opportunities for physical, social and mental health to enable older people to take an active part in society without discrimination and to enjoy an independent and good quality of life [11]." Therefore, the healthy aging perspective considers every sphere of life influencing the individuals' well-being, including social relationships, and it is particularly appropriate for looking at older employees' experiences. Older people at the end of their working life may indeed go through a delicate phase of their existence that could be characterized by poor health but could also be characterized by psychological and relational strains at the workplace. In fact, older workers can be stigmatized in organizational settings [12] and associated with scant motivation, less alert capacity, limited productivity and flexibility [13], more resistance to change and learning [14], less reliability (from poor health), and limited digital skills [15,16]. When older workers feel stereotyped and unappreciated and when they experience effort-reward imbalance, depressive symptoms may develop [17].

In the months before retiring, the individuals' attitudes and expectations about retirement can also be influenced by cultural patterns [18]: in countries where family culture prevails (eg, Italy), people may be more motivated to leave the labor market, because they interpret retirement as an opportunity to devote themselves to family, for example, by taking care of grandchildren or very old and non self-sufficient parents. Conversely, in other countries, where culture is more oriented to productivity as a measure of social success, such as the United States and the United Kingdom, individuals tend to remain in the labor market longer and return to work even after retirement. Moreover, some older employees plan their retirement down to the tiniest details, while others prefer not to make any plans due to fear of being unable to implement them or because they want to enjoy their freedom [18].

Once retired, older people can feel different levels of satisfaction with life, not directly associated with fulfilment of the plans and the expectations they had while working. However, satisfaction seems to be the result of the combination of multiple factors (ie, quantity and quality of interpersonal relationships and affective bonds, physical health, and financial situation within the cultural and social context) [19]. Therefore, retirement may involve both a threat of marginality and a promise of new-found freedom [20]. In the former case, the transition to retirement may be a troubling experience with anxieties, concerns, and social isolation. Conversely, in the latter case, retirement can offer scope for expanded opportunities for a new, active, and positive phase of life.

Older employees' decision to leave work, as well as attitudes, expectations and plans for retirement, are mainly influenced by their health. In fact, the onset of chronic diseases and comorbidities are the first reasons for early retirement and are major barriers for older workers' active participation in the labor market. Indeed, according to the Survey of Health, Ageing and Retirement, in Europe, among people aged 60 to 70 years, those in good health participated in the labor market approximately twice as often as older people with 2 or more chronic diseases [21].

In light of the above, carrying out health promotion and prevention interventions targeting older workers is a prime objective of labor policy for keeping employees active and productive longer [22]. Moreover, according to Cook et al [23],

there is strong evidence that well-constructed health promotion programs for employees in the workplace can be a key strategy to improve workers' health and decrease health care costs.

One strategy to help individuals having healthier lifestyles (ie, increase physical activity, have a healthy diet, and reduce the use of tobacco and alcohol) is offering them health and lifestyle coaching through interpersonal relationship with a trainer [24]. This can motivate the individual to walk or run, eat more vegetables, limit the intake of fats, and reduce the number of cigarettes per day. Literature shows that health and lifestyle coaching programs are feasible and accepted by patients especially when based on a patient-centered approach [25]. This approach is largely adopted for older people with chronic diseases and multimorbidities [26], for prevention [27], and in primary care. Health coaching interventions are in accordance with the concept of the activation of patients, which allows them to partially determine their goals, use self-discovery or active learning processes together with content education to work toward their goals, and self-monitor behaviors to increase accountability and adherence to the program [25,28].

Since the advancement of technology has had an impressive influence on our daily living, recently coaching programs for the promotion of health have been getting increasingly more computer-based; digital health interventions are a promising approach to address older workers' health needs [29]. Effectively, under the umbrella term of *coaching*, we usually discover studies that describe interventions delivered without a human coach but through different technologies used to drive the behavior change process. According to Sherpa Coaching Survey [30], for instance, currently only 32% of coaching is conducted face-to-face, while 68% is delivered through technological tools such as telephone (25%), webcams and Skype (25%), video conferencing in high-definition quality (10%), and e-mail coaching (8%). Despite this promising trend, few digital coaching programs have targeted the specific population of older employees by exploring their experiences with this type of technology. Consequently, studies that prove the efficacy of health programs using technology to deliver coaching services are still needed. Investigations into successful coaching systems point to the concept of user-centered design as a way of understanding the users and their needs in multiple steps of the iterative development process [31]. User-centered design is a multidisciplinary design approach based on active involvement of users to improve understanding of the consumer model. End users' needs, capabilities, and limitations are mapped making use of a variety of methods and tools offered by this approach, and it is indeed a way to ensure that product solutions match the target's demand, especially for health care technologies, where acceptance, usability, and reliability are essential to ensure the effectiveness of the proposed solution [32]. This mapping process is particularly important in the case of older workers since their physical and psychological characteristics are extremely specific compared to those of younger workers. Hence, any health promotion intervention directed at this age group must be matched appropriately with their particular needs and characteristics [23]. This paper, therefore, reports the findings of a systematic literature review of the studies available in the literature focusing on digital health coaching programs for older workers that were developed using the user-centered design approach.

Methods

Aims

This systematic literature review aimed to identify and synthesize published literature focusing on the efficacy of digital health coaching interventions specifically designed for older workers in transition to retirement (or those who just retired) and that used the user-centered design approach. To this purpose, the literature search aimed to answer the following research questions: (1) Which digital coaching interventions are effective or not effective for older employees in transition to retirement or just retired? (2) To what extent are the digital coaching interventions effective for improving the well-being of older employees in transition to retirement (or just retired)? (3) To what extent can this kind of intervention help to prevent poor health?

Eligibility Criteria

Study inclusion criteria were (1) targeting workers aged 50 or older according to the definition of older workers [2] and expanding the latter's age range of 5 years back to include workers in jobs such as night shifts or assembly line for greater part of the working life, and thus entitled to early retirement in most European and high-income countries [33]; (2) written in English; (3) based on user-centered design methodology to assess the impact of the digital coaching intervention through one or more of the most used methods of user-centered design reported in literature [32], such as task analysis, usability testing, field observation, interviews, questionnaires, focus groups, and randomized controlled trials. Moreover, digital coaching interventions were considered eligible if delivered by computer, smartphone or tablet, website, app or software and without the involvement of a human coach. All kinds of well being–related measures were included in this review. Articles were excluded if they (1) did not meet the eligibility criteria (ie, the population of interest and the use of technologies to deliver digital coaching); (2) did not report empirical findings.

Search Strategy

For the literature search, the following databases were used: PubMed, Web of Science, Scopus, and IEEE. The search strategy was conducted without time limitations. Additional articles were obtained from reference lists of included studies and from Google Scholar. Databases were searched using a combination of specific terms that, in the opinion of the authors, were closely related to the topic, such as *older adult workers*, *older adults in retirement*, *transition workers in retirement*, *older employees*, *employees in retirement*, *digital coach**, *virtual coach**, *well-being*, *user-centered design*, *virtual agent*, *avatar*, *virtual personal assistant*. Multimedia Appendix 1 shows the search strategy in detail. The authors performed a search in each database. For each search, the above mentioned terms were combined by the use of the Boolean *AND* operator, restricting the results to articles that contained all the search terms [34,35].

Data Collection and Extraction

According to the predefined criteria, all searches were conducted during April 2019 and May 2019 and then screened independently by the first and the last author. This screening phase was based on analysis of the title and abstract. In August 2019, full-papers of the screened publications were reviewed independently by the first and the last author. Any disagreement was resolved by including the second author in order to reach consensus for all the articles included.

Quality Review

The Mixed Method Appraisal Tool (MMAT; version 2018), developed by Hong et al [36], was used to appraise the quality of the selected studies. In particular, MMAT provides objective tools to rigorously appraise the methodological quality of different categories of studies. The first and third authors independently appraised the methodological quality of each study; the results of each appraisal were compared and any disagreements were solved by including the second author and through discussion among the authors. A quantitative appraisal score was calculated by applying the scoring system proposed by Pluye et al [37] where the presence or absence of criteria may be scored 1 (yes) and 0 (no), respectively. Thereafter, a quality score can be calculated as a percentage: (number of yes responses divided by the number of appropriate criteria) × 100.

Results

General

The search identified 3265 papers, and after duplicates were removed, there were 1341 articles for the initial screening. In accordance with Eden et al [38], the authors carried out the first step of the screening process based on review of titles and abstracts simultaneously. This process resulted in 277 papers. (1064 papers were excluded.) Subsequently, the authors carried out a second screening based on abstract review that resulted in 43 articles (234 papers were excluded) for the full-text selection. The abstracts were reviewed by at least 2 of the 3 researchers (VS, SS, and BD'A), and any disagreements were discussed. In cases where a resolution could not be reached, the third researcher made the final decision. This screening process based on a double review (title and abstract, abstract) allowed the authors to exclude books, books chapters, conference proceedings, or articles that did not meet the eligibility criteria. Full-text papers were downloaded for these records. From this process, the majority (n=41) were excluded because they recruited both young and old employees (n=23), the digital coaching intervention was purely a tracking or monitoring device (n=9), the empirical frameworks were not in line with user-centered design process (n=7), and 2 articles were not available; therefore, 2 papers were included [23,39]. The literature search process is shown in Figure 1.

The specifications and the main findings of the 2 studies matching the inclusion criteria, and thus included in the review, are reported and described in Table 1.

Figure 1. PRISMA flow diagram.

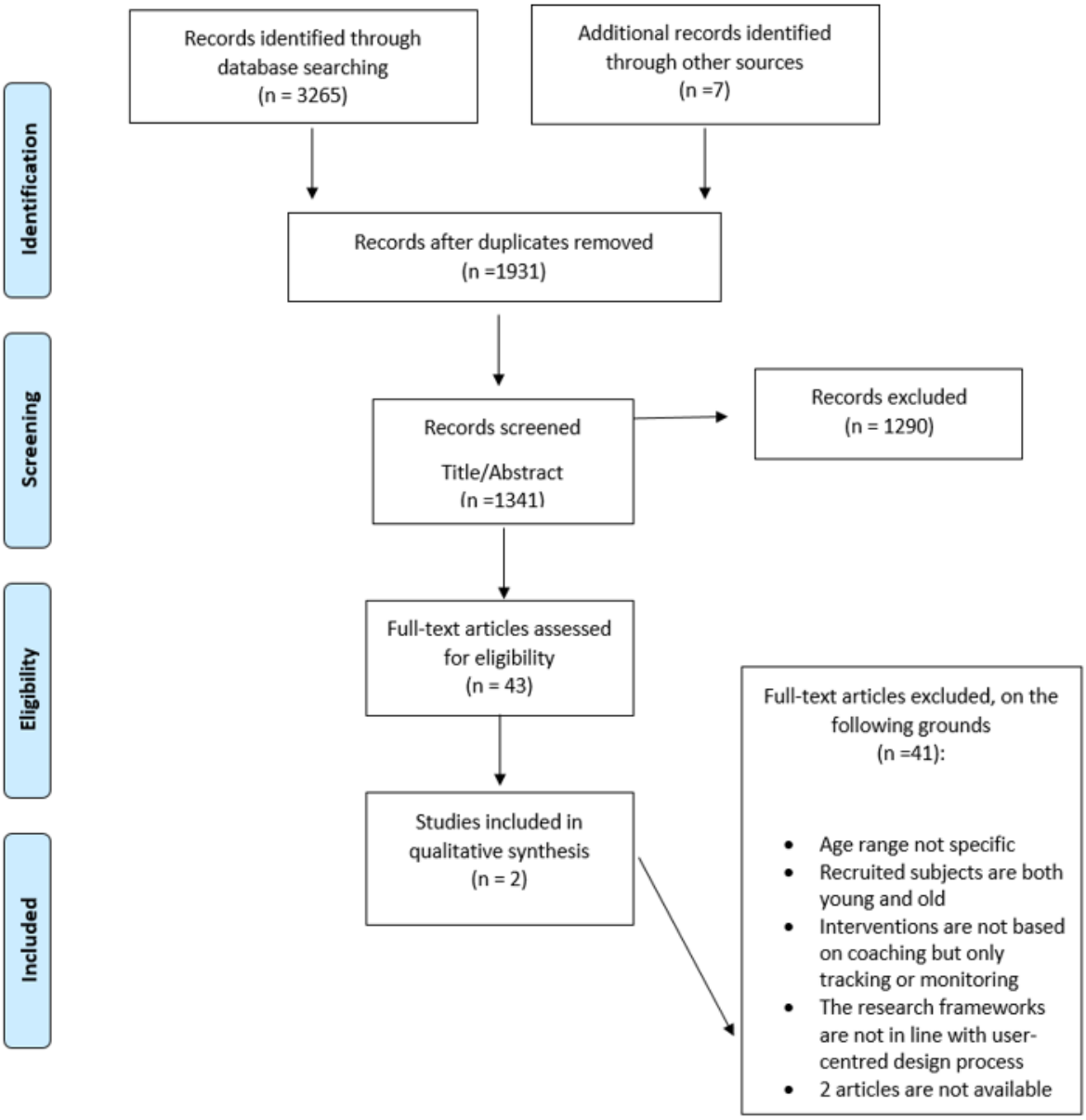


Table 1. Specifications and main findings of the studies included in the systematic literature review.

Reference	Purpose	Method and data collection	Sample/country	Outcomes and measures	Findings
Cook et al [23]	To evaluate the effectiveness of a web-based health program including physical activity, healthy eating, stress management, and tobacco cessation aimed specifically at older workers	Randomized controlled trial; an online survey before and 3 months after the program access	n=278 older adult employees aged 50-68, United States	1. Symptoms of distress 2. Coping with stress 3. Diet outcome expectancies 4. Barriers to a healthy diet 5. Eating practices 6. Overeating self-efficacy 7. Diet change self-efficacy 8. Planning healthy eating 9. Weight and body mass index 10. Exercise habits 11. Exercise self-efficacy 12. Self-efficacy for overcoming barriers to exercise 13. Exercise planning 14. Belief about aging 15. Tobacco use	1. Improvement of diet behavioral change self-efficacy, planning healthy eating 2. Improvement of eating practices, exercise self-efficacy, exercise planning, and aging beliefs
Irvine et al [39]	To evaluate the efficacy of a 12-week internet intervention to help sedentary older adults over 55 years of age adopt and maintain an exercise regimen	Randomized controlled trial; online survey at pretest, at 12 weeks, and at 6 months after program fruition follow-up	n=368 sedentary adults >55 years of age, United States	1. Cardiovascular activities 2. Stretching activities 3. Strengthening activities 4. Balance activities 5. Activities (minutes per week) 6. Short Form-12 physical 7. Short Form-12 mental 8. Weight and body mass index 9. Attitudes/knowledge 10. Self-efficacy 11. Behavioral attentions 12. Motivation to exercise 13. Ability to exercise 14. Barrier to exercise	1. The multivariate model indicated significant treatment effects at posttest and at 6 months 2. Improvement on 13 of 14 outcome measures 3. At 6 months, treatment participants maintained large gains compared to control on all 14 outcome measures.

Quality Review

After calculating the score for each article [37], we synthetized methodological quality results in 3 different categories: low score, <35%; medium score, 36% to 70%; and high score, 71% to 100%.

The selected articles reached a high quality score (100%) as shown in Table 2.

Table 2. Quality score of the selected studies.

Reference	Type of study	Screening score	Randomized controlled clinical trial score	Total score	Appropriate criteria, n	Quantity score, %	Score category
Cook et al [23]	Randomized controlled clinical trial	2	5	7	7	100	High
Irvine et al [39]	Randomized controlled clinical trial	2	5	7	7	100	High

Digital Program Design

The included studies were both digital interventions developed by a multidisciplinary team that focused on a significant user-centered design approach. HealthyPast50 by Cook et al [23] was a multimedia web-based program that contained information and guidance on healthy aging, diet, physical activity, stress management, and tobacco use. The program was developed specifically for adults over 50 years of age through multiple cycles of development and testing that involved older

workers providing feedback on the initiative and rating the prototype content until the final realization. HealthyPast50 was based on a content management system and provided ample graphics, audio, and video contents.

Active After 55 by Irvine et al [39] was an internet-enabled CD-ROM program aimed at boosting the functional ability, mobility, and physical activity of older adults in endurance, stretching, strengthening, and balance enhancement via text and video messages. The program was developed through consultations with professionals experienced in the design and implementation of research-based exercise programs for older adults.

Methodology for Testing and Assessing the Effectiveness of Digital Technology

The effectiveness of these studies was assessed through a randomized controlled trial and with online surveys in different phases of the assessment. The HealthyPast50 study was based on multiple outcome measures before and 3 months after the actual intervention in a sample of 278 older employees aged 50 to 68 years. The users were split randomly into a program group able to log-in to the system and a control group that did not access it. The whole sample was characterized by a high educational level and relatively high computer literacy. The program group could access the web-based program at any time during the 3-month test period at work and outside of work.

The Active After 55 study was based on 3 assessments: pretest, postintervention (12 weeks after pretest), and a 6-month follow-up, with a sample of 368 sedentary adults over 55 years of age. Authors reported that participants tended to be employed, educated, and frequent computer users with at least a middle-class income. Participants were automatically randomized into a treatment group that used the digital health program or a control group that did not have access to it. The intervention consisted in an initial 1-hour session with the Active After 55 program to obtain assistance in designing a personalized plan to follow and then 11 weekly sessions lasting at least 15 minutes in addition to weekly exercises. Both studies used an online survey to gather data in the different assessment phases.

Main Findings of the Selected Studies

In both randomized controlled trials, the intervention group showed significant improvement in healthy behaviors compared that shown by the control group.

In the case of HealthyPast50, the working adults who used the web-based program manifested relevant enhancements over the 3-month test period on 3 out of 15 outcomes: diet behavior change, self-efficacy, and planning of healthy eating. Working adults who were given access to the web-based HealthyPast50 program showed significantly greater improvement on key health constructs over the 3-month test period compared to that shown by the individuals in the control group. In the analysis of the imputed data set, the intervention group performed significantly better than the control group on diet behavior change, self-efficacy, planning healthy eating, and mild exercises. Moreover, there were improvements in eating practices, moderate exercise, and overall exercise but these did not meet the threshold for statistical significance. As for the Active After 55 program, the hypotheses were that the intervention would be linked to changes in the exercise domains of endurance, stretching, strengthening, and balance and that it would be linked to theoretically relevant mediators of behavior change. The treatment group showed significant improvement on 13 out of 14 outcomes at posttest and on all 14 outcomes at the follow-up. The findings were consistent across an array of measures, with a large multivariate effect size at posttest and a medium multivariate effect size at the 6-month follow-up.

Discussion

Principal Findings

This systematic literature review was aimed at exploring the state of the art of research in the field of digital health programs sustaining the well-being of older workers in transition to retirement (or just retired) and using a user-centered design approach. Despite the wide-reaching search and review, only 2 published papers met the inclusion criteria. We provided a qualitative methodological appraisal of the 2 selected articles by means of the MMAT. Every year, a growing number of information and communication technologies emerges with the aim to provide innovative and efficient ways to help older adults in their daily life and to reduce the cost of health care. Nevertheless, there is still a paucity of studies testing the efficacy of such technologies, both for older adults in general and for older workers in transition to retirement. Hence, there is a need for more research in this field. The 2 studies included in the review strengthen the idea that multimedia web-based and internet-enabled CD-ROM programs can be effective in promoting healthy life style behaviors and in preventing poor health condition among older workers near retirement.

The findings of this systematic literature review suggest that digital health programs may help older workers to improve their health and well-being by motivating them to engage in healthy behaviors (eg, healthy diet, physical activity, and less tobacco use). Cook et al [23] and Irvine et al [39], indeed, used mainly physiological and behavioral measures for outcome assessment (eg, cardiovascular activity, body mass index, eating practices), in compliance with a concept of health as an individual physical issue. Conversely, the social, psychological, and relational [17] aspects of aging as well as the cultural patterns [18] of the participants in the studies seemed to not be monitored, and hence, were underestimated. Within the contextual factors neglected by the 2 studies, there is also the transition from work to retirement. During this period, older workers may be required to face age-related stereotypes, to upgrade their digital skills [12,17], and (as retirees) may be asked to reorganize their daily routine and their role within the family. Additionally, they are confronted with potential effects from society on health and general well-being (eg, social marginalization, anxiety, and concerns) [18,20]. In light of the above, future studies based on digital coaching programs should adopt the healthy aging paradigm [11] by adding measures for monitoring social, psychological, and cultural factors as determinants of healthy aging that are in the background of the retirement process and can influence it. This monitoring should be carried out

throughout the whole study life-cycle, from the identification of the individuals' health needs up to the development of the technology and the eHealth plan intervention. This would allow researchers to analyze the data and interpret the findings concerning users' experience with the technology as well as with aging and retirement, from a multidimensional perspective.

A major advantage of eHealth interventions is that such interventions are easily accessible 24/7 and are usable by individuals who may not have access to traditional health promotion otherwise. When these interventions are developed around the real needs, capabilities, and limitations of older workers and tested with rigorous research methodology, the technology-based interventions demonstrate the added value of innovation in coping with healthy aging.

This idea and the review of the 2 papers inspired some suggestions for further investigation: they should use a user-centered approach for designing the technology; test innovative research methodologies; and adopt new technical solutions for high-quality interaction design.

First, digital innovation technologies can be a promising way to cope with the health and well-being deteriorations of older adults in transition from work to retirement, especially when they are designed, developed, and assessed through a user-centered approach. In fact, similar to the approach for health programs that are conducted by a person, it is preferable to adopt a patient-centered approach. Likewise, in health digital programs, it is advisable to involve patients in planning the program and designing the technology, in order to intercept their health needs and translate them into the technology requirements. This might allow patients to set health goals and program phases together with the digital coach just as they would with a human coach, thus increasing adherence to the healthy path [25]. Conversely, the marginal involvement of end users, experts, and stakeholders (who represent the articulated interdependency among individuals, organizational levels, and technological factors), along the entire process from the end-users' needs assessment, through the identification of the technology requirements up to the realization of the product remains one of the major causes of misuse of technologies and a confounding factor in the assessment of systems [40,41]. Even though Cook et al [23] and Irvine et al [39] adopted a user-centered design approach, it seems that they limited this approach to the definition of the contents of the health program and did not use it for the development of the digital solution. In the studies [23,39], a robust method was adopted to evaluate the effectiveness of interventions using randomized controlled trials and online surveys, based on a list of standardized tests, and repeated measurements, such as pretest and posttest (before and 3 months after the end of the HealthyPast50 study and at 12 weeks and 6 months after actual program follow-up in the case of the Active After 55 intervention). Despite the significance of the randomized controlled trial approach, both studies did not provide evidence on the acceptability, usability, and learnability nor on the utility of the web-based interventions. In order to fully address these dimensions, the use of mixed methods is recommended: qualitative open-ended questions in addition to quantitative measures would be useful for capturing the perspective and the personal experiences of the users interacting with the technology.

This observation leads us to the second key point to take into consideration for future studies (ie, the use of innovative and sophisticated research methodologies). In fact, despite increasing research on the impact of technology on older people's health and well-being, most studies on this issue used small samples which neither allowed for randomized controlled trials nor generalization and transferability to other domains. Moreover, a 3-arm randomized controlled trial (ie, including one intervention group using only human coaching, a second intervention group using only digital coaching, and a control group) could offer a real framework for future research in this field. Moreover, studies comparing virtual coaching programs to web-based programs are necessary to assess which approach is more effective and which has a higher level of acceptability by older employees.

The third key point concerns the use of a variety of technical solutions to deliver digital coaching programs addressing older adults. Different workplace interventions have been developed to improve workers' health and well-being using web-based interventions [29] and mobile apps [42]. Nevertheless, the use of embodied conversational agents, and in particular, artificial intelligence virtual coaches (ie, computer software specifically designed to work and act like a human) seems to be particularly appropriate to influence user attitudes or behaviors [43]. The advantage of this virtual coaching is almost similar to an in-person health coach offering self-management through personalized guidance and support available at any time and in any place.

The fourth consideration arising from this systematic literature review concerns the role of positive user experience, measuring people's feelings on interacting technology in a particular context [44,45]. It is common knowledge that if the use of technology entails a negative experience for the user, as a consequence, this leads to the rejection of the system itself [46]. Therefore, it is crucial to design a high-quality interaction involving the intended users in order to identify their needs and derive the technical requirements capable of best meeting these needs [47].

Limitations

The search was performed on 4 databases (ie Scopus, Web of Science, PubMed, and IEEE) accessed during a specific period of time (April 2019 and May 2019); in order to be more inclusive, we added references selected from Google Scholar. Given this procedure, our search could be not exhaustive and, unknowingly and unintentionally, some papers may have been omitted. Another limitation could be the keywords used for search and their combination with the Boolean *AND* operator. The selection of the terms with the Boolean operator are the results of the authors' knowledge, which is not exhaustive. Furthermore, the narrow eligibility criteria, requiring a focus on older adults in retirement, reduced the number of studies selected.

Comparison With Prior Work

To our best knowledge, no other reviews of digital programs are reported in the literature nor are there any on digital programs based on virtual coaching techniques targeting older workers near retirement. We could, therefore, only compare the 2 studies reviewed with literature that would be desirable for increasing knowledge in this field.

First, the reviewed studies were aimed at promoting healthy behaviors, limiting the onset of chronic diseases, and limiting the health-related costs for companies, thereby improving older worker productivity. The technologies tested by these studies were aimed at affecting the individuals' behaviors toward physical activity, stress management, diet, and tobacco use. Future digital programs addressing older workers who are close to retirement should instead take into account not only physical health aspects related to aging itself but also psychological and social aspects that can be influenced by the transition from work to retirement. This goal could be reached by encouraging older workers to adopt strategies to counteract the side effects of retirement, such as the loss of social contacts, a sedentary lifestyle, and a decrease in intellectual stimuli. These strategies might lead older individuals onto a path of awareness of the existential change they are experiencing; to adopt a different planning of the day in order to avoid feelings of usefulness; to have attitudes more oriented to the community for the construction of bonds and friendships replacing those lost when they left work (eg, volunteering). Furthermore, in order to ensure a social dimension in the training, several individuals might share the same digital health coach. The latter might promote interpersonal relationships by helping individuals meet in person for a walk or for a social event, thereby bridging the gap between virtual and real dimensions.

Conclusions

Alone, the recent retirement policies that have postponed retirement age cannot cope with the growing number of workers who ask to retire early due to age-related chronic diseases and multimorbidity. Moreover, workers close to retirement may need to be supported to better face this existential change.

Several studies demonstrated that technology can be very effective in promoting healthy lifestyles among older workers [23,29,39], but they often neglected the psychological and relational aspects of working in old age. Conversely, healthy aging is a multidimensional concept [11] which future research studies aiming at developing and testing digital health programs targeting older workers and retirees should refer to. A healthy aging model could be considered as a compass orienting the choice of multiple health dimensions for improvement by means of technology, including social engagement, mental health, and cultural patterns, to consider during identification of the user' needs, technical requirements identification, and design of eHealth interventions.

The results of this systematic literature review demonstrate the difficulties of assessing the efficacy of digital coaching in itself and for older employees. While the 2 papers [23,39] suggest that workplace digital health programs show interesting results for improving various aspects of older employees' well-being, the literature around this issue remains in an embryonic state. This gap needs to be filled by future research studies adopting a user-centered approach, innovative methodologies for assessing the technology effectiveness, new technical solutions, and high-quality interaction designs.

Acknowledgments

This research was cofunded by Active and Assisted Living Program (reference no. aal-2018-5-92-CP). This study was also partially supported by Ricerca Corrente funding from the Italian Ministry of Health to *Istituto di Ricovero e Cura a Carattere ScientificoIstituto Nazionale Ricovero e Cura per Anziani* (Scientific Institute for Research, Hospitalization, and Healthcare National Institute of Health and Science on Aging).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords and boolean operator (AND) applied in the 4 searches used in each database.
[DOCX File , 14 KB - jmir_v22i9e17809_app1.docx]

References

1. World Population Prospects 2019. Department of Economic and Social Affairs Population Dynamics. URL: https://population.un.org/wpp/Publications/Files/WPP2019_DataBooklet [accessed 2019-09-18]
2. McCarthy J, Heraty N, Cross C, Cleveland J. Who is considered an 'older worker'? extending our conceptualisation of 'older' from an organisational decision maker perspective. Human Resource Management Journal 2014 May 27;24(4):374-393 [FREE Full text] [doi: 10.1111/1748-8583.12041]
3. Employment statistics. Eurostat. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Employment_statistics#Employment_rates_by_sex.2C_age_and_educational_attainment_level [accessed 2019-09-21]

4. Employment rate: aged 55-64: all persons for the United States. FRED Federal Reserve Bank of St Louis. URL: https://fred.stlouisfed.org/series/LREM55TTUSM156S [accessed 2020-06-24]
5. Wang M, Shultz KS. Employee retirement: a review and recommendations for future investigation. Journal of Management 2009 Oct 06;36(1):172-206. [doi: 10.1177/0149206309347957]
6. Carone G, Eckefeldt P, Giamboni L, Laine V, Pamies-Sumner S. Pension reforms in the EU since the early 2000's: achievements and challenges ahead. discussion paper 42. European Economy Discussion Papers 2017;42:1-64 [FREE Full text] [doi: 10.2139/ssrn.2964933]
7. Zacher H, Kooij D, Beier M. Successful aging at work: empirical and methodological advancements. Work, Aging and Retirement 2018 Apr:123-128 [FREE Full text] [doi: 10.1093/workar/way002]
8. Rowe JW, Kahn RL. Human aging: usual and successful. Science 1987 Jul 10;237(4811):143-149. [doi: 10.1126/science.3299702] [Medline: 3299702]
9. Kleineidam L, Thoma MV, Maercker A, Bickel H, Mösch E, Hajek A, et al. What is successful aging? a psychometric validation study of different construct definitions. Gerontologist 2019 Jul 16;59(4):738-748. [doi: 10.1093/geront/gny083] [Medline: 30016435]
10. Pike E. The agenda of the active ageing, old folk devils and a new moral panic. Sociology of Sport Journal 2011;28(2):209-225. [doi: 10.1123/ssj.28.2.209]
11. World report on aging and health. World Health Organization. Geneva, Switzerland; 2015 Sep 30. URL: https://www.who.int/ageing/events/world-report-2015-launch/en/ [accessed 2020-09-01]
12. Harris K, Krygsman S, Waschenko J, Laliberte Rudman D. Ageism and the older worker: a scoping review. Gerontologist 2018 Mar 19;58(2):e1-e14. [doi: 10.1093/geront/gnw194] [Medline: 28082278]
13. Shiu E, Hassan LM, Parry S. The moderating effects of national age stereotyping on the relationships between job satisfaction and its determinants: a study of older workers across 26 countries. Brit J Manage 2015 Feb 11;26(2):255-272. [doi: 10.1111/1467-8551.12091]
14. Gaillard M, Desmette D. (In)validating stereotypes about older workers influences their intentions to retire early and to learn and develop. Basic and Applied Social Psychology 2010 Feb 23;32(1):86-98. [doi: 10.1080/01973530903435763]
15. Ruggs EN, Hebl MR, Singletary Walker S, Fa-Kaji N. Selection biases that emerge when age meets gender. Journal of Managerial Psych 2014 Nov 04;29(8):1028-1043. [doi: 10.1108/jmp-07-2012-0204]
16. Finkelstein LM, Ryan KM, King EB. What do the young (old) people think of me? content and accuracy of age-based metastereotypes. European Journal of Work and Organizational Psychology 2013 Dec;22(6):633-657. [doi: 10.1080/1359432x.2012.673279]
17. Lunau T, Wahrendorf M, Dragano N, Siegrist J. Work stress and depressive symptoms in older employees: impact of national labour and social policies. BMC Public Health 2013 Nov 21;13(1):1086 [FREE Full text] [doi: 10.1186/1471-2458-13-1086] [Medline: 24256638]
18. Principi A, Santini S, Socci M, Smeaton D, Cahill KE, Vegeris S, et al. Retirement plans and active ageing: perspectives in three countries. Ageing and Society 2016 Aug 22;38(1):56-82. [doi: 10.1017/s0144686x16000866]
19. Principi A, Smeaton D, Cahill K, Santini S, Barnes H, Socci M. What happens to retirement plans, and does this affect retirement satisfaction? Int J Aging Hum Dev 2020 Mar;90(2):152-175 [FREE Full text] [doi: 10.1177/0091415018796627] [Medline: 30185060]
20. Weiss R. The Experience of Retirement. Ithaca, New York: ILR/Cornell University Press; 2005.
21. Socci M, Casanova G, Romanini A, Principi A. Policies and services that foster social participation of older people in poor health condition. Mobilising the Potential of Active Ageing in Europe. Italy; 2015. URL: https://mopact.group.shef.ac.uk/wp-content/uploads/2013/10/D5.4-Country-report-Italy.pdf [accessed 2019-11-12]
22. Magnavita N. Obstacles and future prospects: considerations on health promotion activities for older workers in Europe. Int J Environ Res Public Health 2018 May 28;15(6):1096 [FREE Full text] [doi: 10.3390/ijerph15061096] [Medline: 29843423]
23. Cook RF, Hersch RK, Schlossberg D, Leaf SL. A web-based health promotion program for older workers: randomized controlled trial. J Med Internet Res 2015 Mar 25;17(3):e82 [FREE Full text] [doi: 10.2196/jmir.3399] [Medline: 25830503]
24. Wolever RQ, Simmons LA, Sforzo GA, Dill D, Kaye M, Bechard EM, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. Glob Adv Health Med 2013 Jul;2(4):38-57. [doi: 10.7453/gahmj.2013.042] [Medline: 24416684]
25. Neuner-Jehle S, Schmid M, Grüninger U. The "Health Coaching" programme: a new patient-centred and visually supported approach for health behaviour change in primary care. BMC Fam Pract 2013 Jul 17;14:100 [FREE Full text] [doi: 10.1186/1471-2296-14-100] [Medline: 23865509]
26. Bodenheimer T, Willard-Grace R, Ghorob A. Expanding the roles of medical assistants: who does what in primary care? JAMA Intern Med 2014 Jul;174(7):1025-1026. [doi: 10.1001/jamainternmed.2014.1319] [Medline: 24820220]
27. Barker A, Cameron P, Flicker L, Arendts G, Brand C, Etherton-Beer C, et al. Evaluation of RESPOND, a patient-centred program to prevent falls in older people presenting to the emergency department with a fall: A randomised controlled trial. PLoS Med 2019 May;16(5):e1002807 [FREE Full text] [doi: 10.1371/journal.pmed.1002807] [Medline: 31125354]

28. Hibbard J, Gilburt H. Supporting people to manage their health: an introduction to patient activation. London: The King's Fund; 2014 Jun 15. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/supporting-people-manage-health-patient-activation-may14.pdf [accessed 2020-09-01]
29. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018;4:2055207618770861. [doi: 10.1177/2055207618770861] [Medline: 29942631]
30. Sasha Corporation. The 2019 Sherpa Executive Coaching Survey. Cincinnati, OH: Sasha Corporation; 2019.
31. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. ISO 9241-210:2010. 2010. URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075 [accessed 2020-06-16]
32. Czaja S, Boot W, Charness N, Rogers W. Individual differences. In: Designing for older adults: Principles and creative human factors approaches. 3rd edition. Boca Raton, Florida: CRC press; 2019.
33. Retirement regimes for workers in arduous or hazardous jobs in Europe - a study of national policies. European Commission. 2016. URL: http://praha.vupsv.cz/fulltext/ul_1998.pdf
34. Cronin P, Ryan F, Coughlan M. Undertaking a literature review: a step-by-step approach. British Journal of Nursing 2008 Jan;17(1):38-43. [doi: 10.12968/bjon.2008.17.1.28059]
35. Branley D. Doing a literature review. In: Seale C, editor. Researching Society and Culture 3rd edition. London: Sage; 2012:77-96.
36. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F. Cargo M Mixed Methods Appraisal Tool (MMAT), version 2018. Registration of Copyright (#1148552). URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2020-06-22]
37. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
38. Eden J, Levit L, Berg A, Morton S. Finding what works in health care: standards for systematic reviews. Institute of Medicine of the National Academies. 2011. URL: https://www.nihlibrary.nih.gov/sites/default/files/Finding_What_Works_in_Health_Care_Standards_for_Systematic_Reviews_IOM_2011.pdf [accessed 2020-06-18]
39. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]
40. van GJEWC, Nijland N, van LM, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
41. Kip H, van GL. Holistic development of eHealth technology. In: van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R. editors. eHealth Research, Theory and Development. A MultiDisciplinary Approach. London: Routledge; 2018:131-166.
42. de Korte EM, Wiezer N, Janssen JH, Vink P, Kraaij W. Evaluating an mHealth App for Health and Well-Being at Work: Mixed-Method Qualitative Study. JMIR Mhealth Uhealth 2018 Mar 28;6(3):e72 [FREE Full text] [doi: 10.2196/mhealth.6335] [Medline: 29592846]
43. van WA, Vinkers C, van HA. Developing a virtual coach for chronic patients: a user study on the impact of similarity, familiarity and realism. In: Proceedings of the 11th International Conference on Persuasive Technology. 2016 Presented at: 11th International Conference on Persuasive Technology; April 2016; Austria p. 263-265. [doi: 10.1007/978-3-319-31510-2_23]
44. Hartson R, Pyla P. The UX Book: Process and Guidelines for Ensuring a Quality User Experience. San Francisco, CA: Morgan Kaufmann Publishers Inc; 2012.
45. Hassenzahl M. User experience design. In: Soegaard M, Dam RF, editors. The Encyclopedia of Human-Computer Interaction. 2nd ed. Aarhus, Denmark: The Interaction Design Foundation Publisher; 2013.
46. Alenljung B, Lindblom J, Andreasson R, Ziemke T. User experience in social human-robot interaction. In: Rapid Automation: Concepts, Methodologies, Tools, and Applications. Hershey, PA: IGI Global Rapid Automation; 2019:1468-1490.
47. Robinson J, Lanius C, Weber R. The past, present, and future of UX empirical research. Commun Des Q Rev 2018 Feb 16;5(3):10-23. [doi: 10.1145/3188173.3188175]

Abbreviations

MMAT: Mixed Methods Appraisal Tool

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by M Lozano-Lozano, A Ferrario; comments to author 12.06.20; revised version received 10.07.20; accepted 27.07.20; published 24.09.20.

Please cite as:
Stara V, Santini S, Kropf J, D'Amen B
Digital Health Coaching Programs Among Older Employees in Transition to Retirement: Systematic Literature Review
J Med Internet Res 2020;22(9):e17809
URL: http://www.jmir.org/2020/9/e17809/
doi:10.2196/17809
PMID:32969827

Review

Internet and Computer-Based Cognitive Behavioral Therapy for Anxiety and Depression in Adolescents and Young Adults: Systematic Review and Meta-Analysis

Carolien Christ[1,2,3], MSc; Maria JE Schouten[2], MSc; Matthijs Blankers[2,4], PhD; Digna JF van Schaik[1,3], MD, PhD; Aartjan TF Beekman[1,3], MD, PhD; Marike A Wisman[5], MSc; Yvonne AJ Stikkelbroek[6,7], PhD; Jack JM Dekker[2,8], PhD

[1]Department of Psychiatry, GGZ inGeest, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Department of Research, Arkin Mental Health Care, Amsterdam, Netherlands

[3]Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[4]Trimbos institute-The Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

[5]Department of Youth and Family, Arkin Mental Health Care, Amsterdam, Netherlands

[6]Department of Child and Adolescent Studies, Universiteit Utrecht, Utrecht, Netherlands

[7]Depression Expert Center for Youth, Mental Health Care Oost-Brabant, Boekel, Netherlands

[8]Department of Clinical Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Carolien Christ, MSc
Department of Psychiatry
GGZ inGeest
Amsterdam UMC, Vrije Universiteit Amsterdam
Oldenaller 1
Amsterdam, 1081 HJ
Netherlands
Phone: 31 20 590 1339
Email: c.christ@ggzingeest.nl

Abstract

Background: Anxiety and depressive disorders are prevalent in adolescents and young adults. However, most young people with mental health problems do not receive treatment. Computerized cognitive behavior therapy (cCBT) may provide an accessible alternative to face-to-face treatment, but the evidence base in young people is limited.

Objective: The objective was to perform an up-to-date comprehensive systematic review and meta-analysis of the effectiveness of cCBT in treating anxiety and depression in adolescents and young adults compared with active treatment and passive controls. We aimed to examine posttreatment and follow-up effects and explore the moderators of treatment effects.

Methods: We conducted systematic searches in the following six electronic databases: PubMed, EMBASE, PsycINFO, CINAHL, Web of Science, and Cochrane Central Register of Controlled Trials. We included randomized controlled trials comparing cCBT with any control group in adolescents or young adults (age 12-25 years) with anxiety or depressive symptoms. The quality of included studies was assessed using the Cochrane risk-of-bias tool for randomized trials, version 2.0. Overall quality of evidence for each outcome was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. Posttreatment means and SDs were compared between intervention and control groups, and pooled effect sizes (Hedges *g*) were calculated. Random-effects meta-analyses were conducted using Comprehensive Meta-Analysis software. Subgroup analyses and meta-regression analyses were conducted to explore whether age, guidance level, and adherence rate were associated with treatment outcome.

Results: The search identified 7670 papers, of which 24 studies met the inclusion criteria. Most included studies (22/24) had a high risk of bias owing to self-report measures and/or inappropriate handling of missing data. Compared with passive controls, cCBT yielded small to medium posttreatment pooled effect sizes regarding depressive symptoms (g=0.51, 95% CI 0.30-0.72, number needed to treat [NNT]=3.55) and anxiety symptoms (g=0.44, 95% CI 0.23-0.65, NNT=4.10). cCBT yielded effects similar to those of active treatment controls regarding anxiety symptoms (g=0.04, 95% CI −0.23 to 0.31). For depressive symptoms, the

nonsignificant pooled effect size favored active treatment controls (g=−0.70, 95% CI −1.51 to 0.11, P=.09), but heterogeneity was very high (I^2=90.63%). No moderators of treatment effects were identified. At long-term follow-up, cCBT yielded a small pooled effect size regarding depressive symptoms compared with passive controls (g=0.27, 95% CI 0.09-0.45, NNT=6.58). No other follow-up effects were found; however, power was limited owing to the small number of studies.

Conclusions: cCBT is beneficial for reducing posttreatment anxiety and depressive symptoms in adolescents and young adults compared with passive controls. Compared with active treatment controls, cCBT yielded similar effects regarding anxiety symptoms. Regarding depressive symptoms, however, the results remain unclear. More high-quality research involving active controls and long-term follow-up assessments is needed in this population.

Trial Registration: PROSPERO CRD42019119725; https://tinyurl.com/y5acfgd9.

(J Med Internet Res 2020;22(9):e17831) doi:10.2196/17831

KEYWORDS

cognitive behavior therapy; internet; anxiety; depression; youth; meta-analysis

Introduction

Anxiety and depressive disorders are common in children and adolescents [1,2]. Symptoms of anxiety and depression in childhood and adolescence predict a range of mental health problems later in life, including adult anxiety and depressive disorders and substance use disorders [3-8]. Moreover, anxiety and depressive disorders in young people are associated with an increased risk of self-harm and suicide [5,9], the second most common cause of death among youth aged between 10 and 24 years [10].

Among children and adolescents aged up to 18 years, global prevalence rates are estimated at 6.5% for anxiety disorders and 2.6% for depressive disorders [11]. The prevalence of mental disorders increases during the transition from childhood to adolescence [12,13], with prevalence rates in adolescents (ie, age 12-19 years) estimated at 10.7% for anxiety disorders and 6.1% for depressive disorders [13]. The incidence of most anxiety disorders peaks during adolescence, whereas the incidence of depressive disorders starts to rise during adolescence [14] and peaks in young adulthood (ie, age 19-25 years) [15-17]. Given the high incidence and burden of anxiety and depressive disorders in young people, early intervention in both adolescents and young adults is of utmost importance.

Adolescents and young adults with anxiety or depressive disorders are commonly treated with cognitive behavioral therapy (CBT), which is a widely-used treatment that has been proven to be effective in this population [18-21]. However, the majority of adolescents and young adults with mental health problems do not receive treatment [22-25]. Among their reasons for low treatment utilization are limited availability of youth mental health services, perceived stigma associated with mental illness, perceived lack of time or resources, and preference for self-help [24,26,27]. These barriers to treatment utilization may partly be overcome by computerized mental health interventions involving psychological treatment delivered via the internet and/or digital devices. Compared with face-to-face treatment, computerized interventions may provide more flexible access in terms of time, location, and availability; greater privacy and anonymity; and more independence [28,29]. The internet is ubiquitous in the lives of young people, who have shown positive attitudes toward computerized mental health interventions [30]. Therefore, computerized treatment provides an accessible and feasible alternative to face-to-face treatment for this group [31,32].

Numerous randomized controlled trials (RCTs) and meta-analyses in adult populations with anxiety and depressive disorders have shown that CBT may be effectively delivered via the internet or digital devices [33,34]. The effects of these so-called computerized CBT (cCBT) interventions have been demonstrated to be comparable to the effects of face-to-face CBT in adults [35]. In children and young people, cCBT has been found to be effective in treating mental health problems as well [36-42]. Despite these promising results, however, the evidence base on cCBT in young people remains limited compared with research in adults. The number of studies is still small, and the quality of RCTs is often low [39].

To date, three meta-analyses have shown cCBT [36,37] and internet-based mental health interventions [43] to be effective in reducing anxiety and depressive symptoms in young people aged 12 to 25 years. Ebert et al [36] found cCBT to be superior to passive control conditions for both anxiety (g=0.68, 95% CI 0.45-0.92, P<.001; k=7) and depression (g=0.76, 95% CI 0.23-2.66, P<.001; k=4). Active control conditions were not included in their meta-analysis. Similarly, Pennant et al [37] found positive effects of cCBT on both anxiety (standardized mean difference [SMD]=−0.77, 95% CI −1.45 to −0.09, k=6) and depression (SMD=−0.62, 95% CI −1.13 to −0.11, k=7) compared with passive controls. Compared with face-to-face CBT, their meta-analysis showed similar effects for cCBT on anxiety (SMD=−0.04, 95% CI −0.36 to 0.28, P=.89; k=3), but a large effect in favor of face-to-face CBT on depression (SMD=1.65, 95% CI 0.88-2.41, P<.001; k=2). However, these meta-analyses included only a small number of studies that were all published up to 2013. A more recent meta-analysis in children and adolescents up to 18 years with depressive and/or anxiety symptoms showed that cCBT interventions yielded a medium effect size compared with waiting list controls (g=0.66, 95% CI 0.42-0.90, P<.001, k=17) [44]. This study reported neither separate effects of cCBT on depression and anxiety symptoms nor effects of cCBT compared with face-to-face CBT. Importantly, none of these meta-analyses reported mid-term or long-term effects [36,37,44].

To our knowledge, the study of Välimäki et al [43] is the only meta-analysis that not only reported posttreatment effects, but also described short-term and long-term follow-up effects. Posttreatment effects showed the positive effects of internet-based interventions on depressive symptoms (P=.02, median=1.68, 95% CI 0.25-3.11, k=10) and anxiety symptoms (P=.001, median=1.47, 95% CI 0.59-2.36, k=8) compared with any control group. The authors found significant long-term effects of internet-based interventions aimed at reducing depressive symptoms 6 months after treatment (P=.01, median=1.78, 95% CI 0.37-3.20, k=3), but no mid-term effects (ie, 3-5 months after treatment). Regarding anxiety symptoms, they found no mid-term effects in the only two available studies, and no study reported long-term results on anxiety [43]. However, their meta-analysis included both cCBT and various other internet-based mental health interventions (eg, positive psychology), and did not specifically analyze the effects of cCBT. Furthermore, the effects of internet-based interventions were not reported separately compared with active treatment controls and passive controls. In addition, the authors used a narrow search string, which did not include anxiety disorders or interventions aimed at decreasing anxiety symptoms. Hence, it remains unclear whether cCBT is effective in treating young people with anxiety and depressive disorders in the long term, compared with active treatment and passive controls.

In adults, individual participant data meta-analyses on internet-based interventions have demonstrated several predictors of better treatment outcomes, among which are older age [45] and higher treatment adherence [33]. In addition, level of guidance (ie, the level of therapist support provided during cCBT) appears to be associated with larger treatment effects in adults, as studies on guided internet-based interventions have generally demonstrated larger effect sizes than studies on unguided interventions [33,45,46]. Although previous meta-analyses on cCBT in children and young people have attempted to identify moderators of treatment effects, the results remain mixed. Some found evidence for a moderating role of age [36,37,42], whereas others did not [38,44]. With regard to guidance, evidence remains mixed as well [37,42,44]. To our knowledge, previous meta-analyses in young people did not examine whether treatment adherence is associated with cCBT effect sizes.

Research on cCBT in young people with anxiety or depressive symptoms is a rapidly developing field, and all previous meta-analyses are limited to studies of at least 2 years old [36-38,43,44]. In addition, the most recent meta-analyses focused on other age groups [38,41,42,44] or did not separately report effects for either anxiety and depressive disorders [38,44] or cCBT [43]. Moreover, the follow-up effects of cCBT remain largely unknown. Lastly, since most previous meta-analyses in young people did not separately compare cCBT to active treatment and passive controls [36,43], it remains unclear whether cCBT provides an effective alternative to face-to-face treatment in this group. Therefore, our objective was to provide an up-to-date comprehensive systematic review and meta-analysis of the effectiveness of cCBT in treating anxiety and depressive symptoms in adolescents and young adults compared with active treatment and passive control groups, differentiating between posttreatment, short-term follow-up, and long-term follow-up. Furthermore, we aimed to explore whether age, guidance level, and treatment adherence are associated with treatment outcome by conducting subgroup analyses and meta-regression analyses.

Methods

Design

This study was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for reporting systematic reviews and meta-analyses [47]. The systematic review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO, registration number: CRD42019119725).

Search Strategy

We conducted a comprehensive literature search in the following six electronic databases from database inception to September 13, 2019: PubMed, EMBASE, PsycINFO, CINAHL, Web of Science, and Cochrane Central Register of Controlled Trials (CENTRAL). An information specialist was consulted for the search. The search strategy included combinations of relevant medical subject headings and text-based search terms covering computerized (or internet, digital, eHealth, online, smartphone, or web-based), CBT (or cognitive, behavior, therapy, treatment, or intervention), and adolescent (or child, young person, teenager, or youth). The complete search strings are documented in Multimedia Appendix 1. In addition, we manually searched reference lists of included studies and relevant previous reviews, and searched international trial registers for eligible studies, which resulted in one additional record.

Study Selection

We included RCTs in which computer-based, internet-based, or smartphone-based cognitive behavioral therapy targeting anxiety, depression, or both was compared to an active treatment control condition or passive control condition. The study population involved adolescents or young adults with a mean age between 12 and 25 years and elevated symptoms of anxiety or depressive disorder (ie, either a formal diagnosis or an elevated score on a standardized self-report measure representing at least a mild-to-moderate symptom level). We only included studies with an English abstract available, those that were published in peer-reviewed journals or were PhD theses, and those that contained outcome data on a continuous anxiety or depressive symptom measure that allowed for calculation of effect sizes. If effect sizes could not be calculated, authors were contacted to retrieve the necessary information.

The intervention needed to be primarily delivered via technology (ie, computer, internet, or smartphone). Interventions were categorized as CBT if (1) they were explicitly described as such by the authors of the article and we found no reason to disagree or (2) all authors of this review agreed that the description of the main intervention components could be regarded as CBT. The control condition was defined as active treatment control (ie, face-to-face CBT or treatment as usual [TAU]) or passive control (ie, waiting list/no treatment or information control).

Studies in which the control condition involved an active self-help website (ie, including both psychoeducation and exercises) focused specifically on anxiety or depression were excluded. Studies in which the control condition involved a monitoring control website that did not include active self-help content were included. Comorbid psychiatric or medical disorders were not used as an exclusion criterion.

Two authors (CC and MS) conducted the study selection in a stepwise manner. First, titles and abstracts of all studies were independently screened for potential eligibility. Any disagreements were discussed until consensus was reached. Second, the full papers of all included abstracts were independently screened according to the inclusion and exclusion criteria. In case of discrepancy or uncertainty regarding inclusion, a third author (MB) was consulted until consensus was reached.

Data Extraction

Information on study characteristics, participant characteristics, and mental health outcomes was extracted from each study and included in an Excel spreadsheet. Data extraction was conducted by one reviewer (CC) and checked by a second reviewer (MS). Study characteristics included authors, country, year of publication, study design, recruitment setting (ie, clinical, general population, or schools), inclusion and exclusion criteria, primary outcome measures, and descriptions of the experimental intervention and comparator, including focus of the intervention, information on guidance, number of treatment modules, and adherence rates. Participant characteristics included sample size, mean age, gender, primary diagnostic type (ie, anxiety, depression, or both; either based on diagnosis or an elevated symptom level), and baseline symptom levels. Means and SDs of the outcome measures of anxiety and depressive symptoms at posttreatment assessment were extracted. If available, means and SDs at short-term follow-up (ie, 1-5 months) and long-term follow-up (ie, 6-12 months) were extracted as well.

If possible, we utilized effect sizes of the intention-to-treat sample; if these were not available, we used effect sizes of the completer sample. In case of multiple outcome measures of anxiety and depressive symptoms, we selected the primary outcome measure as stated by the authors. If the authors did not specify any primary outcome measure of anxiety or depressive symptoms, we selected a well-validated and widely-used outcome measure of these symptoms that was used at every time point of the study (ie, both at posttest and follow-up, if applicable). If both an active treatment control and passive control were utilized in a single RCT, outcomes from both conditions were extracted. Our main meta-analyses were conducted separately for active treatment control (ie, face-to-face CBT or face-to-face TAU) and passive control (ie, waiting list, information control, or no treatment). In our subgroup analyses, data from all control groups per study were included. As the inclusion of multiple comparisons of one study in a meta-analysis violates the assumption of independence, we divided the *n* of the intervention group evenly across comparators, which is a procedure recommended by the Cochrane guidelines [48].

Quality Assessment

The quality of each included study was assessed following the guidelines provided by the Cochrane risk-of-bias tool for randomized trials, version 2.0 (RoB 2) [49]. Risk of bias was examined in the following five domains: (1) bias arising from the randomization process; (2) bias due to deviations from intended interventions; (3) missing outcome data; (4) bias in measurement of the outcome; and (5) bias in selection of the reported result. Each domain was rated as either low risk of bias, some concerns, or high risk of bias. A total score was calculated for each study by adding up the following values for each domain: “0” for low risk of bias, “1” for some concerns, and “2” for high risk of bias.

The overall quality of the evidence for each outcome was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [50]. The quality of each outcome was assessed for the following domains: (1) risk of bias; (2) inconsistency of results (ie, heterogeneity); (3) indirectness of evidence; (4) imprecision of results; and (5) suspected publication bias. In case of limitations in one domain, the evidence for each outcome was downgraded by one or two levels. Subsequently, the overall evidence for each outcome across domains was categorized as high, moderate, low, or very low, representing the level of certainty of the effect estimates. Both RoB 2 and GRADE assessments were conducted independently by two reviewers (CC and MS), and any disagreements were resolved by discussion until consensus was reached. Cohen κ coefficients were calculated to determine the interrater reliability.

Statistical Analysis

A random-effects meta-analysis was conducted with the Comprehensive Meta-Analysis software (CMA version 3), using the SMD to calculate pooled mean effect sizes (Hedges *g*). Effect sizes were calculated by subtracting the mean posttest score of the treatment group from the mean score of the comparison group, and dividing the result by the pooled standard deviation of the two groups. Posttreatment means and SDs were compared between the intervention and control groups. Effect sizes of 0.2, 0.5, and 0.8 are considered to be small, medium, and large, respectively [51]. In addition, we calculated the number needed to treat (NNT), using the Kraemer & Kupfer [52] formula. The NNT indicates the total number of patients who need to be treated in order to achieve one additional positive outcome [53].

Heterogeneity was assessed by calculating the I^2 statistic, which indicates how much overall variance should be attributed to between-study variance, with a value of 25% representing low heterogeneity, 50% representing moderate heterogeneity, and 75% representing high heterogeneity. In addition, we calculated the 95% CIs around I^2 by using the noncentral chi-square approach in the “heterogi” module of STATA [54,55].

Subgroup analyses were conducted to examine the influence on the difference between intervention and control conditions for (1) the diagnostic focus of the intervention (ie, anxiety, depression, or both); (2) age group (ie, adolescents with mean age ≤18 years or young adults with mean age >18 years); (3)

the level of guidance (ie, guided or self-guided); (4) the adherence rate, defined as the percentage of participants in the intervention group who completed all treatment modules at posttreatment (ie, low, defined as ≤50%, or high, defined as >50%); (5) recruitment type (ie, clinical, community, or university/school); and (6) the number of treatment modules (ie, <5, 5-9, or 10-14 modules). Subgroup analyses were conducted across studies with interventions focused on anxiety, depression, or both, with multiple control groups per study included. Subgroup analyses were conducted using the mixed-effects model, in which the effect sizes within the subgroups are pooled with the random-effects model, whereas the fixed-effects model is used to test for significant differences between the subgroups. Subgroup analyses involving age, guidance level, and adherence rates were planned a priori based on the literature. In addition, a subgroup analysis on the diagnostic focus of the intervention was planned a priori to test whether it was justified to conduct all subgroup analyses on the total set of studies, including interventions aimed at anxiety, depression, or both. Recruitment type and number of treatment modules were examined post-hoc as these reflect potential sources of heterogeneity. For all six subgroup analyses, a Bonferroni-corrected α level of $P<.008$ was used to account for multiple testing. In addition, bivariate meta-regression analyses were conducted to explore the associations of the mean age of study participants, adherence rate, and risk of bias with effect sizes. Analyses with age and adherence rate were planned a priori, whereas risk of bias level was included post-hoc.

Publication bias was examined as follows. First, the funnel plot of effect sizes was visually inspected. Second, the Duval and Tweedie trim and fill procedure was used to calculate an adjusted pooled effect size that accounts for missing studies due to publication bias [56]. Third, the Egger test was used to quantify the bias captured by the funnel plot [57]. In accordance with the Cochrane guidelines [48], publication bias was only examined in meta-analyses with at least 10 studies.

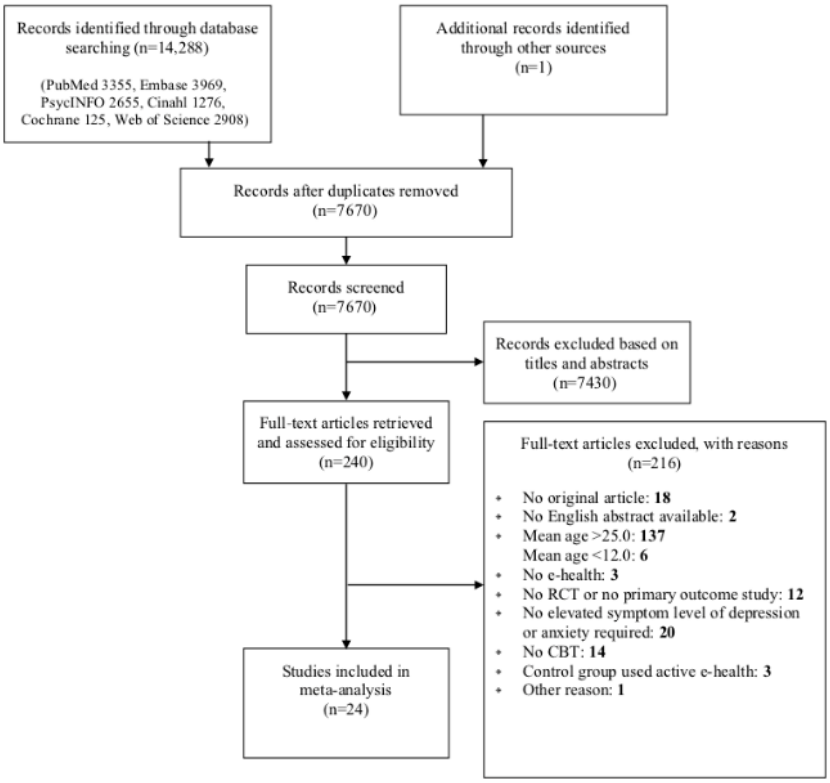
Results

Systematic Review

Included Studies

The database search resulted in 7670 articles, of which we retrieved the full text of 240 articles. Twenty-four studies met all inclusion criteria and were included in the systematic review and meta-analysis (Figure 1). Interrater agreement of inclusion was strong (98.3%; Cohen κ=0.90, $P<.001$). The included studies were published between 2009 and September 2019. Most were conducted in Australia (n=5) and the United Kingdom (n=4). All studies reported posttreatment effects, whereas short-term and long-term follow-up data of both intervention and control conditions were only reported in three and five studies, respectively. The sample sizes of the RCTs ranged from 19 to 257 (mean 92.75, median 70). Twelve studies were primarily aimed at adolescents (age 12-19 years), eight studies were aimed at young adults (age 19-25 years), and four studies had a mixed sample. The mean age varied between 13.31 [58] and 24.40 years [59]. Most studies were conducted in samples of university students (n=8) or community samples (n=7), whereas four studies were conducted in clinical samples, four studies in secondary schools or educational programs, and one study in a mixed sample [60]. Studies targeted participants with a diagnosis or elevated symptoms of depressive disorder (n=10), participants with a diagnosis or elevated symptoms of anxiety disorder (n=8), or participants with elevated symptoms of depressive and/or anxiety disorder (n=6). In total, 19 studies compared cCBT to a waiting list or no treatment control condition, of which five studies also included a face-to-face CBT control condition. Four studies compared cCBT to a placebo condition (information control or attention control), and one study compared cCBT to TAU. Selected characteristics of the included studies are presented in Multimedia Appendix 2.

Figure 1. Flow chart.



Description of the Interventions

Seventeen studies investigated cCBT programs delivered via the internet (also known as iCBT), such as MoodGym [60-62] and BRAVE [63,64]. Of these interventions, most were completed at the respondent's home (n=12), whereas five were completed at a treatment or research site. Seven studies investigated cCBT programs delivered via a computer program or CD-ROM, such as SPARX [65-67], Woebot [68], and Stressbusters [58]. Of these interventions, four were completed at the respondent's home (n=4) and three were completed at school or a treatment site. The regular treatment components of cCBT were psychoeducation, behavioral activation, cognitive restructuring, exposure, problem-solving, and homework assignments.

Treatment Duration and Intensity

The number of treatment modules ranged from 3 to 12 in 21 included studies (mean 7.1, median 7). Two studies did not report the exact number of modules [59,69]. One study examined the Woebot intervention [68], which does not consist of different modules, but delivers cCBT by 1 to 20 (median 12) automated conversations and mood tracking in an instant messenger app. The 24 included interventions were completed over a period of 2 to 16 weeks (mean 7.5, median 7).

Guidance

In 14 studies, participants were guided through the intervention by a therapist or researcher. The other 10 interventions were self-guided (ie, unguided) [58,59,65-72]. Guidance was provided through telephone and/or email contact [63,64,73-78], chat sessions [79], or face-to-face guidance during the participant's completion of the modules [60-62,80,81]. In general, guidance consisted of monitoring progress and providing support, encouragement, and clarification. In nine studies, guidance additionally included providing personalized feedback on completed exercises [63,64,74-80].

Adherence

Adherence rates were reported in 19 studies. Only 10 studies reported the most common measure of adherence (ie, the number of completed sessions divided by the maximum number of sessions). In these studies, adherence ranged from 32.2% to 100% (mean 76.91%, median 78%). An alternative measure of adherence, namely the percentage of participants in the intervention group who completed all treatment modules, was reported in 19 studies, with adherence rates ranging from 0% to 100% (mean 57.12%, median 60%). Adherence rates for each study are presented in Multimedia Appendix 3.

Quality of Evidence

Based on the Cochrane RoB 2 tool [49], 22 out of 24 studies had an overall high risk of bias, and the remaining two studies were rated as "some concerns." The overall high risk of bias in this vast majority of included studies was mainly due to an increased risk of bias in the measurement of the outcome, caused by the use of self-report measures or the unblinded use of observer-rated measures. Furthermore, 13 out of 24 studies also had an increased risk of bias due to missing outcome data, since >5% of their data were missing and no sufficiently appropriate analysis (eg, multiple imputation) was used to handle the missing data. Lastly, one study was rated as having an increased risk of bias due to deviations from the intended intervention [77]. Interrater reliability for the risk of bias was very good (κ=0.89, P<.001). Figure 2 demonstrates the authors'

conclusions regarding the risk of bias across studies. Multimedia Appendix 4 presents the risk of bias classifications per domain assigned to each study. It should be noted that self-report measures are widely used in psychological treatment studies, especially in studies on computerized treatment. Therefore, the current rating may be too strict. Without taking into account the risk of bias in the measurement of the outcome, 13 out of 24 studies had an overall high risk of bias.

The overall quality of the evidence for each outcome was assessed using the GRADE approach [50]. The quality rate for each outcome is shown in Tables 1-3. In summary, although the overall quality of some outcomes was moderate, the overall quality of most outcomes was low. Since almost all studies were associated with a high risk of bias based on the RoB tool, all outcomes were downgraded one level for this domain. Many outcomes were downgraded one additional level for inconsistency, because of substantial heterogeneity in the meta-analysis. One outcome was downgraded one additional level for imprecision of results due to a small sample size. Interrater reliability for the quality of evidence was very good (κ=0.87, P<.001).

Figure 2. Risk of bias graph.

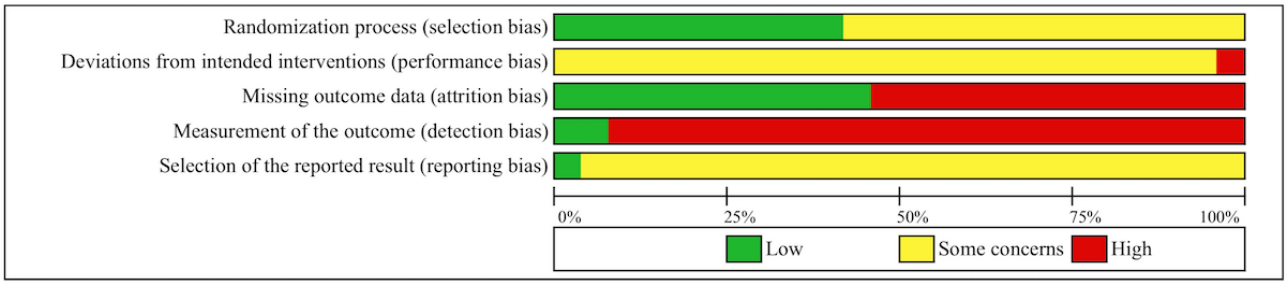


Table 1. Effect sizes regarding depressive symptoms in the meta-analysis of studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at posttreatment.

			Effect size		Heterogeneity			
Variable	N_{com}[a]	N_{par}[b]	*g*	95% CI	I^2	95% CI	Grade[c]	NNT[d]
All studies								
Active treatment controls	5	403	−0.55	−1.18 to 0.08	87.52	73 to 94	++	3.31
Passive controls	20	1604	0.52[e]	0.33 to 0.71	68.69	50 to 80	++	3.5
One outlier removed[f]	19	1558	0.46[e]	0.29 to 0.63	58.49	31 to 75	+++	3.91
Studies aimed at depression								
Active treatment controls	4	351	−0.70	−1.51 to 0.11	90.63	79 to 96	+	2.63
Passive controls	13	1162	0.60[e]	0.35 to 0.85	73.27	54 to 85	++	3.05
One outlier removed[f]	12	1116	0.51[e]	0.30 to 0.72	61.76	28 to 80	++	3.55

[a]N_{com}: number of comparisons.

[b]N_{par}: number of participants.

[c]+: very low quality; ++: low quality; +++: moderate quality.

[d]NNT: number needed to treat.

[e]P<.001.

[f]Outlier Sethi (2013) excluded.

Table 2. Effect sizes regarding depressive symptoms in the subgroup analyses of studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at posttreatment.

			Effect size		Heterogeneity			
Variable	N_{com}[a]	N_{par}[b]	*g*	95% CI	I^2	95% CI	*P*	NNT[c]
Diagnostic focus[d]							.63	
Anxiety	8	464	0.32[e]	0.04 to 0.60	50.78	0 to 78		5.56
Depression	9	1152	0.44[f]	0.21 to 0.67	75.44	53 to 87		4.1
Both	4	182	0.55[e]	0.13 to 0.96	0	0 to 85		3.31
Age group[d]							.29	
Adolescents	12	1027	0.34[g]	0.14 to 0.54	61.51	28 to 79		5.26
Young adults	9	771	0.50[f]	0.28 to 0.73	39.7	0 to 72		3.62
Guidance[d]							.56	
Guided	10	771	0.46[f]	0.23 to 0.70	56.96	13 to 79		3.91
Self-guided	11	1027	0.37[g]	0.16 to 0.58	58.97	20 to 79		4.85
Adherence[d]							.77	
Low	6	779	0.39[e]	0.09 to 0.69	77.56	50 to 90		4.59
High	10	737	0.48[f]	0.23 to 0.72	62.06	25 to 81		3.76
Recruitment type[d]							.53	
Clinical	4	376	0.24	−0.11 to 0.59	0	0 to 85		7.46
Community	5	462	0.50[g]	0.18 to 0.82	78.12	47 to 91		3.62
University/school	12	960	0.44[f]	0.23 to 0.65	50.27	3 to 74		4.1
Number of modules[d]							.21	
<5	3	138	0.65[g]	0.20 to 1.11	0	0 to 90		2.82
5-9	12	1097	0.49[f]	0.29 to 0.69	69.11	44 to 83		3.68
10-14	3	377	0.14	−0.23 to 0.52	0	0 to 90		12.82

[a]N_{com}: number of comparisons.

[b]N_{par}: number of participants.

[c]NNT: number needed to treat.

[d]Outliers Sethi (2010) and Sethi (2013) excluded.

[e]$P<.05$.

[f]$P<.001$.

[g]$P<.01$.

Table 3. Bivariate meta-regression analyses regarding depressive symptoms in studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at posttreatment.

Variable	N_{com}[a]	*b*	95% CI	*P*
Mean age[b]				
Intercept	21	0.14	−0.55 to 0.84	.69
Mean age	21	0.01	−0.02 to 0.05	.44
Adherence[b]				
Intercept	16	0.39	−0.02 to 0.79	.06
Adherence	16	0.01	−0.01 to 0.01	.75
Risk of bias[b]				
Intercept	21	−0.07	−0.68 to 0.53	.81
Risk of bias	21	0.11	−0.01 to 0.23	.08

[a]N_{com}: number of comparisons.

[b]Outliers Sethi (2010) and Sethi (2013) excluded.

Meta-Analysis

Effects of cCBT on Depressive Symptoms at Posttreatment

The pooled effect size of cCBT for depressive disorders, anxiety disorders, or both regarding depressive symptoms at posttreatment compared with active treatment controls was g=−0.55 (95% CI −1.18 to 0.08, P=.09, k=5; Table 1), and heterogeneity was high (I^2=87.52%, 95% CI 73-94). Compared with passive controls, the pooled effect size of cCBT was g=0.52 (95% CI 0.33-0.71, P<.001, k=20), and heterogeneity was moderate (I^2=68.69%, 95% CI 50-80). Removing one extreme outlier with an effect size of g=1.93 [60] resulted in a somewhat smaller mean effect size of g=0.46 (95% CI 0.29-0.63, P<.001, k=19), with a lower, though still moderate, heterogeneity (I^2=58.49%, 95% CI 31-75).

In studies aimed specifically at depressive disorders or both depressive and anxiety disorders, the nonsignificant pooled effect size of cCBT compared with active treatment controls regarding depressive symptoms was g=−0.70 (95% CI −1.51 to 0.11, P=.09, k=4), corresponding to an NNT of 2.63 in favor of active treatment controls. Heterogeneity was very high (I^2=90.63%, 95% CI 79-96). When compared with passive controls, cCBT yielded a significant medium effect size of g=0.60 (95% CI 0.35-0.85, P<.001, k=13), corresponding to an NNT of 3.05. Heterogeneity was moderate to high (I^2=73.27, 95% CI 54-85). Removing the extreme outlier [60] again resulted in a somewhat smaller mean effect size of g=0.51 (95% CI 0.30-0.72, P<.001, k=12), with a lower, though still moderate, heterogeneity (I^2=61.76, 95% CI 28-80) and a corresponding NNT of 3.55. Inspection of the funnel plot and the Duval and Tweedie trim and fill procedure showed no indication of publication bias, and the Egger test of the intercept was not significant (P=.74), indicating no need to adjust for missing studies. Multimedia Appendix 5 and Multimedia Appendix 6 provide forest plots of effect sizes regarding depressive symptoms for active treatment controls and passive controls, respectively.

A series of subgroup analyses (Table 2) was conducted across studies focused on depression, anxiety, or both, with multiple control groups per study included and two extreme outliers excluded [60,62]. Heterogeneity remained moderate in most subgroups. Effects in all but two subgroups were significantly different from zero, and all were in favor of cCBT. We found no indication that the diagnostic focus of the intervention, age group, level of guidance, adherence rate, type of recruitment, or number of treatment modules was associated with differential effect sizes. Lastly, bivariate meta-regression analyses (Table 3) showed no significant association of the mean age of study participants (b=0.01; 95% CI −0.02 to 0.05, P=.44), adherence (b=0.01; 95% CI −0.01 to 0.01, P=.74), or risk of bias (b=0.11; 95% CI −0.01 to 0.23, P=.08) with effect size regarding depressive symptoms.

Effects of cCBT on Anxiety Symptoms at Posttreatment

Regarding anxiety symptoms at posttreatment, the pooled effect size of cCBT for anxiety disorders, depressive disorders, or both compared with active treatment controls was g=0.06 (95% CI −0.13 to 0.26, P=.53, k=5; Table 4). Heterogeneity was low, although the wide 95% CI indicated some uncertainty regarding the exact level of heterogeneity (I^2=0.00%, 95% CI 0-79). Compared with passive controls, cCBT yielded a significant pooled effect size of g=0.49 (95% CI 0.29-0.68, P<.001, k=21), and heterogeneity was moderate (I^2=68.17%, 95% CI 50-80). Removing one extreme outlier with an effect size of g=1.94 [60] resulted in a slightly smaller mean effect size of g=0.42 (95% CI 0.25-0.59, P<.001, k=20), with a lower, though still moderate, heterogeneity (I^2=57.42%, 95% CI 30-74).

When only including studies with cCBT aimed specifically at anxiety disorders or at both anxiety and depressive disorders, the nonsignificant pooled effect size regarding anxiety symptoms compared with active treatment controls remained similar (g=0.04, 95% CI −0.23 to 0.31, P=.79, k=4). Heterogeneity was low, although the wide 95% CI again indicated uncertainty

regarding the exact level of heterogeneity (I^2=0.00%, 95% CI 0-85). Compared with passive controls, cCBT yielded a pooled effect size of g=0.59 (95% CI 0.34-0.84, P<.001, k=16), corresponding to an NNT of 3.09, and heterogeneity was moderate (I^2=67.83%, 95% CI 46-81). Again, removing the extreme outlier [60] resulted in a smaller mean effect size of g=0.50 (95% CI 0.29-0.71, P<.001, k=15), with an NNT of 3.62 and a lower, though still moderate, heterogeneity (I^2=52.57%, 95% CI 15-74). Inspection of the funnel plot and the Duval and Tweedie trim and fill procedure showed a minor indication of publication bias, but the Egger test of the intercept was not significant (P=.11). Adjusting for missing studies using the Duval and Tweedie trim and fill procedure resulted in a slightly smaller overall effect size of g=0.44 (95% CI 0.23-0.65), corresponding to an NNT of 4.10. Multimedia Appendix 7 and Multimedia Appendix 8 provide forest plots of effect sizes regarding anxiety symptoms for active treatment and passive control conditions, respectively.

A series of subgroup analyses were conducted across studies focused on anxiety, depression, or both, with multiple comparators per study included and one outlier excluded [60]. Heterogeneity was moderate in most subgroups (Table 5). Effects in most subgroups were different from zero, and all were in favor of cCBT. We found no indication that the diagnostic focus of the intervention, age group, type of guidance, adherence rate, type of recruitment, or number of treatment modules was associated with differential effect sizes. Lastly, bivariate meta-regression analyses (Table 6) showed no significant association of the mean age of study participants (b=0.02; 95% CI −0.01 to 0.06, P=.21), adherence (b=0.00; 95% CI −0.01 to 0.00, P=.65), or risk of bias (b=0.04; 95% CI −0.05 to 0.13, P=.36) with effect size regarding anxiety symptoms.

Table 4. Effect sizes regarding anxiety symptoms in the meta-analysis of studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at posttreatment.

			Effect size		Heterogeneity		Grade[c]	NNT[d]
Variable	N_{com}[a]	N_{par}[b]	g	95% CI	I^2	95% CI		
All studies								
Active treatment controls	5	390	0.06	−0.13 to 0.26	0	0 to 79	++	29.41
Passive controls	21	1570	0.49[e]	0.29 to 0.68	68.17	50 to 80	++	3.68
One outlier removed[f]	20	1524	0.42[e]	0.25 to 0.59	57.42	30 to 74	+++	4.27
Studies aimed at anxiety								
Active treatment controls	4	203	0.04	−0.23 to 0.31	0	0 to 85	++	45.45
Passive controls	16	868	0.59[e]	0.34 to 0.84	67.83	46 to 81	++	3.09
One outlier removed[f]	15	822	0.50[e]	0.29 to 0.71	52.57	15 to 74	+++	3.62
Trim and fill adjusted values			0.44	0.23 to 0.65				4.1

[a]N_{com}: number of comparisons.

[b]N_{par}: number of participants.

[c]++: low quality; +++: moderate quality.

[d]NNT: number needed to treat.

[e]P<.001.

[f]Outlier Sethi (2013) excluded.

Table 5. Effect sizes regarding anxiety symptoms in the subgroup analyses of studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at posttreatment.

			Effect size		Heterogeneity			
Variable	N_{com}[a]	N_{par}[b]	*g*	95% CI	I^2	95% CI	*P*	NNT[c]
Diagnostic focus[d]								
Anxiety	12	687	0.47[e]	0.25 to 0.69	34.28	0 to 67	.39	3.85
Depression	6	889	0.23	−0.03 to 0.50	71.68	34 to 88		7.69
Both	6	211	0.33	−0.04 to 0.69	61.89	7 to 84		5.43
Age group[d]								
Adolescents	13	1031	0.25[f]	0.06 to 0.44	44.27	0 to 71	.08	7.14
Young adults	11	756	0.51[e]	0.29 to 0.73	51.47	3 to 76		3.55
Guidance[d]								
Guided	15	958	0.41[e]	0.21 to 0.61	48.01	5 to 71	.47	4.39
Self-guided	9	829	0.30[g]	0.07 to 0.53	60.28	17 to 81		5.95
Adherence[d]								
Low	8	894	0.44[e]	0.19 to 0.68	38.77	0 to 73	.61	4.1
High	11	655	0.27[g]	0.04 to 0.51	55.6	13 to 77		6.58
Recruitment type[d]								
Clinical	3	267	0.06	−0.34 to 0.46	0	0 to 90	.28	29.41
Community	8	620	0.43[f]	0.17 to 0.68	58.97	10 to 81		4.2
University/school	13	900	0.40[e]	0.20 to 0.61	51.22	7 to 74		4.5
Number of modules[d]								
<5	5	167	0.47[g]	0.05 to 0.89	68.76	20 to 88	.91	3.85
5-9	12	1051	0.38[f]	0.16 to 0.60	63.37	32 to 80		4.72
10-14	5	492	0.29	−0.04 to 0.61	12.58	0 to 82		6.17

[a]N_{com}: number of comparisons.

[b]N_{par}: number of participants.

[c]NNT: number needed to treat.

[d]Outlier Sethi (2013) excluded.

[e]*P*<.001.

[f]*P*<.01.

[g]*P*<.05.

Table 6. Bivariate meta-regression analyses regarding anxiety symptoms in studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at posttreatment.

Variable	N_{com}[a]	*b*	95% CI	*P*
Mean age[b]				
Intercept	24	−0.07	−0.77 to 0.63	.84
Mean age	24	0.02	−0.01 to 0.06	.21
Adherence[b]				
Intercept	19	0.42	0.10 to 0.74	.01
Adherence	19	0.00	−0.01 to 0.00	.65
Risk of Bias[b]				
Intercept	24	0.13	−0.38 to 0.65	.61
Risk of Bias	24	0.04	−0.05 to 0.13	.36

[a]N_{com}: number of comparisons.

[b]Outlier Sethi (2013) excluded.

Short-Term Follow-Up Effects

Three studies reported short-term follow-up effects (ie, up to 5 months posttreatment) for cCBT on depressive symptoms. The pooled effect size for studies with cCBT aimed specifically at depressive disorders or at both depressive and anxiety disorders compared with active treatment controls was not significant (*g*=0.12, 95% CI −0.11 to 0.35, *P*=.29, *k*=2; Table 7). Compared with passive controls, the pooled effect size showed no significant difference between cCBT and control conditions either (*g*=0.19, 95% CI −0.08 to 0.46, *P*=.16, *k*=2). Although effect sizes were in favor of cCBT, these results indicated that cCBT is not superior to controls at short-term follow-up. However, owing to the small number of comparisons, the statistical power to detect small differences was limited. Heterogeneity was low (I^2=0.00%), but the number of studies was too small to enable calculation of 95% CI. No studies reported short-term follow-up effects for cCBT on anxiety symptoms.

Table 7. Effect sizes regarding depressive and anxiety symptoms in the meta-analysis of studies comparing computerized cognitive behavior therapy in adolescents and young adults with active treatment and passive controls at short-term follow-up (1-5 months) and long-term follow-up (6-12 months).

			Effect size		Heterogeneity			
Variable	N_{com}[a]	N_{par}[b]	*g*	95% CI	I^2	95% CI	Grade[c]	NNT[d]
Depressive symptoms								
Short-term follow-up								
Active treatment controls	2	288	0.12	−0.11 to 0.35	0	N/A[e]	+++	14.71
Passive controls	2	211	0.19	−0.08 to 0.46	0	N/A[e]	++	9.43
Long-term follow-up[f]								
Passive controls	3	461	0.27[g]	0.09 to 0.45	0	0 to 90	+++	6.58
Anxiety symptoms								
Long-term follow-up[h]								
Active treatment controls	2	140	0.08	−0.41 to 0.56	50.61	N/A[e]	++	21.74

[a]N_{com}: number of comparisons.

[b]N_{par}: number of participants.

[c]++: low quality; +++: moderate quality.

[d]NNT: number needed to treat.

[e]N/A: not applicable; calculation of 95% CI not possible because *df*=1.

[f]Only one study with active treatment controls available.

[g]*P*<.01.

[h]No studies with passive controls available.

Long-Term Follow-Up Effects

Three studies reported long-term follow-up effects (ie, 6-12 months posttreatment) for cCBT on depressive symptoms. The pooled effect size indicated cCBT aimed at depressive symptoms or both depressive and anxiety symptoms to be effective compared with passive controls at long-term follow-up (g=0.27, 95% CI 0.09-0.45, P=.004, k=3), corresponding with an NNT of 6.58. Heterogeneity was low, although the wide 95% CI indicated uncertainty regarding the exact level of heterogeneity (I^2=0.00%, 95% CI 0-90). As only one study [65] reported long-term follow-up effects for cCBT on depressive symptoms compared with active treatment controls, meta-analysis was not possible.

Only two studies reported long-term follow-up effects (ie, 6-12 months posttreatment) for cCBT aimed at anxiety or both anxiety and depression on anxiety symptoms. The pooled effect size showed no significant effect for cCBT compared with active treatment controls (g=0.08, 95% CI −0.41 to 0.56, P=.75, k=2) at long-term follow-up. No study reported long-term follow-up effects for cCBT on anxiety symptoms compared with passive controls.

Discussion

Principal Findings

This study provides an up-to-date meta-analysis examining the effects of cCBT on anxiety and depressive symptoms in adolescents and young adults compared with active treatment and passive controls, differentiating between posttreatment and follow-up. Our results indicate that cCBT is beneficial for reducing anxiety and depressive symptoms at posttreatment in adolescents and young adults compared with passive controls, with small to medium effect sizes. For cCBT aimed at depressive disorders or depressive and anxiety disorders, we found a pooled effect size of g=0.51 regarding depressive symptoms, which corresponds to an NNT of 3.55. For cCBT aimed at anxiety disorders or anxiety and depressive disorders, we found an effect size of g=0.50 regarding anxiety symptoms. After adjustment for missing studies owing to a minor indication of publication bias, the effect size lowered slightly to g=0.44, corresponding to an NNT of 4.10. Compared with active treatment controls, the pooled effect size regarding depressive symptoms was in favor of controls (g=−0.70). However, the effect size was not significant and heterogeneity was very high. For anxiety symptoms, cCBT and active treatment controls showed similar effects (g=0.04). Subgroup analyses did not reveal any differences between groups; however, owing to the small number of studies, the statistical power to detect small differences was limited. Meta-regression analyses showed no associations between age, adherence rate, or risk of bias and effect sizes.

Overall, this study shows robust evidence of the effectiveness of cCBT in reducing anxiety and depressive symptoms in adolescents and young adults compared with passive controls. Our results are largely in line with those of previous studies on cCBT in children and young people aged up to 25 years [36], adolescents and young adults aged 12 to 25 years [37] and 10 to 25 years [43], and children and adolescents aged up to 18 years [44]. However, these studies generally reported somewhat larger effect sizes (range 0.62-0.77) regarding both depressive symptoms and anxiety symptoms [36], anxiety symptoms [37], or depressive and/or anxiety symptoms [44] compared with passive controls [36,37,44]. Similarly, research in adults found larger effect sizes for cCBT regarding depression and anxiety compared with passive controls (g=0.90) [34]. Compared with the effects of traditional face-to-face CBT in children and adolescents with anxiety disorders against waiting list controls (NNT=3.0) [19], we found a somewhat lower NNT for cCBT against passive controls (NNT=4.10) after adjusting for potential publication bias. Effect sizes in our study were similar to those found in a meta-analysis comparing face-to-face CBT for depression (0.60) and anxiety disorders (0.48) to passive controls in college and university students [21].

Our results suggest that the effects of cCBT do not differ from those of active treatment controls (ie, face-to-face CBT or face-to-face TAU) regarding anxiety symptoms, but may be inferior to active treatment controls regarding depressive symptoms, although the effect size was not significant. These findings are in line with those of previous meta-analyses in youth across three early studies that were also included in the current meta-analysis [37,41]. Although our meta-analysis included three additional studies, the number of RCTs comparing cCBT with face-to-face treatment in adolescents and young adults remains small. With regard to depressive symptoms, heterogeneity was very high (I^2=90.63), and the pooled effect size should be interpreted with caution. Hence, more research directly comparing both treatments is needed to determine whether cCBT is effective compared with face-to-face treatment controls in adolescents and young adults. Research in adults has shown largely equivalent effects of cCBT on both anxiety and depressive symptoms compared with face-to-face CBT [34,35]. However, the number of studies directly comparing cCBT with face-to-face CBT in adults remains limited as well.

This study also aimed to investigate the effectiveness of cCBT at short-term and long-term follow-ups. However, the number of studies reporting follow-up effects was limited. Regarding long-term effects, cCBT was effective in reducing depressive symptoms compared with passive controls, with a small effect size (g=0.27). Our results indicated no long-term follow-up effect for cCBT on anxiety symptoms compared with active treatment controls. Meta-analyses of short-term follow-up data on depressive symptoms indicated similar effects for cCBT compared with active treatment and passive controls. Only Välimäki et al [43] investigated the follow-up effects of cCBT and other internet-based interventions in adolescents and young adults, reporting mid-term and long-term effects on depressive symptoms and mid-term effects on anxiety symptoms. However, they only reported mean differences, and no standardized effect size or NNT. In addition, they did not separately examine the effects of cCBT. Moreover, their selection of studies was based on interventions aimed at depression and not anxiety. Therefore, their results are not easily comparable to those of the current study.

Importantly, owing to the small number of studies reporting follow-up effects, the power to detect small effect sizes was limited in both this study and the study of Välimäki et al [43]. In contrast, the number of studies reporting follow-up effects of cCBT in adults is substantially larger. A recent meta-analysis found 29 trials that reported short-term follow-up effects and 15 trials that reported long-term follow-up effects [34]. cCBT for depressive disorder or anxiety disorder was found to be effective at short-term follow-up (ie, 3-6 months) and long-term follow-up (ie, 9-18 months) compared with posttreatment effect sizes, with small effect sizes across disorders (ie, g=0.15 and g=0.22, respectively). In contrast with the small effect sizes identified in our study, Andersson et al [82] found very large effect sizes for cCBT regarding depressive or anxiety symptoms in adults (g=1.31 across 10 studies) at long-term follow-up of 2 to 5 years compared with mainly passive controls. However, the authors noted that it was unclear whether randomization remained intact over the follow-up period. In summary, in order to determine the long-term effects of cCBT in adolescents and young adults, it is of great importance that future studies include follow-up assessments. Studies comparing cCBT with active treatment control conditions should aim to maintain randomization during the entire follow-up period.

Furthermore, this study aimed to explore whether respondents' age, guidance level, and treatment adherence were associated with effect sizes. No moderators of treatment effects could be identified. We found no differences in effect sizes for adolescents and young adults regarding anxiety or depressive symptoms, and no association between respondents' mean age and effect sizes. Previous studies that examined the moderating role of age in meta-analyses among youth reported mixed results. Pennant et al [37] found a higher effect size in young adults compared with adolescents regarding anxiety symptoms, but not depressive symptoms. However, the authors noted that these groups also differed in terms of symptom level, which may have caused the difference in effect sizes. Ebert et al [36] and Podina et al [42] found a higher effect size in adolescents compared with children [36], whereas others [38,44] did not find evidence for such a moderating role of age. Regarding the absence of an association between guidance level and effect sizes, our results correspond with those of Pennant et al [37] in the same age groups. Studies in children and adolescents found mixed results, with Podina et al [42] reporting higher effect sizes for lower levels of guidance and Grist et al [44] reporting higher effect sizes for higher levels of guidance. The lack of an association between adherence rates and effect sizes in our study contrasts findings in adults with depression and anxiety [33,83]. However, most studies included in this meta-analysis did not report the most common operationalization of treatment adherence, and several did not report any information on treatment completion. Post-hoc subgroup analyses and meta-regression analyses found no association of the diagnostic focus of the intervention, risk of bias, recruitment type, or number of sessions with effect sizes.

This study included a thorough evaluation of the risk of bias and quality of evidence, which indicated an overall high risk of bias in 22 out of 24 studies, and, accordingly, low to moderate overall quality of evidence. The high risk of bias was mainly due to an increased risk of bias in measurement of the outcome caused by self-report or unblinded use of observer-rated measures. However, as self-report measures allow both treatment and outcome measures of studies on computerized interventions to be completed entirely from the participant's home, they are commonly used in studies on cCBT. As such, using the Cochrane RoB tool 2.0, a high risk of bias in measurement of the outcome is inevitable in many studies on computerized interventions for depression and anxiety. Nevertheless, studies ideally should aim to complement self-report measures with blinded observer-rated outcomes, although measures of anxiety and depressive symptoms will always remain subjective to some extent, even when observer-rated outcomes are used.

Limitations

Several limitations should be considered when interpreting the results. First, the number of studies in the meta-analyses was limited, especially with regard to short-term and long-term follow-up effects. Therefore, the power to detect small effect sizes was limited. Likewise, subgroup analyses consisted of a small number of comparisons, and the lack of relevant differences in most subgroup analyses might be caused by low statistical power. Second, the included studies showed large variations in intervention content, treatment intensity, and outcome measures. Heterogeneity was considerable in the majority of analyses, and pooled effect sizes should be interpreted with caution. Third, most studies had a high risk of bias owing to the use of self-report measures and/or inappropriate handling of missing data. Overall quality was low for most comparisons because of the high risk of bias and, in most cases, considerable heterogeneity. Lastly, almost all studies were conducted in high-income countries, and most studies in young adults were conducted among university students. Hence, generalizability of these results to other populations may be limited.

Future Directions

In the rapidly growing field of computerized mental health treatment in adolescents and young adults, new interventions are developed at a fast pace. Since the publication of the most recent previous meta-analysis in adolescents and young adults [43], six new studies were published, which have been included in our meta-analysis. However, the evidence base in young people remains limited compared with the large body of research in adults, and the quality of RCTs is often low. In addition, most RCTs have compared cCBT to passive control conditions, which appears to lead to an overestimation of effects. It is of utmost importance to compare the effects of cCBT with gold standard face-to-face treatment in order to determine whether cCBT can provide an equally effective alternative. Furthermore, more rigorous high-quality research in accordance with the CONSORT and CONSORT eHealth guidelines for conducting and reporting RCTs [84,85] is needed. In particular, future studies should minimize risk of bias by appropriately handling missing data and, ideally, complementing the use of self-report questionnaires with blinded observer-rated measures. Future research should also include larger sample sizes and longer follow-up periods in which randomization is maintained in case of active control groups [82] and should report adherence rates.

Finally, future studies should investigate the effect of cCBT in lower-educated samples, as well as young people from low-income countries, for whom face-to-face mental health treatment is often unavailable [86-88]. When high-quality evidence in adolescents and young adults accumulates, future researchers will be able to draw stronger conclusions on the effectiveness of cCBT compared with both active treatment and passive controls and to determine differences in effect sizes for various subgroups and populations.

Conclusions

This meta-analysis provides robust evidence for the effectiveness of cCBT in the treatment of anxiety and depressive disorders in adolescents and young adults compared with passive controls, with small to medium posttreatment effect sizes. Furthermore, our results indicate that effects of cCBT are similar to those of active treatment controls in reducing anxiety symptoms. Regarding depressive symptoms, however, the results remain unclear, since heterogeneity was high and the number of studies comparing cCBT with active treatment controls was small. No moderators of treatment effects could be identified. cCBT appears to be a promising treatment option for young people, of whom most do not receive face-to-face treatment [23-25]. Importantly, this study also demonstrates the need for more methodologically high-quality research in this population, including active treatment control groups and long-term follow-up assessments.

Acknowledgments

The authors wish to thank Mrs CHM Planting, information specialist, for her advice and assistance regarding the search strategy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings.
[DOCX File , 21 KB - jmir_v22i9e17831_app1.docx]

Multimedia Appendix 2
Selected characteristics of the included studies examining the effects of computerized cognitive behavior therapy for depression and anxiety in adolescents and young adults.
[DOCX File , 19 KB - jmir_v22i9e17831_app2.docx]

Multimedia Appendix 3
Adherence rates of the included studies examining the effects of computerized cognitive behavior therapy for depression and anxiety in adolescents and young adults.
[DOCX File , 20 KB - jmir_v22i9e17831_app3.docx]

Multimedia Appendix 4
Risk of bias classification for each study.
[PNG File , 312 KB - jmir_v22i9e17831_app4.png]

Multimedia Appendix 5
Forest plot with standardized posttreatment effect sizes (Hedges g) regarding depressive symptoms for computerized cognitive behavior therapy compared with active treatment control conditions.
[PNG File , 62 KB - jmir_v22i9e17831_app5.png]

Multimedia Appendix 6
Forest plot with standardized posttreatment effect sizes (Hedges g) regarding depressive symptoms for computerized cognitive behavior therapy compared with passive control conditions.
[PNG File , 99 KB - jmir_v22i9e17831_app6.png]

Multimedia Appendix 7
Forest plot with standardized posttreatment effect sizes (Hedges g) regarding anxiety symptoms for computerized cognitive behavior therapy compared with active treatment control conditions.
[PNG File , 62 KB - jmir_v22i9e17831_app7.png]

Multimedia Appendix 8

Forest plot with standardized posttreatment effect sizes (Hedges g) regarding anxiety symptoms for computerized cognitive behavior therapy compared with passive control conditions.
[PNG File , 112 KB - jmir_v22i9e17831_app8.png]

References

1. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci 2009;11(1):7-20 [FREE Full text] [Medline: 19432384]
2. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. The Lancet 2007 Apr;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7]
3. Johnson D, Dupuis G, Piche J, Clayborne Z, Colman I. Adult mental health outcomes of adolescent depression: A systematic review. Depress Anxiety 2018 Aug;35(8):700-716. [doi: 10.1002/da.22777] [Medline: 29878410]
4. Fergusson DM, Woodward LJ. Mental health, educational, and social role outcomes of adolescents with depression. Arch Gen Psychiatry 2002 Mar;59(3):225-231. [doi: 10.1001/archpsyc.59.3.225] [Medline: 11879160]
5. Fergusson DM, Horwood LJ, Ridder EM, Beautrais AL. Subthreshold depression in adolescence and mental health outcomes in adulthood. Arch Gen Psychiatry 2005 Jan;62(1):66-72. [doi: 10.1001/archpsyc.62.1.66] [Medline: 15630074]
6. McLeod GF, Horwood LJ, Fergusson DM. Adolescent depression, adult mental health and psychosocial outcomes at 30 and 35 years. Psychol Med 2016 May;46(7):1401-1412. [doi: 10.1017/S0033291715002950] [Medline: 26818194]
7. McKenzie M, Jorm AF, Romaniuk H, Olsson CA, Patton GC. Association of adolescent symptoms of depression and anxiety with alcohol use disorders in young adulthood: findings from the Victorian Adolescent Health Cohort Study. Med J Aust 2011 Aug 01;195(3):S27-S30. [Medline: 21806515]
8. Wolitzky-Taylor K, Dour H, Zinbarg R, Mineka S, Vrshek-Schallhorn S, Epstein A, et al. Experiencing core symptoms of anxiety and unipolar mood disorders in late adolescence predicts disorder onset in early adulthood. Depress Anxiety 2014 Mar;31(3):207-213. [doi: 10.1002/da.22250] [Medline: 24577995]
9. Hawton K, Saunders KE, O'Connor RC. Self-harm and suicide in adolescents. Lancet 2012 Jun 23;379(9834):2373-2382. [doi: 10.1016/S0140-6736(12)60322-5] [Medline: 22726518]
10. Patton GC, Coffey C, Sawyer SM, Viner RM, Haller DM, Bose K, et al. Global patterns of mortality in young people: a systematic analysis of population health data. The Lancet 2009 Sep;374(9693):881-892. [doi: 10.1016/s0140-6736(09)60741-8]
11. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 2015 Mar;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]
12. Avenevoli S, Swendsen J, He JP, Burstein M, Merikangas KR. Major depression in the national comorbidity survey-adolescent supplement: prevalence, correlates, and treatment. J Am Acad Child Adolesc Psychiatry 2015 Jan;54(1):37-44.e2 [FREE Full text] [doi: 10.1016/j.jaac.2014.10.010] [Medline: 25524788]
13. Costello E, Copeland W, Angold A. Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults? J Child Psychol Psychiatry 2011 Oct;52(10):1015-1025 [FREE Full text] [doi: 10.1111/j.1469-7610.2011.02446.x] [Medline: 21815892]
14. Ormel J, Raven D, van Oort F, Hartman CA, Reijneveld SA, Veenstra R, et al. Mental health in Dutch adolescents: a TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. Psychol Med 2015 Jan;45(2):345-360. [doi: 10.1017/S0033291714001469] [Medline: 25066533]
15. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
16. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
17. Klein DN, Glenn CR, Kosty DB, Seeley JR, Rohde P, Lewinsohn PM. Predictors of first lifetime onset of major depressive disorder in young adulthood. J Abnorm Psychol 2013 Feb;122(1):1-6 [FREE Full text] [doi: 10.1037/a0029567] [Medline: 22889243]
18. Bennett K, Manassis K, Duda S, Bagnell A, Bernstein GA, Garland EJ, et al. Treating child and adolescent anxiety effectively: Overview of systematic reviews. Clin Psychol Rev 2016 Dec;50:80-94. [doi: 10.1016/j.cpr.2016.09.006] [Medline: 27744168]
19. James AC, James G, Cowdrey FA, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev 2013 Jun 03(6):CD004690. [doi: 10.1002/14651858.CD004690.pub3] [Medline: 23733328]
20. Zhou X, Hetrick SE, Cuijpers P, Qin B, Barth J, Whittington CJ, et al. Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: A systematic review and network meta-analysis. World Psychiatry 2015 Jun;14(2):207-222 [FREE Full text] [doi: 10.1002/wps.20217] [Medline: 26043339]

21. Huang J, Nigatu YT, Smail-Crevier R, Zhang X, Wang J. Interventions for common mental health problems among university and college students: A systematic review and meta-analysis of randomized controlled trials. J Psychiatr Res 2018 Dec;107:1-10. [doi: 10.1016/j.jpsychires.2018.09.018] [Medline: 30300732]
22. Mojtabai R, Olfson M, Han B. National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults. Pediatrics 2016 Dec;138(6):e20161878 [FREE Full text] [doi: 10.1542/peds.2016-1878] [Medline: 27940701]
23. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
24. Cadigan JM, Lee CM, Larimer ME. Young Adult Mental Health: a Prospective Examination of Service Utilization, Perceived Unmet Service Needs, Attitudes, and Barriers to Service Use. Prev Sci 2019 Apr;20(3):366-376 [FREE Full text] [doi: 10.1007/s11121-018-0875-8] [Medline: 29411197]
25. Merikangas KR, He J, Burstein M, Swendsen J, Avenevoli S, Case B, et al. Service utilization for lifetime mental disorders in U.S. adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2011 Jan;50(1):32-45 [FREE Full text] [doi: 10.1016/j.jaac.2010.10.006] [Medline: 21156268]
26. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
27. Babajide A, Ortin A, Wei C, Mufson L, Duarte CS. Transition Cliffs for Young Adults with Anxiety and Depression: Is Integrated Mental Health Care a Solution? J Behav Health Serv Res 2020 Apr;47(2):275-292. [doi: 10.1007/s11414-019-09670-8] [Medline: 31428923]
28. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
29. MacDonell KW, Prinz RJ. A Review of Technology-Based Youth and Family-Focused Interventions. Clin Child Fam Psychol Rev 2017 Jun;20(2):185-200 [FREE Full text] [doi: 10.1007/s10567-016-0218-x] [Medline: 27787701]
30. Sweeney GM, Donovan CL, March S, Forbes Y. Logging into therapy: Adolescent perceptions of online therapies for mental health problems. Internet Interv 2019 Mar;15:93-99 [FREE Full text] [doi: 10.1016/j.invent.2016.12.001] [Medline: 30792959]
31. Cuijpers P, Ebert D, Reijnders M, Stikkelbroek Y. Technology-assisted treatments for mental health problems in children and adolescents. In: Evidence-Based Psychotherapies for Children and Adolescents. New York: The Guildford Press; 2017:555-576.
32. Clinical Practice Guideline for the Treatment of Depression Across Three Age Cohorts. American Psychological Association. 2019. URL: https://www.apa.org/depression-guideline [accessed 2019-10-24]
33. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
34. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
35. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
36. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
37. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
38. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: A systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
39. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
40. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: A systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
41. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]

42. Podina IR, Mogoase C, David D, Szentagotai A, Dobrean A. A Meta-Analysis on the Efficacy of Technology Mediated CBT for Anxious Children and Adolescents. J Rat-Emo Cognitive-Behav Ther 2015 Nov 17;34(1):31-50. [doi: 10.1007/s10942-015-0228-5]
43. Välimäki M, Anttila K, Anttila M, Lahti M. Web-Based Interventions Supporting Adolescents and Young People With Depressive Symptoms: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2017 Dec 08;5(12):e180 [FREE Full text] [doi: 10.2196/mhealth.8624] [Medline: 29222079]
44. Grist R, Croker A, Denne M, Stallard P. Technology Delivered Interventions for Depression and Anxiety in Children and Adolescents: A Systematic Review and Meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
45. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
46. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
47. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
48. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Cochrane handbook for systematic reviews of interventions. Hoboken: John Wiley & Sons; 2019.
49. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
50. Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group. 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2019-10-24]
51. Cohen J. Statistical power analysis for the behavioral sciences. Cambridge: Academic Press; 2013.
52. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry 2006 Jun 01;59(11):990-996. [doi: 10.1016/j.biopsych.2005.09.014] [Medline: 16368078]
53. Laupacis A, Sackett DL, Roberts RS. An Assessment of Clinically Useful Measures of the Consequences of Treatment. N Engl J Med 1988 Jun 30;318(26):1728-1733. [doi: 10.1056/nejm198806303182605]
54. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007 Nov 03;335(7626):914-916 [FREE Full text] [doi: 10.1136/bmj.39343.408449.80] [Medline: 17974687]
55. von Hippel PT. The heterogeneity statistic I(2) can be biased in small meta-analyses. BMC Med Res Methodol 2015 Apr 14;15:35 [FREE Full text] [doi: 10.1186/s12874-015-0024-z] [Medline: 25880989]
56. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
57. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
58. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: Randomised controlled trial. Behav Res Ther 2015 Oct;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
59. Botella C, Gallego MJ, Garcia-Palacios A, Guillen V, Baños RM, Quero S, et al. An Internet-based self-help treatment for fear of public speaking: a controlled trial. Cyberpsychol Behav Soc Netw 2010 Aug;13(4):407-421. [doi: 10.1089/cyber.2009.0224] [Medline: 20712499]
60. Sethi S. Treating Youth Depression and Anxiety: A Randomised Controlled Trial Examining the Efficacy of Computerised versus Face-to-face Cognitive Behaviour Therapy. Australian Psychologist 2013 Feb 14;48(4):249-257 [FREE Full text] [doi: 10.1111/ap.12006]
61. Ellis L, Campbell A, Sethi S, O'Dea B. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. Journal of CyberTherapy and Rehabilitation 2011;4(4):461-467.
62. Sethi S, Campbell A, Ellis L. The Use of Computerized Self-Help Packages to Treat Adolescent Depression and Anxiety. Journal of Technology in Human Services 2010 Aug 31;28(3):144-160 [FREE Full text] [doi: 10.1080/15228835.2010.508317]
63. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
64. Waite P, Marshall T, Creswell C. A randomized controlled trial of internet-delivered cognitive behaviour therapy for adolescent anxiety disorders in a routine clinical care setting with and without parent sessions. Child Adolesc Ment Health 2019 Sep;24(3):242-250. [doi: 10.1111/camh.12311] [Medline: 32677216]
65. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RC, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: A school-based program

(Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016 May;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
66. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
67. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
68. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
69. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
70. Bowler JO, Mackintosh B, Dunn BD, Mathews A, Dalgleish T, Hoppitt L. A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: effects on anxiety, depression, attentional control, and interpretive bias. J Consult Clin Psychol 2012 Dec;80(6):1021-1033 [FREE Full text] [doi: 10.1037/a0029932] [Medline: 22963595]
71. Ip P, Chim D, Chan KL, Li TM, Ho FK, Van Voorhees BW, et al. Effectiveness of a culturally attuned Internet-based depression prevention program for Chinese adolescents: A randomized controlled trial. Depress Anxiety 2016 Dec;33(12):1123-1131. [doi: 10.1002/da.22554] [Medline: 27618799]
72. McCall HC, Richardson CG, Helgadottir FD, Chen FS. Evaluating a Web-Based Social Anxiety Intervention Among University Students: Randomized Controlled Trial. J Med Internet Res 2018 Mar 21;20(3):e91 [FREE Full text] [doi: 10.2196/jmir.8630] [Medline: 29563078]
73. Day V, McGrath PJ, Wojtowicz M. Internet-based guided self-help for university students with anxiety, depression and stress: a randomized controlled clinical trial. Behav Res Ther 2013 Jul;51(7):344-351. [doi: 10.1016/j.brat.2013.03.003] [Medline: 23639300]
74. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-Guided, Internet-Delivered Cognitive-Behavioral Therapy for Adolescents With Obsessive-Compulsive Disorder: A Randomized Controlled Trial. J Am Acad Child Adolesc Psychiatry 2017 Jan;56(1):10-19.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]
75. Richards D, Timulak L, Rashleigh C, McLoughlin O, Colla A, Joyce C, et al. Effectiveness of an internet-delivered intervention for generalized anxiety disorder in routine care: A randomised controlled trial in a student population. Internet Interv 2016 Nov;6:80-88 [FREE Full text] [doi: 10.1016/j.invent.2016.10.003] [Medline: 30135817]
76. Stjerneklar S, Hougaard E, McLellan LF, Thastum M. A randomized controlled trial examining the efficacy of an internet-based cognitive behavioral therapy program for adolescents with anxiety disorders. PLoS One 2019;14(9):e0222485 [FREE Full text] [doi: 10.1371/journal.pone.0222485] [Medline: 31532802]
77. Tillfors M, Andersson G, Ekselius L, Furmark T, Lewenhaupt S, Karlsson A, et al. A randomized trial of Internet-delivered treatment for social anxiety disorder in high school students. Cogn Behav Ther 2011;40(2):147-157. [doi: 10.1080/16506073.2011.555486] [Medline: 25155815]
78. Wuthrich VM, Rapee RM, Cunningham MJ, Lyneham HJ, Hudson JL, Schniering CA. A randomized controlled trial of the Cool Teens CD-ROM computerized program for adolescent anxiety. J Am Acad Child Adolesc Psychiatry 2012 Mar;51(3):261-270. [doi: 10.1016/j.jaac.2011.12.002] [Medline: 22365462]
79. Topooco N, Berg M, Johansson S, Liljethörn L, Radvogin E, Vlaescu G, et al. Chat- and internet-based cognitive-behavioural therapy in treatment of adolescent depression: randomised controlled trial. BJPsych Open 2018 Jul;4(4):199-207 [FREE Full text] [doi: 10.1192/bjo.2018.18] [Medline: 29988969]
80. Stallard P, Richardson T, Velleman S, Attwood M. Computerized CBT (Think, Feel, Do) for depression and anxiety in children and adolescents: outcomes and feedback from a pilot randomized controlled trial. Behav Cogn Psychother 2011 May;39(3):273-284. [doi: 10.1017/S135246581000086X] [Medline: 21272393]
81. van der Zanden R, Kramer J, Gerrits R, Cuijpers P. Effectiveness of an online group course for depression in adolescents and young adults: a randomized trial. J Med Internet Res 2012 Jun 07;14(3):e86 [FREE Full text] [doi: 10.2196/jmir.2033] [Medline: 22677437]
82. Andersson G, Rozental A, Shafran R, Carlbring P. Long-term effects of internet-supported cognitive behaviour therapy. Expert Rev Neurother 2018 Jan;18(1):21-28. [doi: 10.1080/14737175.2018.1400381] [Medline: 29094622]
83. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
84. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]

85. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
86. Arjadi R, Nauta MH, Chowdhary N, Bockting CL. A systematic review of online interventions for mental health in low and middle income countries: a neglected field. Glob Ment Health (Camb) 2015;2:e12 [FREE Full text] [doi: 10.1017/gmh.2015.10] [Medline: 28596860]
87. World Health Organization. mhGAP: Mental Health Gap Action Programme: scaling up care for mental, neurological and substance use disorders. Geneva: WHO Press; 2008.
88. World Health Organization. mhGAP Intervention Guide for mental, neurological and substance use disorders in non-specialized health settings. Geneva: WHO Press; 2011.

Abbreviations

CBT: cognitive behavior therapy
cCBT: computerized cognitive behavior therapy
CMA: Comprehensive Meta-Analysis software
GRADE: Grading of Recommendations Assessment, Development and Evaluation
NNT: number needed to treat
RCT: randomized controlled trial
RoB 2: risk-of-bias tool for randomized trials, version 2.0
SMD: standardized mean difference
TAU: treatment as usual

Edited by G Eysenbach; submitted 15.01.20; peer-reviewed by A Radomski, L Valentine, M Simmonds-Buckley; comments to author 14.03.20; revised version received 23.05.20; accepted 21.06.20; published 25.09.20.

Please cite as:
Christ C, Schouten MJE, Blankers M, van Schaik DJF, Beekman ATF, Wisman MA, Stikkelbroek YAJ, Dekker JJM
Internet and Computer-Based Cognitive Behavioral Therapy for Anxiety and Depression in Adolescents and Young Adults: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(9):e17831
URL: https://www.jmir.org/2020/9/e17831
doi:10.2196/17831
PMID:32673212

Viewpoint

Doctors Routinely Share Health Data Electronically Under HIPAA, and Sharing With Patients and Patients' Third-Party Health Apps is Consistent: Interoperability and Privacy Analysis

Mark Savage[1*], JD; Lucia Clara Savage[2*], JD

[1]Center for Digital Health Innovation, University of California, San Francisco, CA, United States

[2]Omada Health, San Francisco, CA, United States

[*]all authors contributed equally

Corresponding Author:
Mark Savage, JD
Center for Digital Health Innovation
University of California, San Francisco
1700 Owens Street
Suite 541
San Francisco, CA, 94158
United States
Phone: 1 415 225 1676
Email: Mark.Savage@ucsf.edu

Abstract

Since 2000, federal regulations have affirmed that patients have a right to a complete copy of their health records from their physicians and hospitals. Today, providers across the nation use electronic health records and electronic information exchange for health care, and patients are choosing digital health apps to help them manage their own health and health information. Some doctors and health systems have voiced concern about whether they may transmit a patient's data upon the patient's request to the patient or the patient's health app. This hesitation impedes shared information and care coordination with patients. It impairs patients' ability to use the state-of-the-art digital health tools they choose to track and manage their health. It undermines the ability of patients' family caregivers to monitor health and to work remotely to provide care by using the nearly unique capabilities of health apps on people's smartphones. This paper explains that sharing data electronically with patients and patients' third-party apps is legally consistent under the Health Insurance Portability and Accountability Act (HIPAA) with routine electronic data sharing with other doctors for treatment or with insurers for reimbursement. The paper explains and illustrates basic principles and scenarios around sharing with patients, including patients' third-party apps. Doctors routinely and legally share health data electronically under HIPAA whether or not their organizations retain HIPAA responsibility. Sharing with patients and patients' third-party apps is no different and should be just as routine.

(J Med Internet Res 2020;22(9):e19818) doi:10.2196/19818

KEYWORDS

digital health; privacy; interoperability; mobile phone, smartphone; electronic health records; EHR; patient access; patient engagement; Health Insurance Portability and Accountability Act; HIPAA; Health Information Technology for Economic and Clinical Health Act; HITECH; covered entity; business associate; protected health information; PHI; digital health applications; apps

Introduction

Since 2000, federal regulations have affirmed that patients have a right to a complete copy of their health records from their physicians and hospitals. As the nation transitions to electronic health records (EHRs), electronic information exchange, and health apps that patients choose to help them manage their health and health information, some doctors and health systems have voiced concern about whether they may transmit a patient's data upon the patient's request to the patient or the patient's health app. Physicians worry about their liability under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) if, after transmitting the patient's data to the patient's health app, the app then breaches or improperly uses or discloses the data. This hesitation impedes shared information and care

coordination with patients. It impairs patients' ability to use the state-of-the-art digital health tools they choose to track and manage their health. It undermines the ability of patients' family caregivers to monitor health and to work remotely to provide care, using the nearly unique capabilities of health apps on people's smartphones.

So, on the road from the doctor's office to the patient's third-party app, where are HIPAA's green lights, yellow lights, and red lights for disclosing patients' protected health information as patients direct? We explain in detail why it's a green light all the way, and your patients' health and care are much the better for it because they can be engaged, informed, and shared decision-makers.

Eleven years after the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 [1], clinicians and health systems are increasingly accustomed to transmitting patients' health data electronically to other doctors, hospitals, labs, pharmacies, and payers, within and outside the sending doctor's system, for treatment and reimbursement. As of 2015, 96% of hospitals and 78% of physicians had adopted a certified EHR [2]. In 2017, 88% of hospitals and 36% of doctors were sending patients' health information electronically to care settings and organizations outside the doctors' health systems [3,4].

Many providers seem less comfortable, however, sharing a patient's health data electronically with the patient, and even more providers seem hesitant to share a patient's health data electronically with the patient's chosen health apps, even though patients have these rights. In 2000, HIPAA's Privacy Rule required that physicians provide patients with a copy of their health information in physicians' designated record sets (with some narrow exceptions). In the HITECH Act, Congress requires that physicians who use EHRs give patients electronic copies of their protected health information (PHI) and also requires that physicians who use EHRs follow a direction from a patient to transmit the patient's PHI electronically to any person, entity, or application the patient chooses [1,5,6]. The Office for Civil Rights has posted an excellent set of frequently asked questions documenting the patient's right to a copy of the patient's data and right to have that data sent electronically to any third-party app of the patient's choice [7-9]. These provisions of the HITECH Act, which apply to EHRs such as those we analyze here, are the law of the land [5].

Sometimes, the provider's resistance appears to be information blocking [10-14]. But for many, there is concern and uncertainty about transmitting a patient's data to a health app of unknown security and privacy protection and whether the physician or covered entity may be liable if the patient's app or its developer subsequently breaches or improperly uses or discloses the data.

This analysis should reassure. We explain that sharing health data electronically with patients and patients' third-party apps is required and is entirely consistent with physicians' routine electronic data sharing under HIPAA with other doctors for treatment or with insurers for reimbursement. This paper explains and illustrates basic principles and scenarios around sharing with patients, including patients' third-party apps. In many common scenarios, physicians' organizations retain responsibility under HIPAA after sharing, and in others, they do not. In short, doctors routinely share health data electronically under HIPAA, whether or not their organizations retain HIPAA responsibility [15,16]. Sharing with patients and patients' third-party apps is consistent and should be just as routine, just as banks routinely transmit account information to customers and their smartphone and third-party apps, such as Venmo. To be precise, there is one difference. Doctors' sharing with others for purposes of treatment, payment, and operations is *permitted* under the Privacy Rule [17], but doctor's sharing with patients and patients' third-party apps upon patients' request is *required* by law [18,19].

While this analysis should reassure, we must note a caveat. This overview serves educational purposes only and does not constitute legal advice. The principles and scenarios that follow illustrate generic situations. In actual situations, analysis depends upon specific facts, circumstances, contractual language, and relationships. However, this summary should help considerably to reduce the uncertainty and friction, and the sources we cite should be well known to providers' counsel. Moreover, we only address current legal requirements and practices under HIPAA that providers share patients' health data with their third-party apps upon request. This paper's scope does not cover how medical ethics and current policy debates treat these requirements. However, for deeper reading on ethics and policy proposals regarding health information disclosure, additional information can be found in [20-24].

References in this paper to terms such as a covered entity, business associate, PHI, disclosure, treatment, operations, use, and breach mean those terms as defined by the HIPAA Privacy Rule in 45 CFR §§160.103, 164.402, and 164.501 (2020). A business associate is a contractor or vendor that a covered entity hires to help that covered entity perform a wide range of health care functions that require that business associate to receive or collect, store, access, use, or disclose PHI. By "affiliated," we mean covered entities, their business associates (persons or entities that provide a service for or on behalf of a covered entity other than the provision of health care), and their agents. Conversely, by "unaffiliated," we mean entities or persons that are not legally affiliated under HIPAA, perhaps because they are an independent covered entity or an independent covered entity's business associate.

Part 1: Routine Data Sharing Under HIPAA

In general, when a doctor sends a patient's health data to another provider or system, privately and securely under the circumstances and in the manner allowed by law, the *recipient* is responsible for appropriately securing and handling the PHI after it is received. When a doctor sends a patient's health data to a doctor within the same health system or to the health system itself, the system's EHR, or its app, the *affiliated* health system is the recipient and retains responsibility and liability under HIPAA for any subsequent breach or improper disclosure. This should not surprise. The health system is the covered entity under HIPAA and remains responsible for the privacy and security of the health information in its custody and for sending

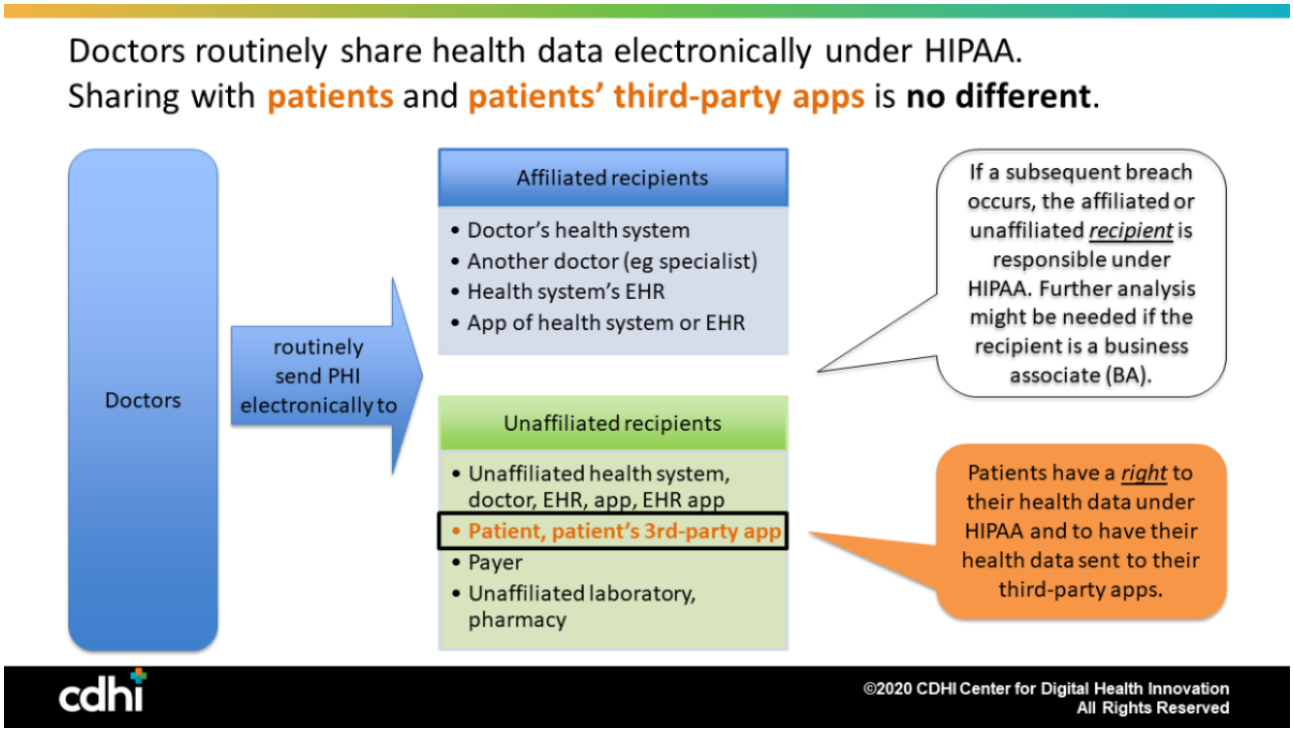
it securely to third parties only when permitted, directed, or authorized by the patient or required by law. Accordingly, the health system must make sure that its business associates, such as its EHR vendors, also abide by these rules.

Conversely, when a doctor sends a patient's health data to an *unaffiliated* doctor, health system, EHR, or EHR's app, the sending doctor's health system does not retain responsibility and liability under HIPAA for any subsequent breach or improper disclosure by that separate covered entity. Instead, the receiving covered entity is responsible under HIPAA for any breach or improper disclosure by itself or its business associates. The same result pertains when the doctor sends the patient's health data to other unaffiliated covered entities such as payers, laboratories, or pharmacies. This conclusion, too, should not surprise. Organizations expect to be responsible for their own mistakes and expect unaffiliated organizations to be responsible for their mistakes in turn. And HIPAA requires that each has systems in place to avoid mistakes in the first place.

The patient and the patient's third-party app are just another *unaffiliated* recipient. When a doctor sends a patient's health data to the patient or to the patient's third-party app at the patient's direction and the patient or third-party app subsequently misuses or allows a breach of the data, the doctor's health system does not retain responsibility and liability under HIPAA for that misuse. The responsibility belongs to the patient or developer of the patient's third-party app, just as the other unaffiliated recipients described in the preceding paragraph were responsible under HIPAA for their subsequent breach or improper disclosure. Sharing with patients and patients' third-party apps is no different.

In short, when doctors electronically send protected health information to affiliated recipients or to unaffiliated recipients and the recipient subsequently has a breach or improper disclosure of the data, the *recipient* is liable under HIPAA for its breach or improper disclosure. Doctors should feel just as comfortable with sharing a patient's health data electronically with the patient herself and her third-party health apps of choice, because the same rule applies (Figure 1).

Figure 1. Routine data sharing under the Health Insurance Portability and Accountability Act of 1996 (HIPAA). EHR: electronic health record; PHI: protected health information.



Part 2: Patients' Third-Party Health Apps

Next, we focus on *third-party* apps in more detail, as they seem to be a source of concern or confusion for some doctors and health systems.

In the typical scenario, a patient selects a third-party health app, perhaps from a smartphone or app store, perhaps to help her and her caregivers manage her chronic conditions more effectively. The patient directs her doctor to send a copy of her PHI to the app. The doctor is part of a health system, which is a covered entity under HIPAA, and the covered entity sends the data to the patient's third-party app as requested.

The relationship between the sending covered entity and the third-party app's developer, regarding the particular exchange of PHI in question, determines responsibility or liability under HIPAA for a subsequent breach or inappropriate disclosure of the patient's information.

When the third-party app's developer is an *unaffiliated* covered entity or its business associate or not a covered entity or business associate at all, any "breach" or "improper" use or disclosure under HIPAA would not subject the sending covered entity to liability under HIPAA. For example, when a doctor discloses PHI to an unaffiliated health system's app for care by a specialist — such as an app for asthma management, heart monitoring,

or fertility tracking — the sending doctor and health system are not liable under HIPAA if the recipient app should subsequently breach or improperly disclose the data. The analysis is the same if hospital A discloses to unaffiliated hospital B, to hospital B's app that hospital B uses to deliver health care, or to an app developed by hospital B's app developer. In each situation, hospital B has its duties and liability under HIPAA to protect the PHI it received.

When the app's developer is instead *affiliated* with the sending covered entity, the health system generally retains liability under HIPAA for misuses or breaches of PHI by apps it uses or paid to develop. However, further analysis may be necessary, to determine whether the app developer and app were acting on behalf of the sending health system for the particular exchange of PHI in question.

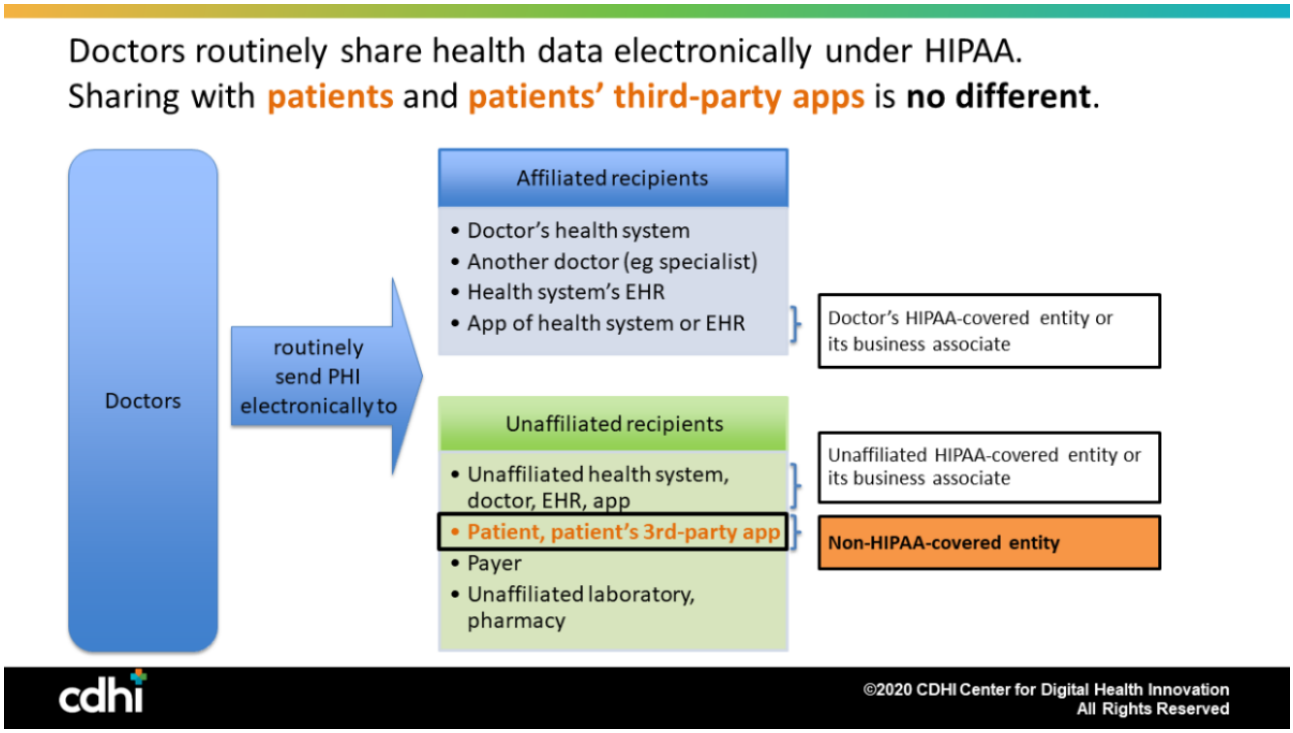
In some rare circumstances, the covered entity may *not* be liable if the app developer's conduct was outside, or unauthorized by, its contract with the covered entity. Conversely, even though the app developer's conduct was outside the terms of the business associate agreement (BAA), the covered entity may nevertheless retain some liability for having failed to oversee its app developer or to take action on some activity it should have known was a misuse of the PHI. These are always fact-specific situations, but the following questions illustrate what types of facts are salient: (1) Does the BAA endorse, permit, or not prohibit the business associate's act in question? (2) Did the sending covered entity know in advance, or should it have known, about the business associate's act? While the sending health system may have no liability under HIPAA for its business associate's breach per se, a covered entity still has duties to report and to address a breach of unsecured PHI once discovered (or once it should have been discovered) and may not ignore suspected inappropriate use of data. Liability may adhere for failure to do so. The health system may also have liability if its business associate's breach entailed noncompliance with HIPAA and the health system knew or should have known about the noncompliance and did not address it. (3) Was the business associate acting as an agent (in the legal sense) of the sending covered entity? Common-law agency generally exists when the sending health system controls or retains authority to control the business associate's actions with interim direction or instructions as the business associate performs services on behalf of the health system [25].

In some scenarios, therefore, an affiliated app developer may be acting outside the scope of the business associate relationship, and the covered entity may not be liable for the app developer's extracontractual activity. However, absent facts that support that conclusion, the general principle remains: The health system retains responsibility under HIPAA for its and its business associate's apps, as usual.

This analysis does not change when the patient directs the health system to send the patient's PHI to the patient's health app. The health system's liability still depends upon the relationship under HIPAA between the sending covered entity and the app's developer. In general, the sending covered entity or its business associate will not be liable under HIPAA for subsequent use or disclosure, unless the app developer is a business associate of and providing services on behalf of the sending covered entity with respect to the disclosure. The patient's directive to send the PHI is not the salient fact; the salient fact is whether a business associate relationship exists between the health system and the app developer [26].

This introduces an important point that is not reflected in Figure 1. In Figure 1, both the affiliated and unaffiliated apps and app developers were covered by HIPAA and had duties under HIPAA to protect the privacy and security of the patient's PHI. For example, Omada Health is a covered entity under 45 CFR §160.103 (2020) and is required by law to comply with HIPAA for any PHI it holds. When the apps the patient chooses are *not* HIPAA-covered entities or are not performing their services as a business associate of a covered entity, HIPAA's requirements and protections do not apply (Figure 2). Here, too, sharing patients' data with their non-HIPAA–covered apps is no different: Doctors routinely share patients' health data with entities not covered by HIPAA, such as public health agencies and researchers. Although HIPAA does not apply, other statutes and regulations, such as the Federal Trade Commission's consumer protection regulations, may apply [24], and this topic has been under active consideration since May 2018 by Congress in various legislative proposals, and more than a half dozen general privacy bills are currently pending. For example, both Senator Wicker's bill (the Consumer Data Privacy Act) and Senator Cantwell's bill (the Consumer Online Privacy Rights Act) propose nationwide consumer privacy laws that would require transparent explanations of privacy practices; the right to delete, correct, or port one's data; choices about collection of certain types of sensitive data, like health data outside HIPAA; and restrictions on use by people other than the original data collector [27,28]. Likewise, while the sending covered entity may have no liability under HIPAA, it may still have liability under other laws or duties, such as medical malpractice or negligence in recommending an app for use.

Figure 2. Routine data sharing with Health Insurance Portability and Accountability Act of 1996 (HIPAA)-covered and non-HIPAA–covered entities. EHR: electronic health record; PHI: protected health information.



Part 3: Twelve Common Scenarios

We illustrate these general principles with 12 common scenarios. In the first set of 3 scenarios, the doctor shares the patient's data with other parts of the same health system, such as an affiliated specialist, the system's EHR, or an app the system uses. The second set (scenarios 4-6) covers examples where the doctor instead shares the patient's data with an unaffiliated health system or covered entity. The third set (scenarios 7-9) focuses on complexities that may arise when the doctor shares the patient's data with an app affiliated with the doctor's health system. In these scenarios, the responsibility for breach or improper use depends upon a closer look at the facts and circumstances. Lastly, the fourth set (scenarios 10-12) covers examples where, pursuant to a patient's direction, the doctor shares the patient's data with a third-party app the patient independently chose. Together, these scenarios illustrate how sharing data with patients' third-party apps sits comfortably and consistently within the range of situations where doctors routinely share patients' data with their own and other health systems.

Sharing Patients' Data Within the Doctor's Health System

As mentioned, we begin with scenarios where doctors are accustomed to sharing a patient's health data. When doctors share health information with affiliated doctors and apps within the same health system or covered entity, they know that, under HIPAA, the health system remains liable for any breach or improper disclosure by the same health system. When doctors share health information with unaffiliated doctors and apps, then the recipient's covered entity or business associate is liable under HIPAA. Doctors already routinely share with both.

Scenarios in which the doctors share the patients' data with other parts of the same health system:

- Scenario 1: A health system's emergency room doctor shares a patient's data with the health system's pulmonologist for the same patient, and after that sharing, there's a breach or improper disclosure of the data. The health system retains liability where one doctor shares the patient's data with another doctor within the same health system or covered entity and that receiving doctor improperly uses or discloses the data under HIPAA.
- Scenario 2: A health system's endocrinologist uses a device or app to share a patient's data with the health system's EHR, and the EHR subsequently has a breach or improper disclosure of the data. The health system retains liability where the doctor shares the patient's data with the health system's EHR (the EHR vendor being a business associate) and the EHR or EHR vendor improperly uses or discloses the data.
- Scenario 3: A health system's cardiologist shares a patient's data from the EHR to a medication management app that the cardiologist has prescribed and the health system developed. The app causes a subsequent breach or improper disclosure of the data. Again, the health system retains liability where the doctor shares the patient's data with a health app that the doctor prescribed and the health system developed to integrate the data with its EHR for the patient's care.

Sharing Patients' Data With Other Health Systems

We next consider scenarios where doctors know that, after sharing patients' health data, their health systems no longer retain liability under HIPAA for a subsequent breach or improper disclosure. When doctors share with unaffiliated doctors and EHRs outside their health system and its business

associates, they usually know that the receiving covered entity assumes the liability under HIPAA for any subsequent breach or improper disclosure.

Scenarios in which the doctors know that, after sharing patients' health data, their health systems no longer retain liability under HIPAA for a subsequent breach or improper disclosure:

- Scenario 4: A health system's doctor shares a patient's data with a different health system's doctor for purposes of a second opinion or shares a patient's billing data with the patient's health insurer for reimbursement, and the recipient doctor or insurer subsequently has a breach or improper disclosure of the data. The sending health system is not liable when the recipient subsequently has a breach or improperly discloses the patient's data. Instead, the separate covered entity is responsible under HIPAA for any breach or improper disclosure by itself or its business associates.
- Scenario 5: A health system's doctor shares a patient's data with a different health system's EHR and that EHR subsequently has a breach or improperly discloses the data. Again, the sending health system is not liable when its doctor shares the patient's data with a separate covered entity's EHR and that EHR subsequently has a breach or improperly discloses the patient's data. Instead, the EHR's vendor is a business associate of the separate covered entity, and the separate covered entity or business associate is responsible under HIPAA for the breach or improper disclosure.
- Scenario 6: A health system's doctor shares the patient's data with a personal health record (PHR) app that the patient has chosen and which the doctor's health system did not develop. The doctor uses the health system's EHR to transmit the health data. The patient's PHR app subsequently has a breach or improper disclosure of the data. The sending health system is likewise not liable for the third-party app's breach, although other laws may make the PHR app liable for its misuse or breach [29].

Sharing Patients' Data With Apps Affiliated With the Doctor's Health System

Next, we consider scenarios where doctors share their patients' data with affiliated apps, but liability for an app's subsequent breach or improper disclosure depends upon the facts and circumstances. In each of the scenarios below, where liability lies will depend in part on whether the app developer's improper use or disclosure fell within or outside the scope of its authority and responsibilities under a BAA with the covered entity, as described in the scenario. Typically, a health system's legal office will negotiate the BAA with an EHR vendor, an app developer, or a staffing agency, and that negotiation will document the rules to resolve these facts and circumstances.

Scenarios in which the doctors share their patients' data with affiliated apps, but liability for an app's subsequent breach or improper disclosure depends upon the facts and circumstances:

- Scenario 7: A health system's doctor shares a patient's data with an app that the health system itself or a business associate developed for the system to integrate with its EHR, and the app subsequently has a breach or improper disclosure of the data. Whether the sending health system is liable depends, for example, on whether the health system developed the app or is paying for the app to be available to patients (liable) or a business associate developed the app (may be liable depending on whether the EHR app's breach or improper disclosure fell within the scope of the app developer's authority and responsibilities under the BAA).
- Scenario 8: A health system's doctor shares a patient's data with the health system's app. The app's developer uses the data to push ads about itself to the patient's friends and family through their social media accounts (which are not regulated by HIPAA). In analyzing liability, even assuming that the app developer was prohibited from this activity, the covered entity might be liable if it knew or should have known that the app developer was using the health information for advertising, as we've already discussed.
- Scenario 9: A health system's doctor shares a patient's data with the health system's EHR, and the EHR vendor uses the data to create a research database [30]. Similarly, whether the sending health system is liable depends on whether the EHR's or EHR vendor's creation of the research database fell within the scope of the EHR developer's authority and responsibilities under the BAA. Does the BAA allow or prohibit such a secondary use by the EHR vendor?

Sharing Patients' Data With Patients' Third-Party Apps

The final 3 scenarios concern an app that the patient chooses and the doctor did not sponsor or pay to make available to the patient. The scenarios illustrate, for example, whether the doctor or patient chose the app for treatment. If the patient selected the app, does the analysis change because the doctor subsequently looked at data from the app or even asked the patient to keep sharing the app's data with the doctor? If the doctor selected the app, but the patient directed the doctor to send the particular data to the app, does the analysis change? We explain why such factors *do not change* the basic principles and results under HIPAA, as already discussed. If the doctor securely sends the data to the patient's app as directed, the physician is not liable under HIPAA for the app's conduct after it receives the patient's data.

Scenarios in which the doctors share their patients' data with patients' third-party apps:

- Scenario 10: A patient selects and uses an unaffiliated third-party app such as a fitness tracker, then visits the doctor. The doctor recommends that the patient keep using the app and send the data to the doctor; the patient uploads the data from the fitness tracker to the doctor's health system through the system's patient portal. The app subsequently has a breach or improper disclosure of the data. The sending doctor's health system is not liable for the patient's app's breach. The patient's third-party app developer is not a business associate of the sending covered entity. The doctor's recommendation and request that the patient show the results to the doctor do not create a

business associate relationship between the covered entity and the patient's third-party app.

- Scenario 11: A patient independently purchases and uses a third-party app such as a diabetes tracker and then visits the doctor (unbeknownst to the patient, the doctor's health system also uses that app and has a business associate relationship with the app's developer). After the patient shares the app's data with the doctor, the doctor recommends that the patient keep using the app and send the data to the doctor. The app subsequently has a breach or improper disclosure of the data. The sending health system is still not liable for the third-party app's breach or improper disclosure. The patient selected the app, and the app was acting on behalf of and providing services for the patient, not the health system nor its doctor. The fact that a business associate relationship independently exists between the health system and the app's developer when the doctor prescribes the app for treatment does not create a business associate relationship here, where the patient purchased the app and the app was acting on behalf of the patient.
- Scenario 12: A patient visits a doctor, and the doctor recommends an affiliated app (the doctor's health system and the app developer have a contract to provide the app and integrate the app's data in the health system's EHR). The patient downloads and uses the app, and the app subsequently has a breach or improper disclosure of the data. The app developer is liable under HIPAA for the breach, and the system may be liable if it knew or should have known of the conditions that led to the breach. This is because the app is providing services on behalf of the covered entity, not the patient. It does not matter whether the doctor "recommended" or "prescribed" the app.

Multimedia Appendix 1 introduces some other factors that might or might not cause a doctor uncertainty about whether to transmit the patient's data and who is liable for a subsequent breach or improper disclosure.

These 12 scenarios all illustrate and bring us back to the common principle in Figure 1. Doctors routinely send protected health information electronically to affiliated recipients (such as doctors in their same health systems and their health systems' EHRs) and to unaffiliated recipients (such as doctors at different health systems, laboratories, pharmacies, and payers). In both situations, the *recipient* is responsible under HIPAA for its breach or improper disclosure. If the recipient is part of the doctor's covered entity, then the doctor's health system retains responsibility under HIPAA. If the recipient is *not* part of the doctor's covered entity, then the recipient's covered entity or business associate is liable under HIPAA. Doctors are already routinely sharing with both.

Patients and patients' third-party apps are no different. Neither patients nor apps they choose independently are recipients affiliated with the doctor or health system, so the doctor's health system does not retain liability if the patient or the patient's third-party app subsequently has a breach or improper disclosure of the data.

Conclusion

Patients have a legal *right* under HIPAA to a copy of their health data and to have their health data sent electronically to a third-party app of their choice. As we have explained, doctors routinely disclose PHI appropriately to other legitimate recipients of that PHI and are not liable under HIPAA for what those recipients do with the PHI. Given these well-established rules and practices, doctors and their health systems should be equally confident in routinely sharing patients' health data electronically with patients and their third-party apps.

Acknowledgments

The authors acknowledge and thank Deven McGraw for her invaluable review of and comments on the principles and scenarios in the appendix.

Conflicts of Interest

MS has received research support from Cisco Systems, Inc.

Multimedia Appendix 1
Reference appendix of real-world scenarios with providers sharing health data with patients and their third-party health apps.
[PDF File (Adobe PDF File), 139 KB - jmir_v22i9e19818_app1.pdf]

References

1. Health Information Technology for Economic and Clinical Health Act of 2009, Pub L 111-5, §§13001-13424, 4001-4302, 123 Stat 115, 226-279, 467-496. 2009 Feb 17. URL: https://www.congress.gov/111/plaws/publ5/PLAW-111publ5.pdf [accessed 2020-05-05]
2. The Office of the National Coordinator for Health Information Technology (ONC), Office of the Secretary, United States Department of Health and Human Services (HHS). 2016 Report to Congress on Health IT Progress: Examining the HITECH Era and the Future of Health IT. 2016 Nov. URL: https://dashboard.healthit.gov/report-to-congress/2016-report-congress-examining-hitech-era-future-health-information-technology.php [accessed 2020-05-04]

3. Pylypchuk Y, Johnson C, Henry J, Ciricean D. Variation in Interoperability among U.S. Non-federal Acute Care Hospitals in 2017: ONC data brief 42. Office of the National Coordinator for Health Information Technology. 2018. URL: https://www.healthit.gov/sites/default/files/page/2018-11/Interop%20variation_0.pdf [accessed 2020-05-04]
4. Patel V, Pylypchuk Y, Parasrampuria S, Kachay L. Interoperability among Office-Based Physicians in 2015 and 2017: ONC data brief 47. Office of the National Coordinator for Health Information Technology. 2019 May. URL: https://www.healthit.gov/sites/default/files/page/2019-05/ONC-Data-Brief-47-Interoperability-among-Office-Based-Physicians-in-2015-and-2017.pdf [accessed 2020-05-05]
5. Ciox Health, LLC v Alex Azar, et al., Case No 18-cv-00040 (APM), 2020 US Dist LEXIS 12801. United States District Court for the District of Columbia. 2020 Jan 23. URL: https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2018cv0040-51 [accessed 2020-08-02]
6. 42 USC §17935. Health Information Technology for Economic and Clinical Health Act of 2009, Pub L 111-5, §13405(e), 123 Stat 115, 264, 268. 2009 Feb 17. URL: https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf [accessed 2020-05-05]
7. Office for Civil Rights, U.S. Department of Health and Human Services. Individuals' Right under HIPAA to Access their Health Information 45 CFR § 164. 2020 Jan 31. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html [accessed 2020-05-04]
8. Office for Civil Rights, U.S. Department of Health and Human Services. Can a covered entity refuse to disclose ePHI to an app chosen by an individual because of concerns about how the app will use or disclose the ePHI it receives?. 2019 May 09. URL: https://www.hhs.gov/hipaa/for-professionals/faq/3012/can-a-covered-entity-refuse-to-disclose-ephi.html [accessed 2020-05-04]
9. Does a HIPAA covered entity that fulfills an individual's request to transmit electronic protected health information (ePHI) to an application or other software (collectively "app") bear liability under the HIPAA Privacy, Security, or Breach Notification Rules (HIPAA Rules) for the app's use or disclosure of the health information it received? Office for Civil Rights, U.S. Department of Health and Human Services. 2019 May 09. URL: https://www.hhs.gov/hipaa/for-professionals/faq/3009/does-a-hipaa-covered-entity-bear-liability.html [accessed 2020-05-04]
10. 42 USC §300jj-52. 21st Century Cures Act, Pub L 114-255, §4004, 130 Stat 1033, 1176-1177. 2016 Dec 13. URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2020-05-04]
11. Office of the National Coordinator for Health Information Technology. Report to Congress on Health Information Blocking. 2015 Apr. URL: https://www.healthit.gov/sites/default/files/reports/info_blocking_040915.pdf [accessed 2020-05-05]
12. Savage L, Gaynor M, Adler-Milstein J. Digital Health Data and Information Sharing: A New Frontier for Health Care Competition? Antitrust Law Journal 2019 Apr;82(2):610-612 (discussing anticompetitive motivations for health information blocking) [FREE Full text]
13. Adler-Milstein J, Pfeifer E. Information Blocking: Is It Occurring and What Policy Strategies Can Address It? Milbank Q 2017 Feb 28;95(1):124-129 (in a national survey of health information organizations, respondents report that hospitals and health systems often or routinely control patient flow by selectively sharing patient health information (22% of respondents) and using HIPAA as a barrier to patient health information sharing when actually HIPAA is not a barrier (15% of respondents)) [FREE Full text] [doi: 10.1111/1468-0009.12247] [Medline: 28266065]
14. U.S. Department of Health and Human Services. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. Federal Register 2019 Mar 04;84(42):7508 (in a national survey of health information organizations, a quarter of respondents reported that hospitals and health systems routinely engage in information blocking) [FREE Full text]
15. Office of the National Coordinator for Health Information Technology, Office for Civil Rights, U.S. Department of Health and Human Services. Permitted Uses and Disclosures: Exchange for Treatment: 45 Code of Federal Regulations (CFR) 164.506(c)(2). 2016 Jan. URL: https://www.healthit.gov/sites/default/files/exchange_treatment.pdf [accessed 2020-05-04]
16. Mello MM, Adler-Milstein J, Ding KL, Savage L. Legal Barriers to the Growth of Health Information Exchange--Boulders or Pebbles? Milbank Q 2018 Mar;96(1):132 ("HIPAA presents no obstacles whatsoever to sharing PHI for treatment and operations purposes and imposes no liability for downstream privacy breaches on disclosing entities that follow the rules.") [FREE Full text] [doi: 10.1111/1468-0009.12313] [Medline: 29504197]
17. 45 CFR §164.502(a)(1)(ii). Electronic Code of Federal Regulations. 2020 Aug 04. URL: https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=1&SID=dd754cb66cdcc1b75e04b011f09b27fd&ty=HTML&h=L&mc=true&r=SECTION&n=se45.2.164_1502 [accessed 2020-08-06]
18. 45 CFR §164.502(a)(2)(i). Electronic Code of Federal Regulations. 2020 Aug 04. URL: https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=1&SID=dd754cb66cdcc1b75e04b011f09b27fd&ty=HTML&h=L&mc=true&r=SECTION&n=se45.2.164_1502 [accessed 2020-08-21]
19. 45 CFR §164.524. Electronic Code of Federal Regulations. 2020 Aug 04. URL: https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=1&SID=dd754cb66cdcc1b75e04b011f09b27fd&h=L&mc=true&r=SECTION&n=se45.2.164_1524 [accessed 2020-08-06]
20. American Medical Association. Code of Medical Ethics Opinion 3.2.1. 2016 Nov 14. URL: https://www.ama-assn.org/delivering-care/ethics/confidentiality [accessed 2020-07-05]

21. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019 Apr 19;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
22. Savage L. Why we must remember where informed consent comes from. International Association of Privacy Professionals. 2018 Sep 14. URL: https://iapp.org/news/a/why-we-must-remember-where-informed-consent-comes-from/ [accessed 2020-07-05]
23. Armontrout J, Torous J, Fisher M, Drogin E, Gutheil T. Mobile Mental Health: Navigating New Rules and Regulations for Digital Tools. Curr Psychiatry Rep 2016 Aug 23;18(10):91. [doi: 10.1007/s11920-016-0726-x] [Medline: 27553979]
24. U.S. Department of Health and Human Services. Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA. 2016 Jun 17. URL: https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf [accessed 2020-07-05]
25. U.S. Department of Health and Human Services. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules; Final Rule. Federal Register 2013 Jan 25;78(17):5580-5581 [FREE Full text] [doi: 10.1177/1942602x10390377]
26. Office for Civil Rights, U.S. Department of Health and Human Services. Health App Use Scenarios & HIPAA. 2016 Feb. URL: https://hipaaqsportal.hhs.gov/community-library/accounts/92/925889/Public/OCR-health-app-developer-scenarios-2-2016.pdf [accessed 2020-05-05]
27. Consumer Online Privacy Rights Act, S 2968, 116th Cong. 2019 Dec 3. URL: https://www.congress.gov/116/bills/s2968/BILLS-116s2968is.pdf [accessed 2020-05-05]
28. United States Consumer Data Privacy Act of 2009, Staff Discussion Draft, 116th Cong. 2019. URL: https://privacyblogfullservice.huntonwilliamsblogs.com/wp-content/uploads/sites/28/2019/12/Nc7.pdf [accessed 2020-05-05]
29. Federal Trade Commission. Health Breach Notification Rule, 16 CFR part 318. 2009 Aug 25. URL: https://www.govinfo.gov/content/pkg/CFR-2016-title16-vol1/pdf/CFR-2016-title16-vol1-part318.pdf [accessed 2020-08-05]
30. Drees J. Epic unveils patient data research initiative, new software. Becker's Hospital Review. 2019 Aug 29. URL: https://www.beckershospitalreview.com/ehrs/epic-unveils-patient-data-research-initiative-new-software.html [accessed 2020-05-04]

Abbreviations

BAA: business associate agreement
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act of 1996
HITECH: Health Information Technology for Economic and Clinical Health Act of 2009
PHI: protected health information
PHR: personal health record

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by D McGraw, J Ropero; comments to author 28.05.20; revised version received 17.07.20; accepted 22.07.20; published 02.09.20.

Please cite as:
Savage M, Savage LC
Doctors Routinely Share Health Data Electronically Under HIPAA, and Sharing With Patients' Third-Party Health Apps is Consistent: Interoperability and Privacy Analysis
J Med Internet Res 2020;22(9):e19818
URL: https://www.jmir.org/2020/9/e19818
doi:10.2196/19818
PMID:32876582

Viewpoint

Telemonitoring for Patients With COVID-19: Recommendations for Design and Implementation

Anna V Silven[1,2*], MD; Annelieke H J Petrus[1,2*], MD, PhD; María Villalobos-Quesada[1,2*], PhD; Ebru Dirikgil[3], MD; Carlijn R Oerlemans[3], BSc; Cyril P Landstra[3,4], MD; Hileen Boosman[5], PhD; Hendrikus J A van Os[1,2,6,7], MD; Marco H Blanker[8], MD, PhD; Roderick W Treskes[9], MD, PhD; Tobias N Bonten[1,2], MD, PhD; Niels H Chavannes[1,2], MD, PhD; Douwe E Atsma[2,9], MD, PhD; Y K Onno Teng[3], MD, PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands
[2]National eHealth Living Lab, Leiden University Medical Center, Leiden, Netherlands
[3]Department of Nephrology, Leiden University Medical Center, Leiden, Netherlands
[4]Department of Endocrinology, Leiden University Medical Center, Leiden, Netherlands
[5]Department of Quality and Patient Safety, Leiden University Medical Center, Leiden, Netherlands
[6]Department of Neurology, Leiden University Medical Center, Leiden, Netherlands
[7]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands
[8]Department of General Practice and Elderly Care Medicine, University Medical Center Groningen, Groningen, Netherlands
[9]Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands
[*]these authors contributed equally

Corresponding Author:
Douwe E Atsma, MD, PhD
Department of Cardiology
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 715262038
Email: d.e.atsma@lumc.nl

Abstract

Despite significant efforts, the COVID-19 pandemic has put enormous pressure on health care systems around the world, threatening the quality of patient care. Telemonitoring offers the opportunity to carefully monitor patients with a confirmed or suspected case of COVID-19 from home and allows for the timely identification of worsening symptoms. Additionally, it may decrease the number of hospital visits and admissions, thereby reducing the use of scarce resources, optimizing health care capacity, and minimizing the risk of viral transmission. In this paper, we present a COVID-19 telemonitoring care pathway developed at a tertiary care hospital in the Netherlands, which combined the monitoring of vital parameters with video consultations for adequate clinical assessment. Additionally, we report a series of medical, scientific, organizational, and ethical recommendations that may be used as a guide for the design and implementation of telemonitoring pathways for COVID-19 and other diseases worldwide.

(J Med Internet Res 2020;22(9):e20953) doi:10.2196/20953

KEYWORDS

telemonitoring; telemedicine; eHealth; digital health; COVID-19

Introduction

The COVID-19 pandemic has created a difficult challenge for global public health [1]. The clinical presentations of COVID-19 are highly variable, ranging from asymptomatic patients to patients with severe symptoms. Several reports indicate that patients who initially present only mild-to-moderate symptoms can show a deterioration to severe symptoms over the course of only a few days [2]. Eventually, approximately 5% of patients need respiratory support and intensive care unit admission. Patients with chronic underlying conditions are at an increased risk of such a severe course of illness when infected with COVID-19. Early identification of hypoxia and/or a deterioration of symptoms may be lifesaving [3-5]. However,

in-hospital monitoring of all patients with an increased risk of severe disease puts even more pressure on health care systems that are already overwhelmed.

Telemonitoring offers the opportunity to closely monitor symptoms and vital parameters while a patient remains at home. As such, telemonitoring may enable early identification of deterioration of symptoms and allows for appropriate treatments for each patient with COVID-19. Additionally, telemonitoring may reduce the number of hospital visits and admissions, thereby decreasing the usage of personal protective material, reducing the pressure on health care personnel, and minimizing the risk of viral transmission. Telemonitoring essentially holds the promise of optimizing care for patients with (suspected) COVID-19 infection while ensuring the sustainability of health care capacity and resources for those who need it most urgently [6-8]. However, the development and implementation of a telemonitoring care pathway in the clinical workflow is challenging. The novelty of COVID-19 and the need for rapid action pose additional difficulties. Based on our extensive experiences with telemonitoring for patients with chronic diseases, we developed and implemented a telemonitoring care pathway for patients with (suspected) COVID-19 [6].

In this paper, we present our telemonitoring care pathway as an example and provide general recommendations for the implementation of telemonitoring care programs for patients with (suspected) COVID-19. Hence, the aim of this paper is to serve as a guide for the design and implementation of telemonitoring care pathways in other health care institutions, in order to optimize care and health care capacity worldwide.

The COVID Box: An Example of a Telemonitoring Care Pathway

Aim and Setting

Telemonitoring for patients with (suspected) COVID-19 can be performed in many ways [9]. At our tertiary care hospital, Leiden University Medical Center (LUMC), telemonitoring is broadly applied for the management of several diseases and has been found to be a safe and effective alternative to standard care [6]. Based on this experience, the “COVID Box” telemonitoring care pathway was developed and implemented as standard care for patients with (suspected) COVID-19 over the course of a few weeks (Figures 1 and 2). The aim of this telemonitoring care pathway is to provide optimal care to patients with (suspected) COVID-19, while avoiding (unnecessary) hospitalizations and reducing the duration of hospitalizations.

Figure 1. Types of patient journeys for individuals with (suspected) COVID-19. In the Netherlands, patients are only assessed at the emergency department after referral by their general practitioner or treating medical specialist (indicated by "a"). The treating physician eventually decides which patients are in need for close monitoring (indicated by "b").

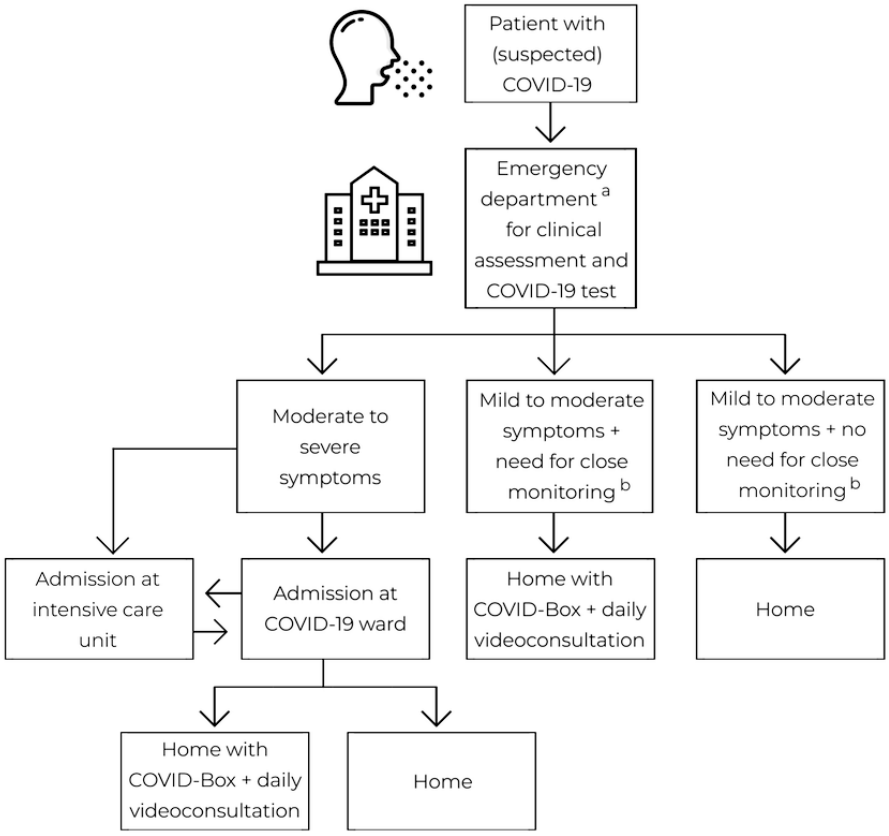
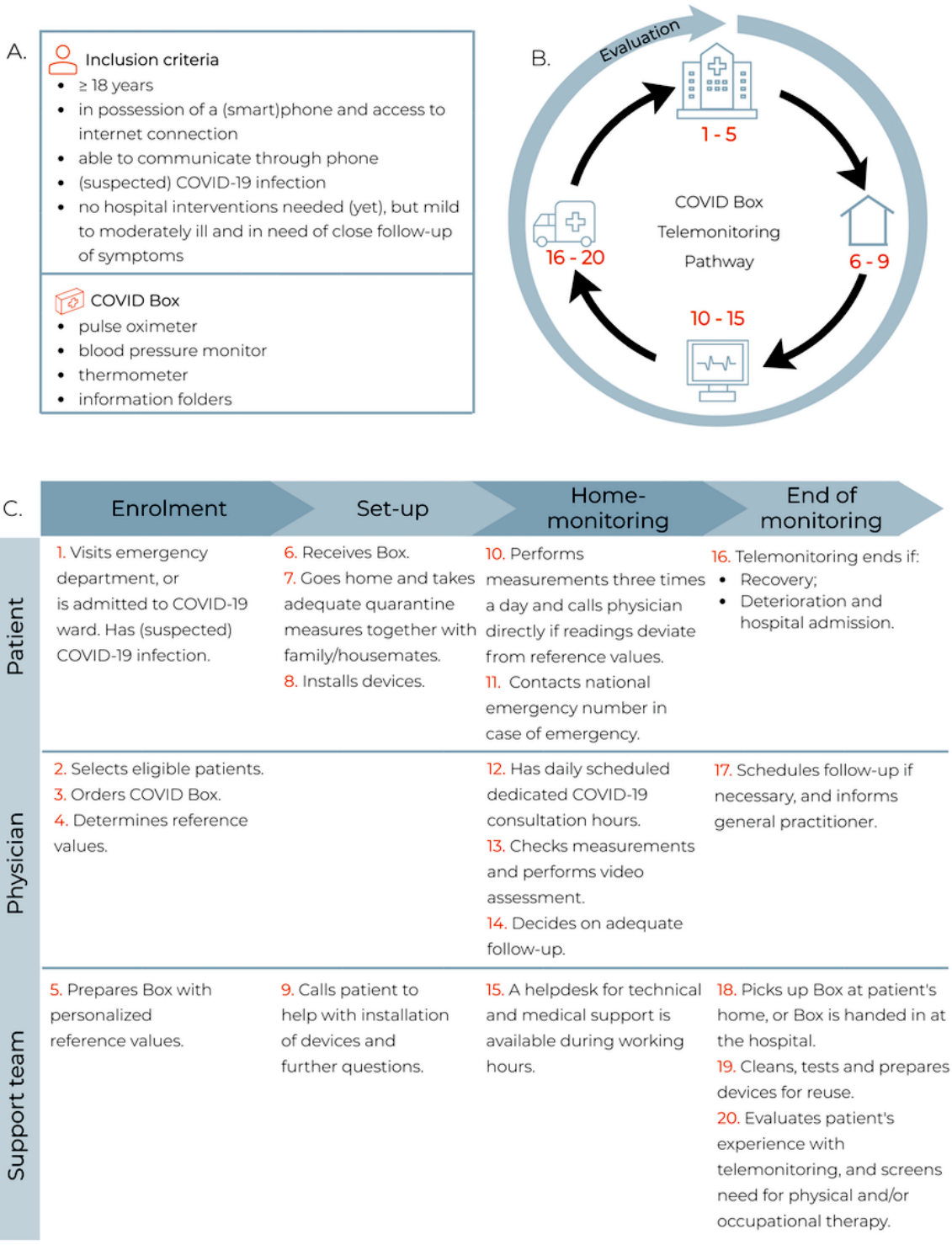


Figure 2. An example of a telemonitoring care pathway—the COVID Box. The COVID Box is a program developed by the Leiden University Medical Center to monitor patients with confirmed or suspected COVID-19 who have an increased risk of severe illness. (A) Inclusion criteria for the COVID Box; (B) general overview of the telemonitoring care pathway including continuous evaluation, where numbers 1-20 correspond to the steps described in (C), which provide a stepwise description of the COVID Box care pathway.



Patient Selection and Eligibility

Patients are selected for telemonitoring when visiting the emergency care department for (suspected) COVID-19 (after referral by their general practitioner), or after admission to the COVID-19 department. Patients are eligible for telemonitoring using the COVID Box in the case of a (suspected) COVID-19 infection and when exhibiting mild-to-moderate symptoms, with a need of close monitoring of vital signs and symptoms, as judged by the treating physician. In addition, psychological and social patient characteristics are taken into account.

Home Monitoring

The COVID Box is literally a box containing a thermometer (Withings Thermo), pulse oximeter (Masimo MightySatRx), blood pressure monitor (Microlife BP B2 Basic or Withings

BPM Connect), and a safety bag (for return of the devices). All devices are approved for medical use in Europe.

Patients receiving the COVID Box are instructed to measure their temperature, oxygen saturation, respiratory frequency, heart rate, and blood pressure three times a day. Patients are instructed on the use of the devices, the desired frequency of measurements, and their personalized reference values. A physician or physician assistant (supervised by a medical specialist) performs a daily video consultation to monitor the patient's symptoms and vital parameters, whereas physical consultations are only performed when the physician or physician assistant suspects a deterioration of symptoms. Patients are instructed to contact the hospital if readings deviate from the personalized reference values or if they feel unsure or unwell. In case of an emergency, patients are advised to contact the national emergency number, and for technical support, the hospital's telemonitoring support team should be contacted.

In the first phase of implementation, measurements of vital parameters are collected during the video consultation and entered manually into the patient's electronic medical record (EMR). In the second phase, patients install an app, developed by the LUMC, which is linked to the devices via Bluetooth and transfers the measurements directly into the patient's EMR.

The telemonitoring care pathway ends when patients no longer need home monitoring because symptoms have sufficiently improved, or when patients are admitted to the hospital. At such time, the devices are either picked up from the patient's home or handed in at the hospital. Subsequently, the devices are sterilized, tested, and then reused.

Scientific Evaluation

Between March 1 and June 15, 2020, 55 patients were monitored at home using the COVID Box. Preliminary results of the evaluation of the telemonitoring program indicate that no adverse events (ie, deaths or emergency hospital admissions) occurred among patients in the telemonitoring care pathway, and that a worsening of symptoms and need for further medical attention could be effectively detected by means of this telemonitoring care program. Eventually, 5 patients (9%) had to be admitted to the hospital due to progression of symptoms. Both patients and health care providers viewed the use of the COVID Box positively. A more detailed overview of the results of the COVID Box will be presented soon.

A prospective registry, which includes all patients who are home monitored using the COVID Box, has been designed. In this registry, baseline patient characteristics, home-monitoring data, patient-reported outcome measures, and clinical outcomes are included. Data from all patients in this registry are periodically evaluated in order to identify any adverse events in a timely manner and improve the telemonitoring care pathway.

Recommendations for the Implementation of a COVID-19 Telemonitoring Care Pathway

The implementation of digital health initiatives has proven to be challenging [10-12]. In the last decade, several frameworks regarding the implementation of digital health have been developed [13-16]. Although these provide useful information, the onset of COVID-19 and the need for a prompt response pose additional challenges [17]. Based on our initial experiences with developing and implementing a COVID-19 telemonitoring care pathway in a tertiary care context and our extensive experience with the implementation of telemonitoring for other diseases, we provide recommendations on clinical, organizational, scientific, and ethical aspects, which can be translated into daily clinical practice (Table 1). These recommendations may also be used to design and implement telemonitoring for the management of other (sub)acute diseases.

Table 1. Main recommendations to facilitate the implementation of a COVID-19 telemonitoring pathway.

Topic	Recommendation
Clinical and organizational aspects	
Aim and setting	• Define the aim of the telemonitoring program • Determine whether telemonitoring will help to achieve this aim in this specific setting
Clinical assessment	• Perform initial in-person assessment (using adequate protective measures) • Evaluate the appropriateness of telemonitoring for each patient, based on clinical, psychological, and social patient characteristics
Monitoring	• Determine which measurements are to be taken • Generate personalized reference values and frequency of measurements for each patient • Prefer video call over contact by telephone • Rely on close supervision of physician rather than automated decision making • Communicate to patients which actions to take in case of an emergency
Integration in clinical workflow	• Involve all stakeholders in the development and implementation process • Establish solid and concise training for health care personnel • Communicate availability of a telemonitoring pathway within the organization • Provide support to avoid extra workload for health care personnel • Organize a technical helpdesk for patients • Integrate readings into the patient's electronic medical record • Evaluate the program continuously
Resources	• Reuse devices to optimize the use of resources and sustainability of the program • Apply for innovation and research grants • Discuss possibilities for reimbursement depending on your health financing or insurance system
Scientific aspects	
Implementation	• Apply scientific evidence obtained from telemonitoring chronic diseases
Evaluation and scientific research	• Perform scientific evaluation in parallel with the implementation • Obtain consent to use patients' data for scientific research
Legal and ethical aspects	
Privacy and data protection	• Ensure intramural and extramural data security and privacy • Comply with legal frameworks and clinical guidelines
Ensuring optimal technical quality	• Guarantee quality of devices and apps
Consent and informed decisions	• Ensure that the choice to use telemonitoring is a jointly made decision by both the patient and the physician; respect autonomy and patient preferences • Establish responsibilities for each party involved: physician, patient, and telemonitoring team • Offer alternative and opt-out options to the patient
Equal opportunities, no discrimination	• Avoid discrimination, offer equal opportunities, and plan alternative nondigital monitoring pathways

Clinical and Organizational Considerations

Aim and Setting

Implementation of a new health care initiative should contribute to solving a relevant problem. In the case of COVID-19, it should optimize care for patients and relieve pressure on health care systems and professionals. In tertiary care hospitals, most patients have substantial comorbidities and vulnerabilities. As such, these patients require close monitoring and are especially suitable for telemonitoring [5]. Although telemonitoring allows for the management of patients in different settings (eg, assessing a patient's condition and estimating the risk of COVID-19 infection in primary care), it should only be used in situations where it will add value, such as improving clinical effectiveness or satisfaction of care [9].

Patient Selection and Eligibility

Telemonitoring may allow for accessible and remote monitoring of a large number of patients with (suspected) COVID-19. Due to the outbreak of this new disease, our knowledge of the epidemiology, pathophysiology, and clinical course of COVID-19 is still evolving. Hence, an initial in-person assessment of the patient is recommended, especially in the case of a complex medical history.

When evaluating whether telemonitoring provides a suitable option, patients need to be assessed in a holistic manner. Psychological characteristics may influence a patient's ability

to adequately interpret measurements of vital signs and symptoms or ask for help if necessary. Social aspects, such as a patient's home environment (eg, the possibility for home quarantine), or safety and social support (eg, for elderly patients or patients living alone) should be considered as well. Additionally, patients should feel comfortable with "digital" disease management and be capable of handling devices and interacting via video connection. Although in The Netherlands the majority of patients has access to internet and telephone, this should be asked explicitly [18,19]. Finally, patients with low (digital) health literacy require extra attention [20,21]. In contrast to telemonitoring for patients with chronic diseases, (suspected) COVID-19 patients are not accustomed to self-managing their disease and may be less likely to feel in control. Ultimately, both the physician and the patient should decide together whether remote monitoring is appropriate.

Home Monitoring

Given the recentness of COVID-19 and the relative inexperience with telemonitoring for managing this disease, no golden standard for how, what, and when to monitor yet exists. The use of video over telephone consultation provides extra information on general demeanor, skin color, and severity of dyspnea, and thus helps to establish a solid clinical impression [9]. The rationale to monitor respiratory rate, saturation, temperature, heart rate, and blood pressure from home is based on early warning scores developed in hospital settings [22-24]; higher scores measured outside hospital settings were also found to be associated with poorer clinical outcomes [25]. Measuring vital signs three times a day ensures recognition of gradual deterioration. Similar to in-hospital patient management, the measurement frequency and cut-off values for vital signs should be determined for each patient based on clinical judgment. If a physician decides that a patient's parameters need to be assessed more than three times a day, this might be an indication that the patient should be admitted to the hospital to be monitored more closely. Future research should determine whether the use of continuous monitoring with wireless patches and/or automated cut-off scores could also be effective.

Integration With the Clinical Workflow

Irrespective of the setting, the workflow of health care professionals and supportive personnel should be adapted to a new telemonitoring care pathway. A shift from "traditional" to "digitally assisted" clinical practice is necessary. This requires experience, time, and adequate guidelines. First, solid but concise training and frequent evaluation of the process are required. Health care professionals should be informed about the availability of a telemonitoring care pathway and need dedicated time in their work schedules to follow up on their home-monitored patients. Active involvement of all stakeholders, including health care professionals and patients, in the development and implementation of telemonitoring is likely to contribute to a better support base. The extra workload associated with the development and implementation of a telemonitoring care pathway should be minimized, for example, through dedicated support teams and a helpdesk for technical problems. Finally, for patients, as well as health care professionals, it is neither efficient nor safe to log in to different systems multiple times a day. Thus, it is essential that data gathered by patients are integrated into the EMR. In addition, EMR integration facilitates the legal obligation of health care professionals to document all clinical information.

Resources

Telemonitoring could contribute to establishing effective, (cost-)efficient and high-quality health care. Nevertheless, it requires the availability of sufficient devices. This can be difficult, since the COVID-19 pandemic resulted in a shortage of resources. Adequate reuse of devices contributes to the optimization of resources and sustainability of the program. In our experience, patients generally hand in the devices in good working order. Additionally, permanent financing is a recurring problem for digital health solutions [15]. Reimbursement policies for telemonitoring programs vary internationally and should be further established [26]. Moreover, innovation budgets may contribute to obtaining the most appropriate devices and services. To generate scientific evidence, the availability of research grants is warranted.

Scientific Considerations

In general, new forms of care need to be evaluated and scientifically validated before large-scale implementation. However, the current COVID-19 pandemic calls for rapid action and scientific data on telemonitoring for (sub)acute diseases and specifically for COVID-19 is scarce. Although this makes development and implementation of such a pathway difficult, we can rely on prior knowledge obtained from telemonitoring for patients with chronic diseases in order to implement telemonitoring programs that are safe and effective [6-8]. Additionally, this lack of scientific evidence underlines the urgent need for the scientific validation of these new forms of care. Scientific validation may not be top priority for health care providers at this time, but lack of validation may lead to a waste of valuable time, energy, and resources if telemonitoring does not lead to the expected benefits [27,28]. As such, scientific validation of safety and efficacy should form a key element in the development and implementation of telemonitoring, and other new forms of care, during a pandemic.

Designing a validation study for COVID-19–related telemonitoring entails several challenges. Randomized controlled trials are difficult in the early phases of a pandemic, and retrospective cohorts, which may serve as control groups, are not available. Under these conditions, assessment of the causal effect of telemonitoring on clinical endpoints is difficult. Scientific validation of the safety and (cost-)effectiveness of telemonitoring by an observational cohort study might be the best option in the early phase of the COVID-19 pandemic. This type of research can be performed in parallel to the implementation process and provides relevant insights into the clinical course of COVID-19. In the future, scientific validation can contribute to the establishment of clinical guidelines for evidence-based telemonitoring.

Ethical Considerations

Telemonitoring involves several parties: physicians, patients, device manufacturers, and health care institutions. When considering ethical aspects, the different roles and

responsibilities of these parties should be taken into account. Additionally, the COVID-19 pandemic demands a careful balance between rapid implementation and thorough ethical and legal evaluation.

Privacy and Data Protection

Telemonitoring brings health care to patients' homes and therefore requires expanding the secure digital perimeter beyond health care facilities. Patients' data need to remain accessible only for the lawful purposes of processing and exclusively to those with granted access.

The primary goal of data collection during telemonitoring is clinical. Personal health data are included in the patient's EMR to directly provide medical treatment. Those granted access are directly involved in the treatment of the patient and are bound to professional confidentiality. Additionally, in our hospital, secondary uses of personal data generated from COVID-19 patients have been anticipated. Accordingly, assumed consent with an opt-out option has been put in place to reuse patients' data for scientific research, a current public health priority [29]. Data obtained from patients will also be used for the evaluation of the COVID Box. From a legal perspective, activities regarding the use of sensitive information must comply with (inter)national regulations.

Ensuring Optimal Technical Quality

The devices and services used for telemonitoring should be user-friendly, trustworthy, validated, and approved, according to national and/or international regulations. Even if certification of quality requires time, it should not be dismissed. For example, at the time of the implementation of the COVID Box, the app had not yet been approved. In order to avoid delaying the implementation without compromising the quality of health care, the team decided to initially enter the data manually in the EMR.

Consent and Informed Decisions

Physicians must be aware that during a health crisis, patients may be under pressure to agree to telemonitoring. Additionally, obtaining informed consent is more complicated for COVID-19–related studies [30]. Patient and physician should engage in a dialogue about the telemonitoring system, the opt-out options, and the alternatives to telemonitoring. The choice to engage in telemonitoring requires patients to commit to a series of daily actions, and they must understand that they will assume a high level of responsibility. Therefore, it should be ensured that telemonitoring actually works in practice for the patient. Experience with current telemonitoring programs at the LUMC has shown that this type of telemonitoring is well accepted by patients and is clinically viable [6].

A telemonitoring system can be either implemented as a monitoring system in which readings are continuously monitored by the physician, or as a monitoring system in which the physician only monitors the readings once a day. The latter asks for patients to adopt an even higher level of responsibility, as they need to contact the physician themselves in case of deviating measurements. The choice for either monitoring system depends on important factors such as the patient population, the clinical features of the disease, and the local and international medical guidelines and legal frameworks.

Equal Opportunities and No Discrimination

Digital health should never lead to inequality or discrimination [31,32]. In the Netherlands, the possibility that patients lack the necessary technical means for telemonitoring is small [18]. However, this situation is not the same in all countries. Additionally, low literacy regarding (digital) health could be a significant barrier to telemonitoring [33,34]. The availability of an intuitive, readily accessible, and inclusive app, which automatically collects measurements, could be an effective and user-friendly solution to avoid inequality and discrimination. Furthermore, keeping an alternative "analogue" monitoring pathway readily available might contribute to offering equal opportunities to all patients.

Finally, a patient's decision to reject telemonitoring may be based on other personal reasons (eg, lack of privacy at home). Independent of the motives, which may or may not be explained to the physician, the autonomous choice of the patient should be respected.

Conclusion

Innovative digital strategies such as telemonitoring have great potential to improve the management of COVID-19. Telemonitoring may optimize care for patients with COVID-19 by detecting clinical deterioration at an early stage. Additionally, telemonitoring reduces the number of hospital visits and admissions, thereby enabling the efficient use of scarce health care resources and lowering the risk of further transmission of the virus. The direct evidence to support the use of telemonitoring for COVID-19 is still being gathered and analyzed, but preliminary data and previous experiences with other diseases indicate that telemonitoring can be an important tool during the pandemic.

This paper presents a specific telemonitoring care pathway for COVID-19 and offers a set of medical, scientific, organizational, and ethical recommendations. These recommendations may be used as a guide for the design and implementation of telemonitoring for patients with a confirmed or suspected case of COVID-19, and may also be used for the design and implementation of telemonitoring care pathways for other diseases.

Acknowledgments

We thank the COVID Box team at the LUMC, especially CJ Hendriks van de Weem, CMN Marges, FH Soltani, GG Glastra van Loon, GJ de Grooth, K Volz-Brons, L van der Plas, MM van Poele, RE Dam, E Stoop, and S Hendriksen.

Authors' Contributions

DA, NH, HO, RT, AS, AP, and MV were responsible for study conceptualization. AS produced an initial outline. The idea was further established with input from ED, CO, CL, HB, MH, TN, and OT. AS, AP, and MV drafted the paper; it was then revised by all other authors. All authors approved the final version of the article to be published. All authors agree to be accountable for all aspects of the work. They ensure that questions related to the accuracy or integrity of any part of the work will be appropriately investigated and resolved.

Conflicts of Interest

None declared.

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19. World Health Organization. 2020 Mar. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-04-30]
2. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar 28;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
3. Gandhi RT, Lynch JB, Del Rio C. Mild or Moderate Covid-19. N Engl J Med 2020 Apr 24:e. [doi: 10.1056/NEJMcp2009249] [Medline: 32329974]
4. Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J Med Virol 2020 Jun 05;92(6):568-576 [FREE Full text] [doi: 10.1002/jmv.25748] [Medline: 32134116]
5. Clinical management of severe acute respiratory infection when COVID-19 is suspected. World Health Organization. 2020 Mar. URL: https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected [accessed 2020-04-26]
6. Treskes RW, van Winden LAM, van Keulen N, van der Velde ET, Beeres SLMA, Atsma DE, et al. Effect of Smartphone-Enabled Health Monitoring Devices vs Regular Follow-up on Blood Pressure Control Among Patients After Myocardial Infarction: A Randomized Clinical Trial. JAMA Netw Open 2020 Apr 01;3(4):e202165 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.2165] [Medline: 32297946]
7. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
8. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. The Lancet 2018 Sep 22;392(10152):1047-1057 [FREE Full text] [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
9. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
10. Alami H, Gagnon MP, Wootton R, Fortin JP, Zanaboni P. Exploring factors associated with the uneven utilization of telemedicine in Norway: a mixed methods study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):180 [FREE Full text] [doi: 10.1186/s12911-017-0576-4] [Medline: 29282048]
11. Choi WS, Park J, Choi JYB, Yang J. Stakeholders' resistance to telemedicine with focus on physicians: Utilizing the Delphi technique. J Telemed Telecare 2019 Jul 01;25(6):378-385. [doi: 10.1177/1357633X18775853] [Medline: 29792080]
12. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
13. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
14. Mendel P, Meredith LS, Schoenbaum M, Sherbourne CD, Wells KB. Interventions in organizational and community context: a framework for building evidence on dissemination and implementation in health services research. Adm Policy Ment Health 2008 Mar;35(1-2):21-37 [FREE Full text] [doi: 10.1007/s10488-007-0144-9] [Medline: 17990095]
15. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
16. Khoja S, Durrani H, Scott RE, Sajwani A, Piryani U. Conceptual framework for development of comprehensive e-health evaluation tool. Telemed J E Health 2013 Jan;19(1):48-53. [doi: 10.1089/tmj.2012.0073] [Medline: 22957502]

17. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
18. Internet; toegang, gebruik en faciliteiten 2019 (The Internet; access, use and facilities 2019). Centraal Bureau voor de Statistiek (Statistics Netherlands). 2019 Oct. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83429NED/table?fromstatweb%20voor%20NL [accessed 2020-04-17]
19. ICT Statistics. International Communication Union. 2019. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2020-04-24]
20. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
21. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. The Lancet Public Health 2020 May;5(5):e249-e250 [FREE Full text] [doi: 10.1016/S2468-2667(20)30086-4] [Medline: 32302535]
22. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
23. Smith GB, Redfern OC, Pimentel MA, Gerry S, Collins GS, Malycha J, et al. The National Early Warning Score 2 (NEWS2). Clin Med (Lond) 2019 May;19(3):260 [FREE Full text] [doi: 10.7861/clinmedicine.19-3-260] [Medline: 31092526]
24. Patel R, Nugawela MD, Edwards HB, Richards A, Le Roux H, Pullyblank A, et al. Can early warning scores identify deteriorating patients in pre-hospital settings? A systematic review. Resuscitation 2018 Nov;132:101-111 [FREE Full text] [doi: 10.1016/j.resuscitation.2018.08.028] [Medline: 30171976]
25. Scott LJ, Redmond NM, Tavaré A, Little H, Srivastava S, Pullyblank A. Association between National Early Warning Scores in primary care and clinical outcomes: an observational study in UK primary and secondary care. Br J Gen Pract 2020 Jun;70(695):e374-e380 [FREE Full text] [doi: 10.3399/bjgp20X709337] [Medline: 32253189]
26. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
27. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
28. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015 May 06;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
29. Novel Coronavirus (nCoV) acute respiratory infection clinical characterisation data tool. World Health Organization. 2020 Feb. URL: https://media.tghn.org/medialibrary/2020/03/ISARIC_COVID-19_CRF_V1.3_24Feb2020.pdf [accessed 2020-04-24]
30. Guidance on the management of clinical trials during the covid-19 (coronavirus) pandemic (Version 3). European Medicines Agency. 2020 Apr. URL: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf [accessed 2020-04-28]
31. van der Kleij RMJJ, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJF, Teichert M, et al. SERIES: eHealth in primary care. Part 1: Concepts, conditions and challenges. Eur J Gen Pract 2019 Oct;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
32. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
33. Heijmans M, Brabers A, Rademakers J. Health literacy in Nederland. Nivel. 2018 May 3. URL: https://nivel.nl/nl/publicatie/health-literacy-nederland [accessed 2020-04-24]
34. Digital health literacy: how new skills can help improve health, equity and sustainability. EuroHealthNet. 2019. URL: https://eurohealthnet.eu/sites/eurohealthnet.eu/files/publications/PP_Digital%20Health%20Literacy_LR.pdf [accessed 2020-04-24]

Abbreviations

EMR: electronic medical record
LUMC: Leiden University Medical Center

Edited by G Eysenbach, G Fagherazzi; submitted 02.06.20; peer-reviewed by R Tonino, E Metting; comments to author 13.07.20; revised version received 29.07.20; accepted 15.08.20; published 02.09.20.

Please cite as:
Silven AV, Petrus AHJ, Villalobos-Quesada M, Dirikgil E, Oerlemans CR, Landstra CP, Boosman H, van Os HJA, Blanker MH, Treskes RW, Bonten TN, Chavannes NH, Atsma DE, Teng YKO
Telemonitoring for Patients With COVID-19: Recommendations for Design and Implementation
J Med Internet Res 2020;22(9):e20953
URL: https://www.jmir.org/2020/9/e20953
doi:10.2196/20953
PMID:32833660

Viewpoint

Regulatory Sandboxes: A Cure for mHealth Pilotitis?

Abhishek Bhatia[1,2*], MSc; Rahul Matthan[3,4*], LLB; Tarun Khanna[1,5*], PhD; Satchit Balsari[1,2,6*], MD, MPH

[1]The Lakshmi Mittal and Family South Asia Institute, Harvard University, Cambridge, MA, United States

[2]FXB Center for Health and Human Rights, Harvard TH Chan School of Public Health, Boston, MA, United States

[3]The Takshashila Institution, Bengaluru, India

[4]Trilegal, Bengaluru, India

[5]Harvard Business School, Boston, MA, United States

[6]Department of Emergency Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Satchit Balsari, MD, MPH
FXB Center for Health and Human Rights
Harvard TH Chan School of Public Health
651 Huntington Avenue, 703C
Boston, MA,
United States
Phone: 1 6174951000
Email: balsari@hsph.harvard.edu

Abstract

Mobile health (mHealth) and related digital health interventions in the past decade have not always scaled globally as anticipated earlier despite large investments by governments and philanthropic foundations. The implementation of digital health tools has suffered from 2 limitations: (1) the interventions commonly ignore the “law of amplification” that states that technology is most likely to succeed when it seeks to augment and not alter human behavior; and (2) end-user needs and clinical gaps are often poorly understood while designing solutions, contributing to a substantial decrease in usage, referred to as the “law of attrition” in eHealth. The COVID-19 pandemic has addressed the first of the 2 problems—technology solutions, such as telemedicine, that were struggling to find traction are now closely aligned with health-seeking behavior. The second problem (poorly designed solutions) persists, as demonstrated by a plethora of poorly designed epidemic prediction tools and digital contact-tracing apps, which were deployed at scale, around the world, with little validation. The pandemic has accelerated the Indian state’s desire to build the nation’s digital health ecosystem. We call for the inclusion of regulatory sandboxes, as successfully done in the fintech sector, to provide a real-world testing environment for mHealth solutions before deploying them at scale.

(J Med Internet Res 2020;22(9):e21276) doi:10.2196/21276

KEYWORDS

COVID-19; mHealth; digital health; design thinking; regulation; intervention; regulatory sandbox

Introduction

Millions of dollars have been spent on digital health interventions in the global south in the past few years, often with the generous support of influential global philanthropic organizations [1]. Many have not had the anticipated impact at scale despite successful pilot studies [2,3]. The COVID-19 pandemic has amplified and accelerated this mobile health (mHealth) pilotitis with unvalidated and untested social distancing scoreboards, apps, and forecasting models proliferating around the world [4]. As communities and businesses struggle with the uncertainties associated with reopening society, a wide range of technological solutions are being proposed—including syndromic surveillance trackers, electronic passes controlling access to building and transport hubs, and contact-tracing apps and devices [5]. These unvetted tools risk detracting policymakers from focusing on the basics of pandemic response [6].

Over the past decade, the implementation of digital health tools has largely suffered from 2 limitations. First, technocentric interventions commonly ignore what Toyama refers to as the law of amplification in his book *Geek Heresy* [7]. Technology, he postulates, is most likely to succeed when it seeks to augment and not alter human behavior. Secondly, end-user needs are poorly understood in the interest of creating new markets, leading to a significant decrease in usage over time. This attrition

is seldom studied but is important while evaluating digital health implementation, as described by Eysenbach [8].

The minimum viable product—the norm in the start-up industry—is anathema in medicine and public health. Speed need not supplant rigor. In this article, we propose a measured iterative approach to developing, validating, and scaling digital health interventions, using illustrative examples from India.

The Law of Amplification

Before the pandemic, there were over already 350,000 apps in the mHealth category across app stores, though few have had any longevity [9]. Most have relied on misguided assumptions of how humans consume and interact with technology and provided false reassurances to users and policymakers [10]. Those that have been somewhat successful and had impact at scale have demonstrated a commitment to understanding the context and have adopted key design thinking principles, including an iterative and inclusive development process [11,12]. They recognize Toyama's law of amplification [7] that states that technology by itself will not fundamentally change behavior; it will merely amplify it.

The motion picture, radio, television, and finally, massive open online courses (MOOCs) were all expected to fundamentally change how we educated students. They did not [13-15]. While the internet substantially eased the pursuit of knowledge, it had, for the most part, augmented but not entirely replaced classrooms—until the pandemic fundamentally modified human behavior. How we seek education and health has now changed; human mobility is restricted, and because students cannot travel to campuses, we are witnessing an explosive growth in online learning tools as technology responds to amplify human behavior [16].

Telemedicine, which was expected to substantially improve access to general and specialized care, has also not scaled as anticipated. Legal, financial, behavioral, and infrastructural limitations were hard to overcome in both high- and low-income settings [17]. The pandemic has, however, disrupted primary and chronic care globally with particularly devastating consequences in resource-poor settings. In the United States, for example, this public health emergency has accelerated negotiations about regulation and remuneration in telemedicine that had otherwise stalled for years [18]. The 3-month-long nationwide lockdown in India precluded hundreds of thousands of patients from receiving routine home care by community health workers or periodic checks, for example, by their endocrinologists and oncologists [19]. While physicians in the developing world have routinely been using phone-based apps, especially WhatsApp, to communicate with patients [20], the clinical and financial imperatives posed by the pandemic prompted many hospitals, both private and public, to adopt telemedicine solutions expeditiously [21]. The government of India released Telemedicine Practice Guidelines on March 25, 2020, allowing teleconsultation for essential non-COVID services, further embedding telemedicine in routine health care delivery [22].

In contrast, COVID-19 mobility dashboards and contact-tracing apps promoted by governments, technology companies, consulting companies, and academics largely failed to alter the course of the epidemic [23]. Decision rooms of health departments were inundated with presentations of prediction models, despite the unavailability of data required to parametrize these models adequately [24]. Policymakers are weighed down by the low signal-to-noise ratio in the data generated by the health system along with the digital solutions presented to them, compelling the WHO director-general to label the phenomenon an "infodemic" [25,26]. These solutions could not amplify what did not exist—the capacity to test, isolate, or treat patients or to provide for citizens to safely stay home [27,28].

The Tyranny of "Tablets"

The discordance between what is needed and what is supplied by the digital health tool ecosystem is not new. A closer look at the national programs digitizing their health data ecosystem in the countries with low- and middle-income reveals common patterns: overburdened healthcare providers enter large volumes of data into tablet computers or web apps in survey forms riddled with radio buttons, small fonts, and hard-to-navigate dropdown menus that are all anachronistic [29]. These tools reflect regulatory reporting and operational requirements originally designed for an analog world. In India, for example, individual programs must collect demographic data de novo, instead of relying on an application programming interface (API)-enabled import from an interoperable sister program. These data are then transferred with varying compliance and quality to dashboards with scant detail on how the summary statistics are calculated [30]. There is little evidence to show that these efforts result in timely action. Throughout, there is a greater emphasis on programmatic needs; and little attention is paid to the needs of the key users, the patient, and the health care provider—leading to user attrition over time [8,31].

Sandboxing: The Need for Rapid Iteration

Governments currently find themselves facing tough choices between moving swiftly to reassure and protect populations by adopting promising though unvalidated technologies and proceeding cautiously, awaiting robust evidence, while the contagion spreads unabated [32]. We believe it is possible to strike a responsible balance. In 2019, the government of India published its sweeping vision for a digital health ecosystem for India's 1.3 billion residents, the National Digital Health Blueprint (NDHB) [33]. The NDHB called for the provision of a regulatory sandbox—a controlled testing environment within which existing regulations may be temporarily relaxed to allow experimentation [34]. In August 2020, the Prime Minister announced the National Digital Health Mission (NDHM), and the National Health Authority (NHA) released its "Draft Health Data Management Policy" for public comment. Simultaneously, the NHA has launched the NDHM Sandbox, allowing integration and validation of third-party software by partnering with NDHM APIs [35]. We believe that these regulatory sandboxes can be the pill needed to combat the bane of mHealth pilotitis, in India and around the world. A sandbox will provide the enabling conditions administrators need to test new solutions in subsets of populations before mandating change at scale. They offer an opportunity to test what are often minimum viable

digital health products in a responsible, controlled, and monitored real-world environment.

Let us consider, for example, follow-up home visits that are required of India's army of accredited social health activists, or ASHAs (which means "hope" in Hindi). Several population health programs require ASHA workers to go door to door to inquire about the health status of children or expectant mothers or missed follow-up appointments. They then trudge back to the primary care centers to upload these data online, when tablet computers and broadband connectivity are not available in remote regions, and thousands of person-hours are lost weekly in this exercise. The (not entirely unfounded) fear of new technologies failing precludes administrators from experimenting or challenging age-old, resource-intensive, and sometimes ineffective practices. A sandbox would, in this case, provide state administrators with the regulatory flexibility needed to temporarily replace home visits with alternative digital solutions. In the context of the COVID-19 pandemic, there is a behavioral impetus to align the ASHA workers' incentives to use digital solutions with patients' needs. It is unsafe for ASHA workers to go door to door during a pandemic. The time saved by operating from home will allow them to either attend to increased household responsibilities resulting from shelter-in-place orders, spend more time with each patient, or call more patients. Patients benefit from access to care despite quarantining orders. A sandbox approach would allow the ASHA workers (and policy makers) to vet multiple solutions simultaneously, say, by comparing response rates among groups of patients self-isolating at home that received either push-notifications (inexpensive), or automated interactive text messaging (with minimal upfront development costs), or phone calls (resource-intensive). What may be effective in a certain demographic may not necessarily be generalizable, warranting constant feedback loops and the ability to rapidly iterate [36].

Such prototyping must begin upstream, incorporating end-user feedback early in the design process, as was successfully evinced during the development of a digital disease surveillance platform at a mass gathering in Maharashtra, India, in 2015. Preparatory consultation workshops with medical officers seconded to serve in pop-up clinics at the 2015 Kumbh Mela resulted in substantial data minimization and workflow acceleration while improving record-keeping [37].

Agility and Caution

We recognize that regulatory sandboxes are at risk for abuse and will lose public trust if they seek to circumvent the nation's laws on data protection and privacy [38]. The sandboxes could, however, themselves serve as powerful tools to test proposed technological solutions against India's evolving data privacy jurisprudence and societal norms. Over the decades, the Indian state has amassed large volumes of health data; these data can be used to accelerate medical science and positively influence population health [39,40], but they can also result in discrimination and harm [41,42]. The ubiquity of mobile devices and cloud-based services will influence a society's willingness to share their data—as will a better understanding of the harms and benefits of the exchange of health data. The sandboxes provide an opportunity to examine the terms on which society will engage with governments and private players to generate, exchange, and use health data and address fundamental questions about health data that are yet not fully answered across different contexts. For example, are data better utilized for research if they are secured at source and queried remotely with a role-based access, as with the Research Patient Data Registry (RPDR) by Partners Healthcare in Massachusetts [43]; or if datasets are exported for a fee with time and purpose limitations as with research data from the Centers for Medicare and Medicaid Services (CMS) database in the United States [44]; or without a fee, as with the MIMIC-III (Medical Information Mart for Intensive Care III) database, a comprehensive dataset of 12 years of deidentified health data from critical care units of Harvard's Beth Israel Deaconess Medical Center [45]? Which of these approaches are best suited to the range of infrastructure and capability differences within a country?

Policymakers have neither the requisite evidence to evaluate the veracity of the promises made by digital health technologies nor the tools to allow rapid iteration during large-scale adoption [3,46]. We believe that the proposed sandboxes can provide an alternative model. These testbeds provide a means to reject what does not work in the communities served, to optimize interventions by tweaking them in a real-world context, and to generate the evidence required to scale up. Regulated with expertise, caution, and consent, they can serve as incubators for digital health innovation. Prior to making unvetted digital technologies a critical pillar of the COVID-19 response, or of the national public health system at large, a controlled launch in willing communities would provide critical insights on the technological feasibility, epidemiological utility, data protection capability, and society's response to these innovations (Textbox 1).

Textbox 1. Objectives of regulatory sandboxes for mHealth.

- Generate evidence for digital health interventions in the real-world context at reasonable scale
- Allow cautious relaxation of select regulations to allow for innovation
- Measure societal response to the proposed regulatory changes (and to the introduced innovations)
- Eliminate (early) interventions that fail to demonstrate impact or engender trust

Will Sandboxing Work?

The idea of testing policy changes before enforcing them is not new. The economic reforms in China from 1979 to 2012, for example, were a result of several decentralized economic experiments across cities and provinces, where hard evidence and technological expertise were relied upon to inform the rapidly changing national economic policy [47,48]. The

regulatory sandboxes described in the NDHB have more recent precedents in the fintech industry. First introduced in the United Kingdom in December 2015, sandboxes have rapidly spread across the Asia Pacific and used in the insurance, payments, and capital market sectors [34,49]. In November 2019, India's 4 financial service regulators invited fintech companies to test their applications in regulatory sandboxes designed to cater to the banking, insurance, securities, and pensions sectors [47,50]. An mHealth regulatory sandbox could similarly catalyze innovation and infuse rigor in digital health implementation by providing a mechanism for evaluating the scientific validity, health impact, contextual relevance, regulatory compliance, and long-term feasibility of applications, prior to scale-up (Textbox 2).

Textbox 2. Evaluation framework for mHealth tools in regulatory sandboxes.

Validation

- Does the innovation do what it says it does?
- Example claim: "Bluetooth technology will help identify most exposures among persons carrying enabled devices."
- Validating this claim would require widespread testing in a target population to calculate the sensitivity and specificity of the contact-tracing app and to consequently alter detection parameters before a large-scale rollout.

Impact

- Does it result in improved clinical or population health outcomes?
- Example claim: "Human mobility dashboards predicting the impact of social distancing measures will help inform containment strategies."
- Impact would need to be shown by a measurable change in the incidence of cases or by comparing outcomes with control group populations after implementing interventions based on information provided by the dashboard. Statistically significant testing strategies would be a prerequisite [51].

Relevance

- Are the improved outcomes a priority for the community?
- Example claim: "Digitizing vertical service delivery programs improves care quality."
- Improved record keeping after digitization may not have an impact on population health if chronic inventory shortages and absenteeism are not addressed.

Compliance

- Do regulatory relaxations result in harm, now or later?
- Example claim: "It is necessary to collect location data to monitor individual movement during the pandemic."
- In compliance with data protection laws, data should be aggregated or anonymized to the lowest possible resolution that allows epidemiologically sound interventions without risking individual or group reidentification or harm.

Feasibility

- Is it easy to use? Does it integrate with existing workflows?
- Example claim: "Digitization of discharge summary notes at the bedside expedites claims of reimbursements."
- Feasible interventions should integrate into existing systems without introducing bottlenecks or resulting in flow reorganization with stakeholder buy-in. Given the volume of patients seen in the outpatient departments at public hospitals (several hundreds), this requirement, for example, may be debilitating in certain contexts. Providers should have the option of completing discharge notes within a reasonable timeframe.

Sustainability

- Can it scale to meet increased demand?
- Example claim: "A successful locally developed, locally deployed electronic medical record system is ready for national expansion."
- Sustainable scale-up proposals would require assessing the flexibility of design features to accommodate for the heterogeneity in workflow, digital literacy, and ability to report needs across jurisdictions as well as the adequacy of decision support at scale. Early findings may signal the need for subsequent RCTs.

Conclusion

The COVID-19 pandemic addresses 1 of the 2 key barriers to successful mHealth adoption—the discordance between the proposed solution and underlying human behavior is reduced as technological solutions are now better aligned with needs. We believe regulatory sandboxes address the other—the need for validation. As the National Health Authority and the Ministry of Health and Family Welfare launch the NDHM in August 2020, incorporating sandboxes is opportune and important [52]. Before prescribing solutions at a national scale, the sandboxes must examine what is permissible under existing laws, where

existing laws fall short, and where additional enabling regulations are needed to advance clinical or population health.

The pandemic warrants urgent intervention, not *any* intervention. In heterogenous populations with large socioeconomic, cultural, and behavioral diversity, this approach is far more effective than ideological adherence. Medicine is after all the original home of randomized controlled trials—now the darling of economists [53]. The profession has, for decades, strived to make the practice evidence-based. It is time to apply the same vigilance to digital health implementation, even during the pandemic. Regulatory sandboxes may present an effective mechanism to do so.

Acknowledgments

SB received support from the Burke Fellowship at the Harvard Global Health Institute. This paper is part of a series of publications on India's evolving digital health ecosystem by the India Digital Health Net at Harvard University.

Conflicts of Interest

None declared.

References

1. Using Data to Save Lives: The Rockefeller Foundation and Partners Launch $100 Million Precision Public Health Initiative. The Rockefeller Foundation. New York; 2019 Sep 25. URL: https://www.rockefellerfoundation.org/news/using-data-save-lives-rockefeller-foundation-partners-launch-100-million-precision-public-health-initiative/ [accessed 2020-06-09]
2. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
3. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
4. Editorial. Show evidence that apps for COVID-19 contact-tracing are secure and effective. Nature 2020 Apr;580(7805):563. [doi: 10.1038/d41586-020-01264-1] [Medline: 32350479]
5. Sarbadhikari S, Sarbadhikari SN. The global experience of digital health interventions in COVID-19 management. Indian J Public Health 2020 Jun;64(Supplement):S117-S124 [FREE Full text] [doi: 10.4103/ijph.IJPH_457_20] [Medline: 32496240]
6. Baheti A. How India's lack of transparency on coronavirus data is hindering the battle against pandemic. Scroll.in. 2020 May 17. URL: https://scroll.in/article/962138/how-indias-lack-of-transparency-on-coronavirus-data-is-hindering-the-battle-against-pandemic [accessed 2020-06-09]
7. Toyama K. Geek Heresy: Rescuing Social Change from the Cult of Technology. New York: PublicAffairs; May 26, 2015.
8. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
9. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. Berlin, Germany: Research2Guidance; 2017 Nov 01. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2020-06-09]
10. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
11. Modi D, Dholakia N, Gopalan R, Venkatraman S, Dave K, Shah S, et al. mHealth intervention "ImTeCHO" to improve delivery of maternal, neonatal, and child care services-A cluster-randomized trial in tribal areas of Gujarat, India. PLoS Med 2019 Oct;16(10):e1002939 [FREE Full text] [doi: 10.1371/journal.pmed.1002939] [Medline: 31647821]
12. Scholten H, Granic I. Use of the Principles of Design Thinking to Address Limitations of Digital Mental Health Interventions for Youth: Viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
13. Ossiannilsson E, Altinay F, Altinay Z. MOOCs as Change Agents to Boost Innovation in Higher Education Learning Arenas. Education Sciences 2016 Aug 06;6(4):25. [doi: 10.3390/educsci6030025]
14. Sra P, Chakraborty P. Opinion of Computer Science Instructors and Students on MOOCs in an Indian University. Journal of Educational Technology Systems 2018 Aug 28;47(2):205-212. [doi: 10.1177/0047239518797085]
15. Dynarski S, Lalani F. Online Courses Are Harming the Students Who Need the Most Help. The New York Times. 2018 Jan 19. URL: https://www.nytimes.com/2018/01/19/business/online-courses-are-harming-the-students-who-need-the-most-help.html [accessed 2020-06-09]
16. Li C. The COVID-19 pandemic has changed education forever. This is how. World Economic Forum. 2020 Apr 29. URL: https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/ [accessed 2020-06-09]
17. Combi C, Pozzani G, Pozzi G. Telemedicine for Developing Countries. A Survey and Some Design Issues. Appl Clin Inform 2016 Nov 02;7(4):1025-1050 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0089] [Medline: 27803948]

18. Medicare Telemedicine Health Care Provider Fact Sheet. Centers for Medicare & Medicaid Services. 2020 Mar 17. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2020-06-09]
19. Cash R, Patel V. Has COVID-19 subverted global health? Lancet 2020 May 30;395(10238):1687-1688 [FREE Full text] [doi: 10.1016/S0140-6736(20)31089-8] [Medline: 32539939]
20. Giordano V, Koch H, Godoy-Santos A, Dias Belangero W, Esteves Santos Pires R, Labronici P. WhatsApp Messenger as an Adjunctive Tool for Telemedicine: An Overview. Interact J Med Res 2017 Jul 21;6(2):e11 [FREE Full text] [doi: 10.2196/ijmr.6214] [Medline: 28733273]
21. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
22. Sarbadhikari S, Nautiyal S. Disruptive Digital Health - Hastened by the Covid-19 Pandemic Internet. RGNUL Student Research Review (RSRR) Blog Series. 2020 Jun 11. URL: http://rsrr.in/2020/06/11/digital-healthcare-covid-19-technology/ [accessed 2020-08-01]
23. Zastrow M. Coronavirus contact-tracing apps: can they slow the spread of COVID-19? Nature 2020 May 19. [doi: 10.1038/d41586-020-01514-2] [Medline: 32433633]
24. Holmdahl I, Buckee C. Wrong but Useful - What Covid-19 Epidemiologic Models Can and Cannot Tell Us. N Engl J Med 2020 Jul 23;383(4):303-305. [doi: 10.1056/NEJMp2016822] [Medline: 32412711]
25. Munich Security Conference. World Health Organization. 2020 Feb 15. URL: https://www.who.int/dg/speeches/detail/munich-security-conference [accessed 2020-06-09]
26. Balsari S, Buckee C C, Khanna T. Which Covid-19 Data Can You Trust? Harvard Business Review. 2020 May 08. URL: https://hbr.org/2020/05/which-covid-19-data-can-you-trust [accessed 2020-06-09]
27. McDonald S. Technology Theatre. Centre for International Governance Innovation. 2020 Jul 13. URL: https://www.cigionline.org/articles/technology-theatre [accessed 2020-08-01]
28. mHealth: New Horizons for Health through Mobile Technologies. Geneva: World Health Organization; Nov 01, 2011.
29. Rajan JV, Moura J, Gourley G, Kiso K, Sizilio A, Cortez AM, et al. Understanding the barriers to successful adoption and use of a mobile health information system in a community health center in São Paulo, Brazil: a cohort study. BMC Med Inform Decis Mak 2016 Nov 17;16(1):146 [FREE Full text] [doi: 10.1186/s12911-016-0385-1] [Medline: 27855685]
30. Ghazisaeidi M, Safdari R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of Performance Dashboards in Healthcare Sector: Key Practical Issues. Acta Inform Med 2015 Oct;23(5):317-321 [FREE Full text] [doi: 10.5455/aim.2015.23.317-321] [Medline: 26635442]
31. Balsari S. Will AI help universalize health care? The BMJ Opinion. 2019 Sep 23. URL: https://blogs.bmj.com/bmj/2019/09/23/satchit-balsari-will-ai-help-universalize-health-care/ [accessed 2020-06-09]
32. Zagury-Orly I, Schwartzstein RM. Covid-19 - A Reminder to Reason. N Engl J Med 2020 Jul 16;383(3):e12. [doi: 10.1056/NEJMp2009405] [Medline: 32343505]
33. Final report on National Digital Health Blueprint (NDHB). Delhi, India: Ministry of Health and Family Welfare, Govt. of India; 2019 Nov 08. URL: https://main.mohfw.gov.in/sites/default/files/Final%20NDHB%20report_0.pdf [accessed 2020-06-09]
34. Matthan R. Regulatory sandbox. Livemint. 2017 Jan 25. URL: https://www.livemint.com/Opinion/jsvqNgipJgYizPaqhYtqEL/Regulatory-sandbox.html [accessed 2020-06-09]
35. Ministry of Health and Family Welfare, Government of India. NDHM Sandbox Environment. URL: https://sandbox.ndhm.gov.in/ [accessed 2020-09-05]
36. Moola S, John O, Bhaumik S. Encouraging health workers to use mHealth for delivering primary healthcare services: policy brief. The George Institute of Global Health. India; 2020 Jun 01. URL: https://www.georgeinstitute.org/encouraging-health-workers-to-use-mhealth-for-delivering-primary-healthcare-services-policy-brief [accessed 2020-06-09]
37. Paliwal A. Tracking India's Deadly Flu Outbreak in Real Time. NOVA Next.: PBS Online; 2015 Oct 28. URL: https://www.pbs.org/wgbh/nova/article/kumbh-mela-flu/ [accessed 2020-06-09]
38. Responses to the Government of India's Joint Parliamentary Committee's Consultation on The Personal Data Protection Bill. The Lakshmi Mittal and Family South Asia Institute at Harvard University. 2020 Feb 01. URL: http://mittalsouthasiainstitute.harvard.edu/wp-content/uploads/2020/03/IDHN-Comments-on-PDP-Bill.pdf [accessed 2020-06-09]
39. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
40. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2019;2:111 [FREE Full text] [doi: 10.1038/s41746-019-0189-7] [Medline: 31754637]
41. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
42. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
43. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a Research Patient Data Repository. AMIA Annu Symp Proc 2006:1044 [FREE Full text] [Medline: 17238663]

44. Herland M, Khoshgoftaar TM, Bauder RA. Big Data fraud detection using multiple medicare data sources. J Big Data 2018 Sep 4;5(1):29. [doi: 10.1186/s40537-018-0138-3]
45. Johnson AE, Stone DJ, Celi LA, Pollard TJ. The MIMIC Code Repository: enabling reproducibility in critical care research. J Am Med Inform Assoc 2018 Jan 01;25(1):32-39 [FREE Full text] [doi: 10.1093/jamia/ocx084] [Medline: 29036464]
46. Agarwal S, Lefevre AE, Labrique AB. A Call to Digital Health Practitioners: New Guidelines Can Help Improve the Quality of Digital Health Evidence. JMIR Mhealth Uhealth 2017 Oct 06;5(10):e136 [FREE Full text] [doi: 10.2196/mhealth.6640] [Medline: 28986340]
47. Khanna T. Learning From Economic Experiments in China and India. AMP 2009 May;23(2):36-43. [doi: 10.5465/amp.2009.39985539]
48. Morrison W. China's Economic Rise: History, Trends, Challenges, and Implications for the United States. Congressional Research Service. 2014 Aug 21. URL: https://digital.library.unt.edu/ark:/67531/metadc770519/ [accessed 2020-06-09]
49. Chen CC. Regulatory Sandboxes in the UK and Singapore: A Preliminary Survey. Rochester, NY: Social Science Research Network; 2019 Sep 14. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3448901 [accessed 2020-06-09]
50. PTI. RBI allows fintech companies, financial institutions to set up regulatory sandbox. The Economic Times. 2019 Aug 13. URL: https://economictimes.indiatimes.com/news/economy/finance/rbi-allows-fintech-companies-financial-institutions-to-set-up-regulatory-sandbox/articleshow/70664351.cms [accessed 2020-06-09]
51. Prabhakaran D, Jha D, Prieto-Merino D, Roy A, Singh K, Ajay VS, et al. Effectiveness of an mHealth-Based Electronic Decision Support System for Integrated Management of Chronic Conditions in Primary Care: The mWellcare Cluster-Randomized Controlled Trial. Circulation 2018 Nov 10;139(3):380-391. [doi: 10.1161/CIRCULATIONAHA.118.038192] [Medline: 30586732]
52. Sharma A. We are 100 per cent technology-driven by nature: Dr Indu Bhushan, CEO, Ayushman Bharat. Express Healthcare. 2020 Mar 30. URL: https://www.expresshealthcare.in/public-health/we-are-100-per-cent-technology-driven-by-nature-dr-indu-bhushan-ceo-ayushman-bharat/417831/ [accessed 2020-06-09]
53. McKenzie D. A Nobel Prize for Development RCTs!. World Bank Blogs: Development Impact. 2019 Oct 14. URL: https://blogs.worldbank.org/impactevaluations/nobel-prize-development-rcts [accessed 2020-06-09]

Abbreviations

ASHA: accredited social health activist
API: application programming interface
CMS: Centers for Medicare and Medicaid Services
mHealth: mobile health
MIMIC-III: Medical Information Mart for Intensive Care III
MOOC: massive open online course
NDHB: National Digital Health Blueprint
NDHM: National Digital Health Mission
NHA: National Health Authority
RPDR: Research Patient Data Registry

Edited by G Eysenbach; submitted 09.06.20; peer-reviewed by S Sarbadhikari, F Lami; comments to author 29.06.20; revised version received 23.07.20; accepted 02.08.20; published 15.09.20.

Please cite as:
Bhatia A, Matthan R, Khanna T, Balsari S
Regulatory Sandboxes: A Cure for mHealth Pilotitis?
J Med Internet Res 2020;22(9):e21276
URL: http://www.jmir.org/2020/9/e21276/
doi:10.2196/21276
PMID:32763889

Viewpoint

Can Disinfection Robots Reduce the Risk of Transmission of SARS-CoV-2 in Health Care and Educational Settings?

Kathrin Cresswell[1], BSc, MSc, PhD; Aziz Sheikh[1], OBE, FRSE, FMedSci

Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Kathrin Cresswell, BSc, MSc, PhD
Usher Institute
The University of Edinburgh
Teviot Place
Edinburgh
United Kingdom
Phone: 44 131 650 8102
Email: kathrin.cresswell@ed.ac.uk

Abstract

We explore the opportunities and challenges surrounding the use of disinfection robots to reduce the risk of SARS-CoV-2 transmission in health care and educational settings. Although there is some potential for deploying robots to help with manual cleaning, the evidence base is mixed, and we highlight that there needs to be work to establish and enhance the effectiveness of these robots in inactivating the virus.

(J Med Internet Res 2020;22(9):e20896) doi:10.2196/20896

KEYWORDS

robotics; disinfection; SARS-CoV-2; COVID-19; risk; transmission; virus

SARS-CoV-2 can be transmitted through droplets and contact with contaminated surfaces [1]. To contain the spread, there is a need for more regular and deeper cleaning of indoor surfaces, for example, in schools, care homes, and health care facilities. There is also a need to reduce human exposure to potentially contaminated surfaces. As a result, there is now a greater interest in cleaning and disinfection robots in these settings [2-4]. Such robots are, for example, currently routinely cleaning the Hong Kong metro, and the Smart Field Hospital in Wuhan uses them in an attempt to reduce the spread of SARS-CoV-2 [5,6].

Existing disinfection robots work through a combination of automated or semiautomated processes. They can clean or disinfect floors and surfaces but increasingly focus on disinfecting whole rooms with increasingly complex distribution systems. These most commonly include machines using UV-C light, which works by altering DNA and RNA so that organisms cannot replicate, and vapor and fogging systems that spray chemical disinfectants.

However, despite their increasing use and demand across settings, evidence of their effectiveness is mixed. There is no existing work exploring the effectiveness of disinfection robots in relation to SARS-CoV-2 and other viruses, and the evidence of the impact of UV-C and vapor on health care–associated infections is also limited. In health care settings, both UV-C light and chemical-based disinfection methods (most commonly hydrogen peroxide vapor) do not demonstrate any significant impact on reduced infection rates, although some studies have identified some positive trends and demonstrated a reduction in surface contamination [7-9]. Not surprisingly, UV-C light and chemicals need to touch a surface to be effective, and this may not always be the case—they have issues with shadows, may not reach all areas of concave surfaces, and their effectiveness reduces with distance [10,11]. This work is further complicated by a lack of evidence around how much contamination actually leads to infection and adverse patient outcomes, but there appears to be a general agreement that both techniques are most effective when combined with manual cleaning [12].

Studies investigating cleaning robots using these techniques are limited. The few existing investigations have found that cleaning robots using UV-C light and hydrogen peroxide can deliver some benefits in reducing microbial surface contamination but only when combined with manual cleaning [13,14]. The study quality is relatively low for both applications with possible commercial biases.

Deploying the current generation of cleaning and disinfection robots in health care settings, care homes, and schools is, therefore, unlikely to be of major benefit, and there needs be

work to establish and enhance the effectiveness of these robots in inactivating SARS-CoV-2. In addition to concerns around effectiveness, these devices are expensive at between US $30,000 and US $135,000 per unit, and organizations need to train staff to deploy and control them [13,15-17]. Disinfectant chemicals and UV-C light can also be dangerous to human health, so people typically need to leave while the robot cleans the room. This is particularly concerning for communal settings but does not preclude the use of UV-C light in enclosed empty spaces. Other factors to consider include disinfection time (some devices take a few hours per room) and issues with physical spaces and navigation (robots are not good at climbing stairs) [14,18].

Floor cleaning robots are likely to be cheaper units that can relatively easily and quickly be adapted (eg, from other types of service robots) and that can focus on one aspect of the physical environment (ie, the floor) while humans work in parallel with them, eliminating issues around disinfection time. There is, therefore, a need to catalyze the development of floor cleaning robots that can regularly clean communal settings, particularly those with a high risk of transmitting nosocomial infections. These devices can augment manual cleaning, for instance, through supporting an already stretched workforce and through reducing the risk of exposure for cleaning staff and those who work in these settings (eg, doctors, nurses, assistants, teachers), particularly in the context of shortages of personal protective equipment. Some have noted issues with the compliance of cleaning protocols promoting use of these robots, and others have highlighted the importance of effective integration with existing routines and operations [19,20], but this is unlikely to be a significant hurdle in a times of global need. If the current generation of cleaning and disinfection robots are viewed as a panacea to reduce the spread of SARS-CoV-2, the resulting overreliance on their performance may jeopardize lives unnecessarily, but this is an area for urgent development that could help with lockdown exit strategies.

Acknowledgments

This work was funded by a Scottish Government Chief Scientist Research Grant. The views expressed are those of the authors.

Conflicts of Interest

None declared.

References

1. van Doremalen N, Bushmaker T, Morris D, Holbrook M, Gamble A, Williamson B, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. N Engl J Med 2020 Apr 16;382(16):1564-1567 [FREE Full text] [doi: 10.1056/NEJMc2004973] [Medline: 32182409]
2. Curcio K. Robots: the future of cleaning? Services Magazine. URL: https://servicesmag.org/online-digital-magazine/digital-archives/item/291-robots-the-future-of-cleaning [accessed 2020-04-15]
3. Cleaning and disinfection robots market growth is accelerating with a CAGR of 159%, during 2019-2025. MarketWatch. URL: https://www.marketwatch.com/press-release/cleaning-and-disinfection-robots-market-growth-is-accelerating-with-a-cagr-of-159-during-2019-2025-2019-09-09?mod=mw_quote_news [accessed 2020-04-15]
4. Azevedo MA. Hospital-disinfecting robots: Xenex sees surge in orders as COVID-19 pandemic escalates. Crunchbase News. 2020 Mar 25. URL: https://news.crunchbase.com/news/hospital-disinfecting-robots-xenex-sees-surge-in-orders-as-covid-19-pandemic-escalates/ [accessed 2020-04-15]
5. Metro operator deploys cleaning robot to fight the coronavirus. Metro Report International. 2020 Mar 11. URL: https://www.railwaygazette.com/technology-data-and-business/metro-operator-deploys-cleaning-robot-to-fight-the-coronavirus/55995.article [accessed 2020-04-15]
6. Ackerman E. Autonomous robots are helping kill coronavirus in hospitals. IEEE Spectrum. 2020 Mar 11. URL: https://spectrum.ieee.org/automaton/robotics/medical-robots/autonomous-robots-are-helping-kill-coronavirus-in-hospitals [accessed 2020-04-15]
7. Anderson DJ, Chen LF, Weber DJ, Moehring RW, Lewis SS, Triplett PF, et al. Enhanced terminal room disinfection and acquisition and infection caused by multidrug-resistant organisms and Clostridium difficile (the Benefits of Enhanced Terminal Room Disinfection study): a cluster-randomised, multicentre, crossover study. Lancet 2017 Feb;389(10071):805-814. [doi: 10.1016/s0140-6736(16)31588-4]
8. Passaretti C, Otter J, Reich N, Myers J, Shepard J, Ross T, et al. An evaluation of environmental decontamination with hydrogen peroxide vapor for reducing the risk of patient acquisition of multidrug-resistant organisms. Clin Infect Dis 2013 Jan;56(1):27-35. [doi: 10.1093/cid/cis839] [Medline: 23042972]
9. Tran K, McCormack S. Non-Manual Ultraviolet Light Disinfection for Hospital Acquired Infections: A Review of Clinical Effectiveness and Guidelines. 2019 Apr 05. URL: https://www.ncbi.nlm.nih.gov/books/NBK544669/ [accessed 2020-04-15]
10. Yang J, Wu U, Tai H, Sheng W. Effectiveness of an ultraviolet-C disinfection system for reduction of healthcare-associated pathogens. J Microbiol Immunol Infect 2019 Jun;52(3):487-493 [FREE Full text] [doi: 10.1016/j.jmii.2017.08.017] [Medline: 28951015]

11. Mana TS, Sitzlar B, Cadnum JL, Jencson AL, Koganti S, Donskey CJ. Evaluation of an automated room decontamination device using aerosolized peracetic acid. Am J Infect Control 2017 Mar 01;45(3):327-329. [doi: 10.1016/j.ajic.2016.10.006] [Medline: 27866755]
12. Miller R, Simmons S, Dale C, Stachowiak J, Stibich M. Utilization and impact of a pulsed-xenon ultraviolet room disinfection system and multidisciplinary care team on Clostridium difficile in a long-term acute care facility. Am J Infect Control 2015 Dec 01;43(12):1350-1353 [FREE Full text] [doi: 10.1016/j.ajic.2015.07.029] [Medline: 26362699]
13. Doll M, Stevens M, Bearman G. Environmental cleaning and disinfection of patient areas. Int J Infect Dis 2018 Feb;67:52-57 [FREE Full text] [doi: 10.1016/j.ijid.2017.10.014] [Medline: 29102556]
14. Doll M, Morgan DJ, Anderson D, Bearman G. Touchless technologies for decontamination in the hospital: a review of hydrogen peroxide and UV devices. Curr Infect Dis Rep 2015 Sep;17(9):498. [doi: 10.1007/s11908-015-0498-1] [Medline: 26252970]
15. UVC disinfection robot. Hipac. URL: https://www.hipac.com.au/shop/uvc-disinfection-robot/index [accessed 2020-04-15]
16. Murray A. Coronavirus: robots use light beams to zap hospital viruses. BBC News. 2020 Mar 20. URL: https://www.bbc.co.uk/news/business-51914722 [accessed 2020-04-15]
17. Randall I. Chinese company unveils $40,000 hospital robot that uses a combination of UV light and liquid disinfectant spray to kill coronavirus pathogens. Daily Mail. 2020 Mar 27. URL: https://www.dailymail.co.uk/sciencetech/article-8159921/Chinese-company-unveils-40-000-hospital-robot-disinfectants-against-coronavirus-pathogens.html [accessed 2020-04-15]
18. Hard surface disinfection system from Surfacide. Bates IT. URL: https://www.batesit.co.uk/helios-disinfection-system [accessed 2020-04-15]
19. Fleming M, Major Y, Gryskevicz M, Fife J, Hassmer L, Masroor N, et al. Will audit and feedback drive compliance with UV robot disinfection? Am J Infect Control 2017 Jun;45(6):S51. [doi: 10.1016/j.ajic.2017.04.083]
20. Stibich M. Reduction of healthcare associated infections through the use of pulsed xenon ultraviolet disinfection. SANIMED. 2016 May 13. URL: https://www.sanimed.ro/upload_poze_documente/files/STUDII%20CLINICE.pdf [accessed 2020-04-15]

Edited by G Eysenbach, G Fagherazzi; submitted 01.06.20; peer-reviewed by JA Sánchez Margallo, N Kim; comments to author 09.08.20; revised version received 10.08.20; accepted 03.09.20; published 15.09.20.

Please cite as:
Cresswell K, Sheikh A
Can Disinfection Robots Reduce the Risk of Transmission of SARS-CoV-2 in Health Care and Educational Settings?
J Med Internet Res 2020;22(9):e20896
URL: http://www.jmir.org/2020/9/e20896/
doi:10.2196/20896
PMID:32903196

Viewpoint

Ethical Challenges and Opportunities Associated With the Ability to Perform Medical Screening From Interactions With Search Engines: Viewpoint

Elad Yom-Tov[1,2], PhD; Yuval Cherlow[3], Rabbi

[1]Microsoft Research, Herzeliya, Israel

[2]Faculty of Industrial Engineering and Management, Technion, Haifa, Israel

[3]The Tzohar Rabbinical Organization, Lod, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
Alan Turing 3
Herzeliya, 4672415
Israel
Phone: 972 747111359
Email: eladyt@microsoft.com

Abstract

Recent research has shown the efficacy of screening for serious medical conditions from data collected while people interact with online services. In particular, queries to search engines and the interactions with them were shown to be advantageous for screening a range of conditions including diabetes, several forms of cancer, eating disorders, and depression. These screening abilities offer unique advantages in that they can serve a broad strata of the society, including people in underserved populations and in countries with poor access to medical services. However, these advantages need to be balanced against the potential harm to privacy, autonomy, and nonmaleficence, which are recognized as the cornerstones of ethical medical care. Here, we discuss these opportunities and challenges, both when collecting data to develop online screening services and when deploying them. We offer several solutions that balance the advantages of these services with the ethical challenges they pose.

(J Med Internet Res 2020;22(9):e21922) doi:10.2196/21922

KEYWORDS

search engines; diagnosis; screening

Introduction

Recent work has demonstrated the ability to screen for serious medical conditions using search engine logs [1-5]. The development and deployment of these abilities can open new opportunities for earlier diagnosis and more equitable care but require careful consideration of the associated ethical challenges. The goal of this paper is to discuss the ethical pros and cons of these capabilities and to set the stage for a broader discussion of these issues.

Search engines are used by the vast majority of internet users to obtain information on a variety of topics, including medicine [6]. Search engine operators collect information on the interaction of users with their services to improve the operation of their search engines, for example, by measuring user satisfaction from specific answers given to them [7]. It is important to stress that the data collected by search engine operators are not collected to improve medical research or improve people's health but to enhance search engine operation.

The data collected by search engine operators include, for example, query text, links shown to the user, time of clicking on the links, duration of reading each link, and mouse movements, which serve as a proxy for eye gaze tracking [8]. The data collected by search engines are usually anonymous, in the sense that specific individuals cannot be easily linked to their data, but, unless specified by the user, multiple searches can be attributed, with high likelihood, to the same user.

As noted above, these data have been shown to be effective for screening people for a variety of medical conditions, such as diabetes [4], several forms of cancer [1-3], eating disorders, and depression [5]. Interactions with search engines are useful for such screening because of a combination of factors, including people's limited knowledge of the association between symptoms and conditions [9]; the fact that many conditions (eg,

ovarian cancer) have benign symptoms, of which only the confluence indicates disease, but psychological biases lead people to focus on only the latest symptom [2]; and people's natural tendency to defer treatment but ask about it online.

These screening capabilities offer unique advantages in that they can serve a broad strata of the society, including people in underserved populations and in countries with poor access to medical services [7]. However, these advantages come at a possible cost to privacy, autonomy, and nonmaleficence, which are recognized as the cornerstones of ethical medical care [10]. Note that the legal aspects of providing (and of not providing) these capabilities are not discussed in this work.

We note that other services, including content providers (eg, Wikipedia and patient groups [11]) and social media platforms (eg, Facebook [12] and Twitter [13]) collect similar data. However, for the reasons described above, we focus on search engines. Moreover, for a broader discussion on the ethics of internet research, readers can refer to the article by Buchanan and Zimmer [14].

In our opinion, the ethical questions that arise from the ability to screen search engine logs should be divided into questions that appear during the development of screening capabilities and questions that should be resolved before medical interventions are provided to people as part of the use of a resulting product. Here, we discuss both these areas.

Development

Incidental Finding

"Incidental finding" [15] refers to the case where, during research on one medical topic, data indicates that a person under study has another medical condition of which he/she is (possibly) unaware. Consider a person who contributes their genetic information to build a new screening test for a specific hereditary condition. Upon examination, researchers realize that this person's genetic information reveals that he/she has another, perhaps common, mutation, which indicates that the person has a serious medical condition that he/she may not know about.

The commonly accepted solution to this challenge in genetic research is to screen for mutations that are common, life-saving, and do not require the person to have a deep understanding of genetics in order to decide whether he/she would like to be treated. If such a mutation is identified, the researcher informs the person that he/she should consult with a genetic counsellor but does not provide advice, as this is not the researcher's specialty. This route is taken also because, if the burden of treatment (or advice) is placed on the researcher, medical research will, in practice, be severely restricted. This is also the reason that, in many cases, ethics committees recommend opting for completely anonymous research, which reduces the ethical burden on the researcher.

We claim that there is similarity between the question of incidental finding in the medical domain and the case where researchers use data collected during people's interactions with search engines to later determine that a user may have a medical condition. This can arise from a simple interaction, such as a query suggesting suicidal ideation, to a more elaborate insight obtained from a predictive model based on interactions with the internet service. However, the analogy is not perfect. People who donate their data for medical research know that their data will be examined for medical purposes, whereas people who use a search engine do not expect their data to be used for medical research. In fact, in many cases, people who use search engines may not realize that their interaction data are being collected. We note in passing that routine experiments, such as Facebook's modification of the order of postings by friends, caused an uproar when they were described in an academic paper [16].

Nevertheless, we argue that these differences should not prevent us from using the insight medical ethics has garnered on the question of incidental findings because people who contribute their data to medical research may not realize that other findings are possible, and on the other hand, as public awareness of search engine data grows, people will realize that these data can provide them with benefits.

Informed Consent and Autonomy

Search engines, as other internet services, have a system of consent that often includes the use of data collected by the search engine for research purposes. People who use a search engine implicitly consent to its use and further can click on the link at the bottom of the search page where its "Terms of Use" are specified. However, it is difficult to refer to this as informed consent in the medical sense. For example, the authors found that in a sample of approximately 116 million users, only 0.05% clicked on the Terms of Use page during a 1-month period. Experience from other web services suggests that even if a pop-up window would require people to consent to their data being used for research, most people would click on the window without considering what they are consenting to [17].

Additionally, Terms of Use are necessarily broad in their description and, we assume, are often broader than informed consent forms signed by people participating in medical research.

Thus, it is still a challenge to develop a form of consent that both satisfies the ethical requirements for data use and does not overburden users in their interactions with the search engine.

Willingness to Provide Search Data for Medical Research

There is often an implicit assumption that people would not want their data, collected for other purposes, to be used for medical research without their specific consent. Gefen et al [18] tried to quantify the value that people assign to their data and found that, in a sample of people from around the world, 99% were willing to provide their search engine data in exchange for monetary compensation lower than US $1500 and 53% were willing to pay to have their data analyzed, even if the value of the analysis would be to the society at large rather than to them directly.

Thus, while a minority of users would not agree to the use of their data regardless of compensation, many would agree, and a relevant portion of the population even sees value in the

availability of these services, which exceeds that of the data itself.

Anonymity

As noted above, most of the search data used for medical research has, to date, been anonymous as far as researchers are concerned. This anonymization is provided through the provision of a random user identifier and by not including information that could easily compromise anonymity (eg, location). However, as shown in the AOL leak [19], a malicious researcher may be able to identify a small number of users when such anonymization is used. Therefore, it may be necessary to assume that data are not fully anonymized to a malicious researcher and perhaps even sometimes to a benevolent researcher.

Companies collecting data may, on the other hand, be able to identify a user. This can happen, for example, if users register to their services with their real name. Thus, even if data are anonymous to a researcher, it could, conceivably, be deidentified by the organization collecting the data. In such a case, the problem of incidental findings can arise, as described above.

Finally, an advantage to having data linked to an individual (either anonymous or identifiable) is the ability of users to control the use of their data, as offered, for example, in the European Union's recent General Data Protection Regulation (GDPR).

Representativeness

The question of representation in internet data appears in several forms. First, there are the questions of who uses the services from which data are collected and whether they faithfully characterize the entire population. Second, not all people use the internet in similar ways to acquire information, which causes another form of bias in the data.

The first source of representation bias could greatly impact populations, especially in financially disadvantaged parts of the society and in countries with lower access to the internet. Although many efforts have been devoted to closing this gap, it still exists. For example, the percentage of people with access to the internet in different countries ranges from almost 0% to 100% [20]. Thus, it is important to account for such representation biases when using the data to build a model that can be useful to all people.

The second source of representation bias is less well known but is no less important. As shown in past work [21], the use of search engines for medical queries, for example, is highly dependent on people's age and gender. Moreover, only around 16% of people use search engines to query for medical information [9], adding to the representation bias.

Summary

Taking the above-mentioned points into consideration, we suggest that in the case of research on medical conditions using search engine data collected for operational purposes, it may be preferable to use anonymized data rather than to obtain consent for the use of identifiable information. If the former route is taken, it is important that ethical committees approve the research, serving in their capacity as representatives of society. We recognize that this is an imperfect solution because of both the inability to seek informed consent and the difficulty for ethics committees to represent search engine users who come from a range of countries and societies, each with its own norms and expectations. However, we view this as a balance between the competing challenges outlined above.

Production

Approaches for Providing Search-Based Screening Information

Once a screening method is developed, it may be put into regular use. This could be done in several ways, which are described below.

Suppose an anonymous search engine user is predicted to have a medical condition (eg, screened positive according to interactions with the search engine for the medical condition). The first and most intrusive way to provide the user with this information is to display a notice at a prominent location on the screen. This is currently done only to people who search for information on how to kill themselves [22] or for related topics. In such cases, a banner notice is displayed with the telephone number of a local helpline.

Another way that could be used is to bias (modify) search results toward suggesting the suspected condition. For example, if a user searches for "constant thirst," instead of showing the regular set of results, users who are predicted to have diabetes will be shown more results that suggest diabetes. A similar "personalized search" is currently part of the service of all major search engine providers (eg, when results are served such that they are relevant to the user's current location). Therefore, such a solution might not be perceived as a major change by users.

The third way we envision to display information is through the use of search advertisements [3]. Advertisements are not part of the main search results ("organic results") and are assessed differently by users [23]. People who search for diagnostic information ("do I have diabetes?") will be shown advertisements that would suggest help in diagnosis ("Worried you have diabetes? Click here to obtain more information"). People who click on the advertisements will be given diagnostic assistance, for example, in the form of clinical questionnaires. As shown recently [3], it is possible to train advertising systems to focus on people who are the most at risk.

We note that advertising in the health domain is currently limited by the policy of advertising systems. On one hand, this prevents abuse of the system by purveyors of unapproved medical services, but also means that any use of this method will often require approval by advertising system managers.

A fourth method to inform people of a possible medical concern is through the normal use of a search engine after first obtaining informed consent to provide these insights. If this method is adopted, users will be shown an informed consent form whenever they are identified as new users by the search engine. The form will offer the users to receive screening information but will default to not receiving the information unless the users

positively indicate their interest in receiving this information. Users who consent will then be given alerts whenever a possible medical condition is predicted, based on their queries and behaviors.

Finally, a system might be built where users register and agree to provide their search data on a continuous basis in exchange for alerts when a medical concern is identified in these searches. The data collection, storage methods, and data use would be clearly described to the user. This is similar to services that analyze people's genetic material, where their data will be the search data (or browsing data, in general) and the analysis will be conducted on an ongoing basis, rather than a single transfer of data. Such a system could be offered by search engine providers or, perhaps preferably, by medical providers or dedicated companies.

Unsolicited Diagnosis

Unsolicited diagnosis [24] or unsolicited medical opinion [25] refers to the case where people may be provided with medical information when they do not expect it. For example [26], consider the case of a dermatologist who is standing at the back of an elevator at the mall and notices that the person standing in front has a mole that the dermatologist thinks is likely cancerous. In this case, the person who has the mole is not expecting to receive a diagnosis from a random person at the mall (though a specialist in this case), and thus, this is a case of unsolicited diagnosis. Medical ethicists have considered the question of whether the medical specialist has a duty to inform the person to seek medical attention and whether the specialist has a right to do so. On one hand, the person is not expecting a diagnosis and there is no doctor-patient relationship between the two. On the other hand, not informing the person may lead to serious and irreparable damage. The conclusion reached by some ethicists [26] is that medical doctors have a duty to offer their unsolicited medical opinion, especially when the medical condition requires urgent attention for treatment. However, doctors need to consider the possible harm of such an intervention. Note that a legal duty to act is very much country-specific, often defined through legislation (ie, "good Samaritan" laws [27]) to protect people who take such action.

We note in passing that the balance between benefit and harm for the individual may differ from that for the society. For example, some conditions currently have no treatment because they cannot be identified early enough and so many people would prefer not to know that they have such conditions. However, suppose search engine data could provide such an early alert [28]. In such a case, if enough people knew they have these conditions, pharmaceutical companies might be compelled to develop treatments. However, as this is a secondary effect, we have not focused on it.

Risk Compensation

Risk compensation (also referred to as moral hazard [29]) describes increased risk taking caused by the perceived usefulness of safety measures. For example, it has been suggested that condom distribution fosters inhibition among HIV-positive people [30].

If internet platforms disclose offering screening services, users may choose to modify their behaviors in ways that could harm them. For example, as noted above, only around 16% of users queried about medical symptoms prior to diagnosis [9]. It is difficult to predict illness for people who do not query, but they may assume that a screening model is examining their queries and will alert them when it is necessary to visit a health provider, thus preferring not to access medical care even when they think they should. This is especially likely in the fifth solution described above, because users who register with a dedicated service expect it to provide such alerts. Therefore, it may be important for such a service to alert users about its inability to screen when it predicts that they will not ask relevant questions.

Cost of Errors

No system is perfect, including those discussed in this paper. The cost of errors is an important factor in whether and how information should be provided to users. A false positive error means that a person is informed (depending on the method of provision described above) of a medical condition when he/she does not have one. This can cause undue stress and result in unnecessary medical procedures [31]. A false negative error means that a person who should have been provided with an alert does not receive one, possibly causing late diagnosis (as described above).

Summary: Advantages and Disadvantages of Different Notification Methods

The first method described above, whereby a notice is shown to the user, is advantageous in that it provides people with immediate, clear, and actionable information. However, we advocate its use in only the most extreme situations (eg, expressed intention of suicide) because it is intrusive and may cause more harm than good in the form of breaching privacy and impinging on people's autonomy.

The second (biasing results), third (advertisements), and fourth (explicit prior informed consent) methods are advantageous in that they do not force information upon users and allow users to decide if they would like to use the offered information. These methods (especially biasing results and advertisements), however, somewhat impinge on privacy and autonomy. Additionally, not all users will recognize the help offered to them, and only some will make use of it even when they recognize it. We note that in the case of advertising, the act of choosing to click on the advertisement should be considered informed consent (assuming that the advertisement is appropriately phrased). We also recognize that obtaining explicit informed consent (eg, the fourth method) can be difficult from a design perspective and can burden users who, for example, use private browsing. For these reasons, we believe that these methods, especially the one using advertisements, correctly balance benefit and harm.

Finally, the fifth method (dedicated system) is clearly superior in terms of autonomy and consent, but based on prior experience, we assume that only a minority of users, probably skewed toward the more affluent parts of society, will use this method. Thus, while extremely beneficial for individuals, it should be considered less useful at the societal level.

Authors' Contributions

EYT conceived this paper. YC and EYT discussed the questions raised herein, suggested relevant past work, and wrote the paper.

Conflicts of Interest

YC declares no conflicts of interest. EYT is an employee of Microsoft, owner of the Bing search engine. The positions described in the paper are those of the authors and not necessarily those of their respective employers.

References

1. White RW, Horvitz E. Evaluation of the Feasibility of Screening Patients for Early Signs of Lung Carcinoma in Web Search Logs. JAMA Oncol 2017 Mar 01;3(3):398-401. [doi: 10.1001/jamaoncol.2016.4911] [Medline: 27832243]
2. Soldaini L, Yom-Tov E. Inferring Individual Attributes from Search Engine Queries and Auxiliary Information. In: WWW '17: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: 26th International Conference on World Wide Web; April 2017; Perth, Australia. [doi: 10.1145/3038912.3052629]
3. Yom-Tov E. Screening for Cancer Using a Learning Internet Advertising System. ACM Trans. Comput. Healthcare 2020 Apr 15;1(2):1-13. [doi: 10.1145/3373720]
4. Hochberg I, Daoud D, Shehadeh N, Yom-Tov E. Can internet search engine queries be used to diagnose diabetes? Analysis of archival search data. Acta Diabetol 2019 Oct;56(10):1149-1154. [doi: 10.1007/s00592-019-01350-5] [Medline: 31093762]
5. Schueller SM, Steakley-Freeman DM, Mohr DC, Yom-Tov E. Understanding perceived barriers to treatment from web browsing behavior. J Affect Disord 2020 Apr 15;267:63-66. [doi: 10.1016/j.jad.2020.01.131] [Medline: 32063574]
6. Fox S, Duggan M. Health Online 2013. Pew Research Center. 2013. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-03-01]
7. Youngho K, Hassan A, White RW, Zitouni I. Modeling dwell time to predict click-level satisfaction. In: WSDM '14: Proceedings of the 7th ACM international conference on Web search and data mining. 2014 Presented at: 7th ACM international conference on Web search and data mining; February 2014; New York, NY, USA. [doi: 10.1145/2556195.2556220]
8. Youngmann B, Yom-Tov E. Anxiety and Information Seeking: Evidence From Large-Scale Mouse Tracking. In: Proceedings of the World Wide Web Conference on World Wide Web (WWW 2018). 2018 Presented at: World Wide Web Conference on World Wide Web (WWW 2018); 2018; Lyon, France. [doi: 10.1145/3178876.3186156]
9. Hochberg I, Allon R, Yom-Tov E. Assessment of the Frequency of Online Searches for Symptoms Before Diagnosis: Analysis of Archival Data. J Med Internet Res 2020 Mar 06;22(3):e15065 [FREE Full text] [doi: 10.2196/15065] [Medline: 32141835]
10. Gillon R. Medical ethics: four principles plus attention to scope. BMJ 1994 Jul 16;309(6948):184-188 [FREE Full text] [doi: 10.1136/bmj.309.6948.184] [Medline: 8044100]
11. Yom-Tov E. Crowdsourced Health: How What You Do on the Internet Will Improve Medicine. Cambridge, Massachusetts, United States: MIT Press; 2016.
12. De Choundhury M, Counts S, Horvitz E, Hoff A. Characterizing and predicting postpartum depression from shared facebook data. In: CSCW '14: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 2014 Presented at: 17th ACM conference on Computer supported cooperative work & social computing; February 2014; Vancouver, BC, Canada p. 626-639. [doi: 10.1145/2531602.2531675]
13. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
14. Buchanan EA, Zimmer M. Internet Research Ethics. 2012. URL: https://plato.stanford.edu/entries/ethics-internet-research/ [accessed 2020-05-19]
15. Christenhusz GM, Devriendt K, Dierickx K. To tell or not to tell? A systematic review of ethical reflections on incidental findings arising in genetics contexts. Eur J Hum Genet 2013 Mar;21(3):248-255 [FREE Full text] [doi: 10.1038/ejhg.2012.130] [Medline: 22739341]
16. Jouhki J, Lauk E, Penttinen M, Sormanen N, Uskali T. Facebook's Emotional Contagion Experiment as a Challenge to Research Ethics. MaC 2016 Oct 10;4(4):75-85. [doi: 10.17645/mac.v4i4.579]
17. Acquisti A, Adjerid I, Balebako R, Brandimarte L, Cranor LF, Komanduri S, et al. Nudges for Privacy and Security. ACM Comput. Surv 2017 Oct 09;50(3):1-41. [doi: 10.1145/3054926]
18. Gefen G, Ben-Porat O, Tennenholtz M, Yom-Tov E. Privacy, Altruism, and Experience: Estimating the Perceived Value of Internet Data for Medical Uses. In: WWW '20: Companion Proceedings of the Web Conference 2020. 2020 Presented at: Web Conference 2020; April 2020; Taipei, Taiwan p. 552-556. [doi: 10.1145/3366424.3383414]
19. Hong Y, Vaidya J, Lu H, Karras P, Goel S. Collaborative Search Log Sanitization: Toward Differential Privacy and Boosted Utility. IEEE Trans. Dependable and Secure Comput 2015 Sep 1;12(5):504-518. [doi: 10.1109/tdsc.2014.2369034]

20. Individuals using the Internet (% of population). The World Bank. 2018. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS [accessed 2020-05-19]
21. Yom-Tov E. Demographic differences in search engine use with implications for cohort selection. Inf Retrieval J 2019 Jan 1;22(6):570-580. [doi: 10.1007/s10791-018-09349-2]
22. Cheng Q, Yom-Tov E. Do Search Engine Helpline Notices Aid in Preventing Suicide? Analysis of Archival Data. J Med Internet Res 2019 Mar 26;21(3):e12235 [FREE Full text] [doi: 10.2196/12235] [Medline: 30912753]
23. Danescu-Niculescu-Mizil C, Broder A, Gabrilovich E, Josifovski V, Pang B. Competing for users' attention: on the interplay between organic and sponsored search results. In: WWW '10: Proceedings of the 19th International Conference on World Wide Web. 2010 Presented at: 19th International Conference on World Wide Web; 2010; Raleigh, North Carolina p. 291-300. [doi: 10.1145/1772690.1772721]
24. Mitchell EW. The ethics of unsolicited diagnosis of mental disorder in acquaintances: benefits and dangers. Psychiatrist 2018 Jan 02;35(8):297-301. [doi: 10.1192/pb.bp.110.032953]
25. Ratzan RM. Unsolicited medical opinion. J Med Philos 1985 May;10(2):147-162. [doi: 10.1093/jmp/10.2.147] [Medline: 4009080]
26. Preller G, Salloch S. Melanoma in the shopping mall: A utilitarian argument for offering unsolicited medical opinions in informal settings. Bioethics 2018 Mar;32(3):193-198. [doi: 10.1111/bioe.12426] [Medline: 29369379]
27. Pardun JT. Good Samaritan Laws: A Global Perspective. Loyola of Los Angeles International and Comparative Law Journal 1998;20(3):591-613.
28. Youngmann B, Allerhand L, Paltiel O, Yom-Tov E, Arkadir D. A machine learning algorithm successfully screens for Parkinson's in web users. Ann Clin Transl Neurol 2019 Dec;6(12):2503-2509 [FREE Full text] [doi: 10.1002/acn3.50945] [Medline: 31714022]
29. Reynolds J. A critical examination of the climate engineering moral hazard and risk compensation concern. The Anthropocene Review 2014 Oct 08;2(2):174-191. [doi: 10.1177/2053019614554304]
30. Shelton JD. Ten myths and one truth about generalised HIV epidemics. The Lancet 2007 Dec;370(9602):1809-1811. [doi: 10.1016/s0140-6736(07)61755-3]
31. Hatch S. Snowball in a Blizzard: A Physician's Notes on Uncertainty in Medicine. New York, NY, USA: Basic Books; 2016.

Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by C Danforth, F Sanz; comments to author 04.08.20; revised version received 09.08.20; accepted 11.08.20; published 16.09.20.

Please cite as:
Yom-Tov E, Cherlow Y
Ethical Challenges and Opportunities Associated With the Ability to Perform Medical Screening From Interactions With Search Engines: Viewpoint
J Med Internet Res 2020;22(9):e21922
URL: http://www.jmir.org/2020/9/e21922/
doi:10.2196/21922
PMID:32936082

Viewpoint

Adoption of Blockchain in Health Care

Mark Gaynor[1], PhD; Janet Tuttle-Newhall[2], MD; Jessica Parker[1], MHA; Arti Patel[1], MHA; Clare Tang[1], MHA

[1]School of Public Health and Social Justice, Saint Louis University, St Louis, MO, United States
[2]Brody School of Medicine at East Carolina University, Greenville, NC, United States

Corresponding Author:
Mark Gaynor, PhD
School of Public Health and Social Justice
Saint Louis University
3545 Lafayette Ave
St Louis, MO, 63104
United States
Phone: 1 314 977 8304
Email: mark.gaynor@slu.edu

Abstract

This study aims to review current issues regarding the application of blockchain technology in health care. We illustrated the various ways in which blockchain can solve current health care issues in three main arenas: data exchange, contracts, and supply chain management. This paper presents several current and projected uses of blockchain technology in the health care industry. We predicted which of these applications are likely to be adopted quickly and provided a supply chain example of tracking the transportation of organs for transplantation.

(J Med Internet Res 2020;22(9):e17423) doi:10.2196/17423

KEYWORDS

blockchain adoption; blockchain technology in health care; supply chain; data management

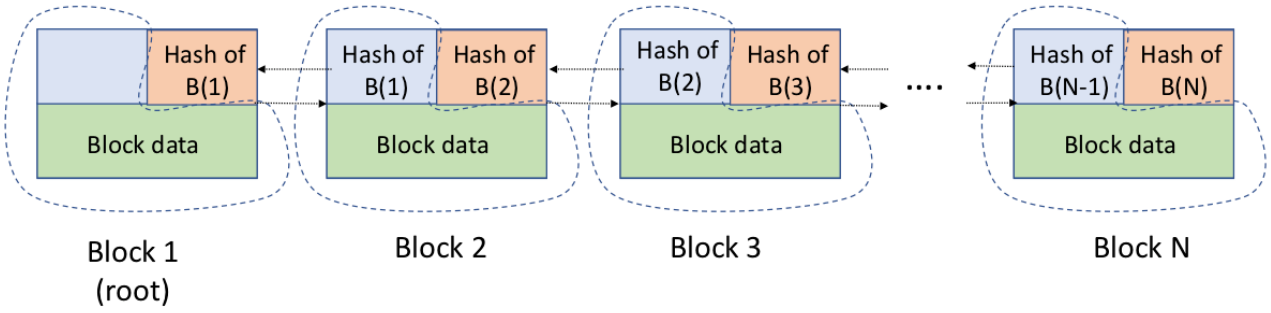
An Introduction to Blockchain Technology

Blockchain technology is a distributed database that records and stores transaction records. Specifically, blockchain is a record of peer-to-peer transactions built from linked transaction blocks that are immutable and shared within a network.[1]. A distribution ledger is "a type of database that is shared, replicated, and synchronized among the members of a network. The distribution ledger records the transactions, such as the exchange of assets or data, among the participants in the network" [2]. Distribution ledgers can be classified as public or private. A public distribution ledger is anonymous in the sense that each user has a copy of the ledger and participates in confirming transactions independently, whereas a private distribution ledger is not anonymous. A permissioned blockchain requires that individuals be given a copy of the ledger and permission from an organization that oversees the ledger to participate in confirming transactions. This technology allows organizations to manage privacy and Health Insurance Portability and Accountability Act concerns as most will be private and require permission.

Blockchain technology, which is currently revolutionizing industries globally, is well suited for identity management, transaction processing, record management, and public health surveillance. This disruptive technology is creating innovative solutions for complex issues in a variety of industries including fishing, diamond, fashion, shipping, banking, and now health care, which are discussed in Multimedia Appendix 1 [3-11]. This paper discusses how this technology may solve some of the largest, most complex, and convoluted problems in the health care industry today. It also introduces a framework to assess the adoption of blockchain technology in health care proposed by Iansiti [12]. This framework is used to justify an example of blockchain application that is likely to be adopted quickly to a transportation chain of custody system for organs intended for transplantation.

Figure 1 illustrates a simple blockchain with N blocks. Each block has data, a hash of the previous block, and a hash of the data in this block. The dashed lines define the region where the block hash covers. Each block is linked to the previous block (except the root block) and every block in the secure chain. If any data in a block are changed, then the hash for that and later blocks will be incorrect. The distributed nature of blockchain means that everybody has a copy of the chain and must also make the same changes to keep the entire blockchain consistent, which is highly unlikely (ie, high Byzantine fault tolerance). This is a model of distributed trust where one or more *bad actors* do not compromise the integrity of the database.

Figure 1. Simple blockchain.



Interactions among participants occur across the network to store, exchange, and view information. The ledger will permanently record data in a sequential chain where *confirmed and validated transaction blocks are linked and chained from the beginning of the chain to the most current block* [2]. Once a transaction is added to the block, it cannot be altered because each block must be verified from all users with access to the ledger, which ensures the integrity of the information being shared. As the blockchain network automatically conducts a self-check, it results in greater transparency and reduced corruptibility.

Blockchain technology does not use any form of centralized authority. The records are public and easily verifiable. In health care, these blocks will not be completely public, but those who are given permission will be granted access to verify whether the information is correct before it becomes part of the blockchain. An audit trail follows every transaction for authentication purposes, and each record has a timestamp and a unique cryptographic signature [2]. The cryptographic signature, also known as a digital signature, allows a user to sign with a private key to track the origination of the transaction information. Blockchain technology uses encryption for security control and authentication via a public and private key. A user's public key is their address on the blockchain, whereas a user's private key is similar to a password that provides access to data. The combination of public and private keys ensures that the data stored on the blockchain are incorruptible and traceable to the origin source while keeping that source anonymous.

A cryptographic hash serves as the digital signature for a block of data. Hashing, along with the use of public and private keys, proves that information in the transaction has not been altered. The goal of blockchain technology is to serve as a digital ledger that will eliminate the intermediaries with the use of cryptographic hashing and timestamps to establish digital trust among users and to allow direct and efficient transfer of data and information [13].

Growing Interest in Blockchain and the Health Care Industry

On the basis of the successful use of blockchain in other industries, as described in Multimedia Appendix 1 [3-11], we examine and predict how health care could benefit from this technology and the speed of adoption of different applications in 3 arenas: data exchange, contracts, and supply chain management. Each arena is depicted below with an illustrated tree and description of the components making up each potential areas for improvement. These applications are derived from a set of relevant papers commissioned by the Office of the National Coordinator [1,2,14-18], other related research [19,20], and the authors' experiences in the health care industry.

Overview of Health Care Problems

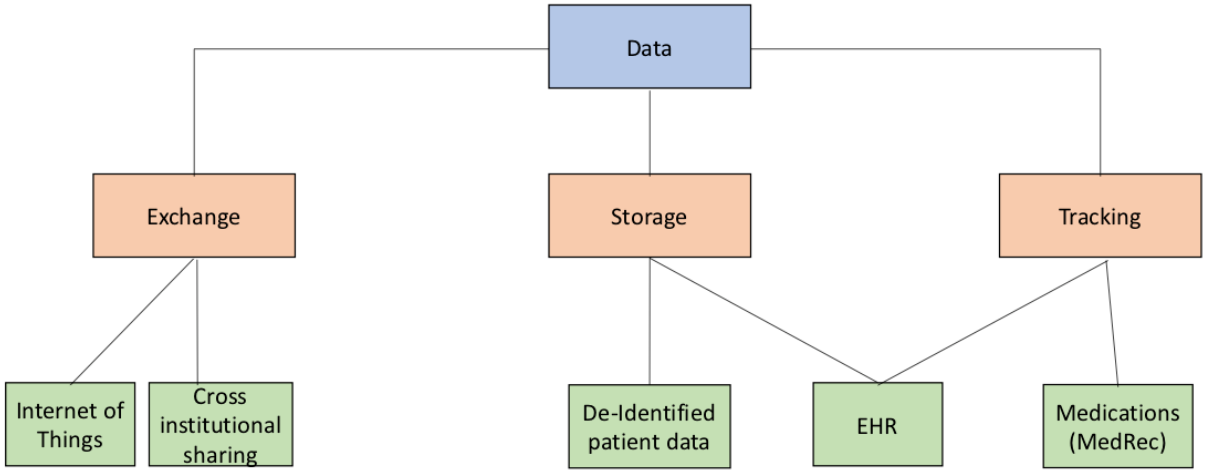
Health care is a complex industry with various influential stakeholders. Blockchain-based technology has the capacity to disrupt the logistics of the health care industry through innovative solutions to the challenges faced in the industry, which include but are not limited to the following: (1) supply chain management, (2) technical issues in data management, (3) smart contracts, (4) confidentiality of personal health information (PHI), (5) enabling and/or assisting implementation and assessment of alternative payment models, and (6) virus outbreak tracking and surveillance.

Data Exchange

Management of the large amount of data collected by health systems presents a technological challenge for health care systems, payors, regulatory agencies, government overseers, and professionals. Challenges in data management specifically include data structure, security, data standardization, data storage and transfers, governance and ownership of the data, inaccuracies, and real-time applicable analytics [21]. The two-part verification system utilized in blockchain technology creates potential solutions to these issues. The private or public key system in blockchain allows for trust to be established to efficiently transfer data in a manner that is traceable and secure.

Figure 2 graphically depicts some opportunities provided by blockchain for exchanging health care data. These applications could allow the health care industry to improve data exchange across all industry areas, including exchange, storage, and record tracking.

Figure 2. Data exchange tree. EHR: electronic health record.



For long-term success, blockchain technology application must supplement the electronic health records (EHRs) that health care systems and providers are currently using. Certainly, for multiple reasons (ie, cost, compliance, and meaningful use requirements to name a few), blockchain technology cannot replace EHRs. Supply chain management for organ transport is a good example of an application where early adoption is likely because the application of this technology is well understood (not novel) and simple to implement (not complex) as it is independent of the EHR function. Blockchain technologies have already been developed on platforms such as Chronicled [22], which tracks a range of products including drugs, blood, and organs.

Likely applications to health care data management and storage include the following:

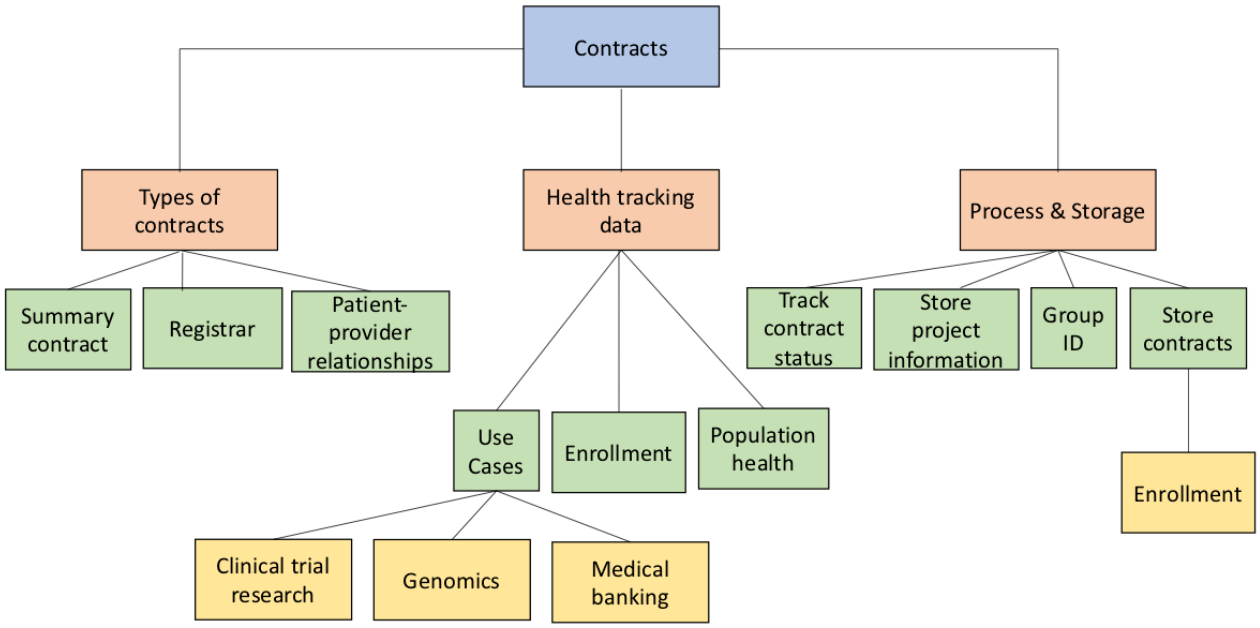
1. The potential to manage data created by a patient encounter with the Internet of Things (IoT) for cross-institutional data sharing. A blockchain infrastructure for the IoT could manage health data from wearable devices [15]. Blockchain could also be used for cross-institutional sharing of health care data via health information exchange networks, application programming interfaces, and the use of standards. Cross-institutional sharing will allow significant increases in clinical and research efficiency through access to data from multiple health care institutions and continuity of care across multiple health care venues based on provider access to consistent patient information.
2. Fast Healthcare Interoperability Resources (FHIR) and the smart security infrastructure ensure data privacy and interoperability with EHRs. These features allow individual institutions to retain operational control of their data and ensures that sensitive patient data are not shared on the blockchain for security and confidentiality measures [17]. Most importantly, FHIR allows for full collaboration among institutions and public health organizations for better care coordination, outcome-based care, population health as well as enabling other diverse data to be shared for research purposes.
3. Blockchain ensures privacy, security, and trust in real-time, distributed data structures. It also ensures provenance, data verification, and data accuracy. For example, linked hashes and public or private key cryptography will help ensure data integrity and tracking of ownership, which can be used to store parts of the EHR (or a pointer to it) on the blockchain to be accessed across multiple health care systems. A public blockchain could be used as an access control manager to health records that are stored off the blockchain. This will allow access to data through a secure user's unique identifier, encrypted link to the health record, and timestamp of each transaction. Blockchain could serve as an immutable audit log where data queried on the blockchain are tracked to ensure that data are only accessed by authorized personnel [16].
4. Blockchain is also attempting to address the confidentiality of PHI through a secure and encrypted data exchange network [2]. Blockchain could enable patients to manage and control their own health information to stay educated and aware of their health care needs. One example is to allow patients to hold custody of their health care data and control where and how the data can be used to share health data in a timely manner for better patient care [21,23]. Strong encryption and an immutable transaction ledger will allow patients to use web-based or mobile apps to view and grant or revoke access to specific parties [23].
5. Finally, an MIT lab has developed a MedRec prototype that uses blockchain technology for data management and record keeping. MedRec manages authentication, confidentiality, accountability, and sharing of sensitive data while facilitating interoperability and making health information technology (IT) convenient and adaptable. Additionally, MedRec is integrated with the organization's existing local data storage solutions to pull aggregated data across networks. "The MedRec prototype provides a proof-of-concept system, demonstrating how principles of decentralization and Blockchain architecture could contribute to secure, interoperable EHR systems" [15]. MedRec can also be used to track medications used by patients and serve as a repository for medication storage and management.
6. Now more than ever, blockchain technology can help monitor the spread of disease, patient data management, insurance, provider directories, and supply chain of food sources. As the Coronavirus disease 2019 pandemic

continues to spread rapidly across the country, this technology can be used to trace cases and transmission patterns. Recently, the US Department of Homeland Security published guidelines for using blockchain to accurately track and trace the movement of goods in the supply chain [24]. IBM Food Trust is currently using distributed ledger technology (DLT) to monitor the movement of food throughout the United States to assess the food supply during this pandemic.

Contracts

Figure 3 illustrates the opportunities where blockchain could be employed to improve contracts in the health care environment. Smart contracts allow users to automate and track certain state transitions. The party receiving new information receives an automated notification and can verify the proposed record before accepting or rejecting the data. This keeps participants informed and engaged in the evolution of the record. Overall, contracts are broken down into 3 branches: types of contracts available to use, health tracking data, and process and storage.

Figure 3. Smart contracts tree.



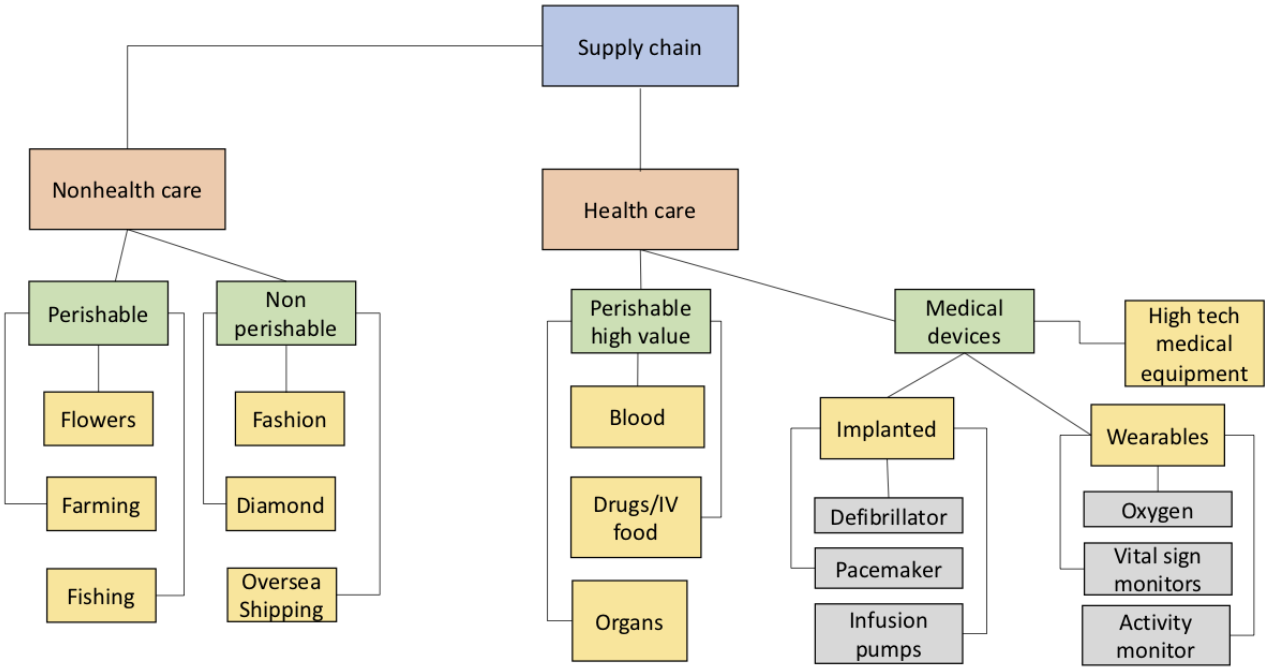
1. The first area of focus is the different types of smart contracts that allow users to automate and track state transitions. The party receiving new information receives an automated notification and can verify the proposed record before accepting or rejecting the data while keeping the participants informed. The 3 types of smart contracts are summary contracts, registrar contracts, and patient-provider relationship contracts [15]. A summary contract functions as a bread crumb trail for participants in the system to locate their medical record history. Providers have references to patients they serve and third parties with whom their patients have authorized data sharing [15]. Smart contracts allow participating parties to build a trusting relationship through increased transparency and tracking with each transition and record evolution.
2. Smart contracts could advance methods used for tracking health data. As previously mentioned, cross-institutional data exchange can allow data to be gathered on patients on a microscopic through to the macroscopic level when comparing patients with similar conditions and/or demographics [17]. Smart contracts could be used to track enrollment for health plans giving providers and patients a better understanding of benefit utilization [14]. In terms of population health, data gathered nationally could be accessed by researchers and institutions to study epidemiology when data (or pointers to these data) are stored in the blockchain, making the data interoperable, secure, and easily accessible and traceable. Smart contracts allow participating parties to understand the utilization of health data while providing secure and auditable tracking of access.
3. One significant benefit of a smart contract is its ability to provide process and storage for contracts, project management, and health data. Contracts and other vital information are stored on the blockchain, establishing a network of trust among the parties involved. This allows health organizations to share patient data that are deidentifiable, ensuring confidentiality. The decentralized database solution provided by blockchain is a key resolution to interoperability and record storage while maintaining health data and other health-related information [2].

Supply Chain

In health care, issues arise with transparency and IT and tracking costs when following items along the supply chain [21]. Blockchain has the power to solve these issues by creating more transparent, instantaneous tracking of high-value items through its shared digital ledger. The final tree in Figure 4 illustrates the opportunities for the use of blockchain in the medical supply

chain. These areas include high-value items, medical devices, and durable medical equipment.

Figure 4. Supply chain tree. IV: intravenous.



1. Supply chain managers responsible for high-value items would benefit from the implementation of a blockchain supply chain ledger. High-value items that hospitals must track include organs for transplantation, blood products, expensive medications, operating room equipment and medical implants such as heart valves, and prosthetic blood vessels or hardware. Furthermore, some of these items, such as organs, blood products, and medications, may have stringent transport requirements such as transport time, temperature for transport, or regulatory transport compliance policies, which can be tracked in a distributed blockchain log. Blockchain's traceability and transparency through the digital ledger can create a more efficient and effective way to manage high-value items along the supply chain. The distributed database will allow each party to come into contact with the high-value item to verify its location, compliance with transport requirements, and document handoffs. This will create transparency of transactions and the irreversible records entered in a distributed supply chain database [12]. This enables high-value items to be tracked in real time, improving inventory management, minimizing courier costs, identifying issues faster along the supply chain, reducing errors, and improving the use of limited resources such as operating room time [19]. Implementing blockchain technologies in the supply chain management of organ transportation should increase the rates of patients receiving allocated organs in a timely manner, and will provide a ledger that cannot be changed. By managing organ transportation from the donor hospital to the recipient with a distributed infrastructure, time sensitivity issues are monitored such that each transport team knows exactly when an organ is being removed, how long transportation will take, and how long after retrieval the organ will arrive at the recipient facility and placed into the transplant recipient. With electronic tracking, we can also monitor which parts of the organ transport system failed or lacked efficiency and work on improving the process. By using blockchain to track the organ in its journey from the donor to recipient, it can improve efficiency throughout the entirety of the process, ensure compliance with regulatory policies, and allow for innovation in organ transport, such as the use of drones for organ delivery.
2. One example of the potential use of blockchain in the health care industry is DonorNet, a secure internet-based system in which organ procurement coordinators (OPCs) send out offers of donated organs to transplant centers with compatible candidates. This system allows organ procurement organizations (OPOs) to add or modify information on donors and donor organs, initiate the donor-recipient matching process, and record organ placement information once the organ is accepted by the transplanting center based on the United Network of Organ Sharing (UNOS) designated policies. The donor-recipient match process ranks all matched, active candidates with specific information entered for a given donor organ. The resulting match list is the guideline by which individual organs are offered to listed transplant candidates.
3. OPO personnel post donor information in an electronic file format for review by transplant personnel. Such files include the OPO's donor information form, ancillary confirmatory information such as ABO blood type, past social history, medical history, consent for donation, serology results for communicable diseases, digitized x-ray images, and other UNOS-mandated information regarding the donor. The intent of DonorNet was to remove an intermediary for organ acceptance and communicate directly with the physician

or surgeon making the organ acceptance decision. By viewing posted source documents in a UNOS-specified and consistent fashion in DonorNet, transplant center personnel can reach an informed decision of whether to accept the organ for their transplant candidate [25,26]. Updating information into DonorNet is not a static process; however, currently, there is no way to notify the potential transplant centers of updated donor information other than the OPC reaching out with a phone call. Blockchain technology would avoid the time delay for a phone call and put in a system of checks and balances such that each party in the organ offer transaction acknowledges updated information, which is an important patient safety issue.

4. Once the organ is accepted by the transplant center, for example, kidney allografts, the retrieval process may occur long distances away from the transplant center. Currently, organs are shipped via transport companies, including commercial airlines and commercial courier services in some circumstances. With the implementation of blockchain technology, organs could be tracked more precisely along their transport route with a recording of each point of data entry and physical contact along with any important information about the condition of the organ during transport (such as pump perfusion numbers). According to the adoption framework being used [12], applications such as tracking access to EHRs that have a high degree of novelty and complexity or coordination will have slow adoption. Applications such as transplant organ availability and transport tracking of these organs are less innovative and require much less coordination with current EHR infrastructure, implying faster adoption. These functions allow for our organ transport example to keep patient information private and secure; however, it allows the transplant center to allocate limited resources, such as operating room time, in a more efficient manner. In our example, the organ and blood type may remain in the public domain for all users to know which organ is being tracked. However, PHIs such as the donor or the recipient may be protected under encryption and only those with a private key will have access to that data.
5. The next branch of the tree, medical devices, including implanted medical device technology, fall under the same branch as wearable medical devices. Implanted technology includes cardiovascular defibrillators, pacemakers, heart valves, and infusion pumps that must be surgically embedded into a patient. For example, in 2017, 465,000 pacemakers in the United States, and an additional 280,000 pacemakers internationally, were recalled by Abbott because of a design flaw in which hackers could access and *modify programming commands to the implanted pacemaker, which could result in patient harm from rapid battery depletion or administration of inappropriate pacing* [27]. The results of this hacking could lead to serious medical events or death. This example demonstrates where the same technology that allows for medical devices to be tracked along the supply chain can provide enhanced security for implanted medical devices. The security provided through the public or private keys, which allow verification to add new blocks or alter blocks to the chain, is valuable through the cryptographic hash function. Wearable medical devices and data from these devices are also included as part of health care. Wearable technology that can be improved through blockchain includes oxygen tanks, glucose monitors, and heart monitors. Implanted technology, including cardiovascular defibrillators, pacemakers, and infusion pumps, must be surgically embedded into a patient. For example, in the case of a glucose monitor, the patient or provider can issue a private key that *could automatically and securely record a patient's blood glucose levels, and then, potentially, communicate with an insulin delivery device to maintain blood glucose at a healthy level* [28]. This tracking of wearable medical devices would allow for information to be tracked by patients and providers in real time.

Finally, the last branch under medical devices is high-tech medical machines. High-tech machines in hospitals include the da Vinci Surgical System, NvisionVLE Imaging System for Advanced optical coherence tomography imaging, and even atom-smashing machines used in cancer treatment. Similar to high-value medical items from the other branch of this tree, tracking these items along the supply chain is important for the machine's maintenance and performance logs. Applying blockchain technology to the maintenance and performance logs would require any person logging in, such as a provider completing a procedure or an engineer completing maintenance, to log in with their private key. A record would be provided for every use, update, or repair performed on these expensive machines while clearly tracking the machines' use throughout their lifecycle.

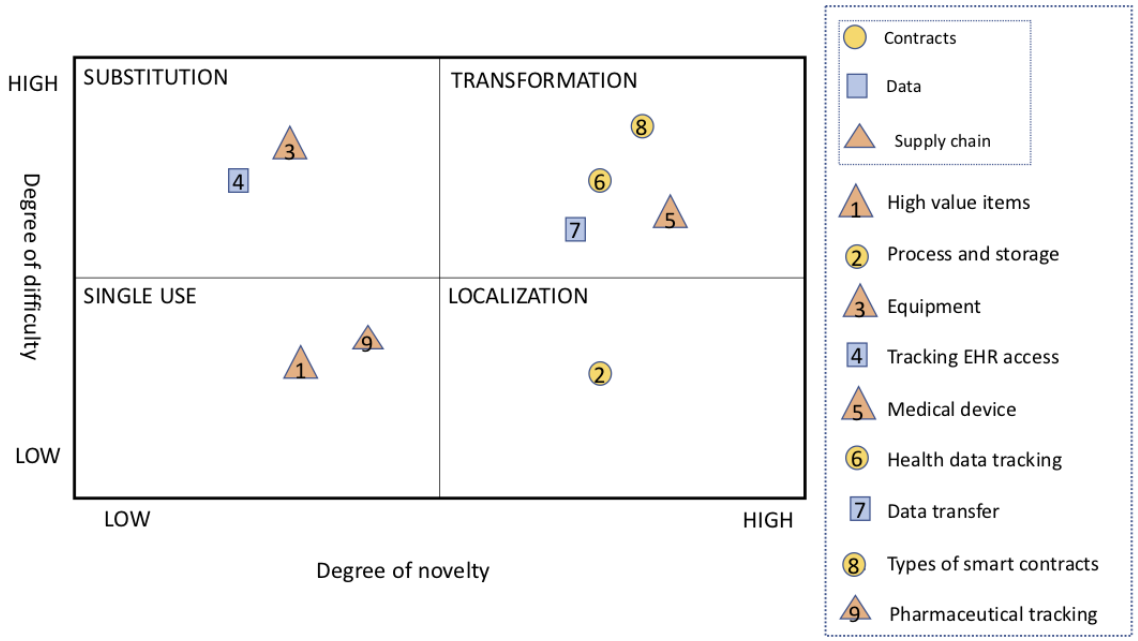
Blockchain has the ability to affect organ transplants not just in supply chain management but also in smart contracts and data exchange. Blockchain acts as a replacement for paper records and allows for a real-time view of each transaction along the supply chain process. As there is no central authority, all users have the authority to add transactions, but they may not change any history. This provides transparency in the process and makes blockchain a safe, interoperable solution to organ transplant and other medical technologies [29].

Predictions for Blockchain Technology

Figure 5 illustrates the use of the framework by Iansiti to display the degree of novelty and complexity of each proposed application and the speed of likely adoption. The first quadrant is single use. These technologies are *low-novelty and low-coordination applications that create better, less costly, highly focused solutions* [12]. The location of each mark in the quadrant indicates the level of novelty and complexity. Email is a non–health care single-use example, which provides an inexpensive alternative to traditional mail. Blockchain technology in its beginnings fell into this quadrant when it was being utilized only as cryptocurrency. In the health care field, high-value items supply chain management fits into the single-use category. This quadrant would be the easiest and quickest to adapt into health care organizations because it is low novelty and requires low levels of coordination and complexity to implement. One example of application in this

quadrant is tracking high-value items such as organs, blood, and IV fluids. The ability to trace real-time movement along the supply chain provides unprecedented traceability and transparency for organizations. Another example is pharmaceutical drug tracking. Recently, SAP and Chronicled teamed up to produce a blockchain solution to detect counterfeit drugs. The application would allow multi-user verification to track prescription drug routing within the SAP collaboration hub [30]. The adoption of a blockchain would allow wholesalers and pharmaceutical companies to increase information sharing across the industry to improve patient safety. The low levels of coordination, complexity, and novelty make these technologies contenders for the first blockchain technology to be implemented in health care. The ease of implementation, transparency, accountability, and efficiency would be immediately beneficial for the installation of these technologies.

Figure 5. Evaluation of the proposed applications using Marco's framework. EHR: electronic health data.



Although blockchain is not a novel concept, the use of blockchain itself may not be necessary in all of health care. Health care values privacy, transparency, and integrity more than the anonymity of user input [31]. Therefore, utilizing parts of blockchain technology, such as tamper-resistant technology, may increase the security of patient data and supply chain of high-cost necessities.

The next quadrant is localization, which *comprises innovations that are relatively high in novelty but need a limited number of users to create immediate value; therefore, it is still relatively easy to promote their adoption* [12]. This indicates that health care organizations would build on their single-use quadrant to branch out into implementing other blockchain-based technologies. Building of the single use of a high-value item supply chain would be the localization of process and storage. This would be the implementation of various technologies under the data exchange and contract trees, which would allow for blockchain technologies to begin storing data and processes on a blockchain network. For example, the storage of signed smart contracts or records of those who have accessed patients' EHRs.

The third quadrant is substitution, which is the predicted utilization of blockchain that will become compelling substitutes or replacements of existing technologies. This technology can only be implemented by building off the foundational technology implemented in the single-use and localization quadrants. This technology is "relatively low in novelty because they build on existing single-use and localized applications... These innovations aim to replace entire ways of doing business." An example of this would be the different cryptocurrencies that have evolved out of Bitcoin. In this quadrant for the health care industry, there is the equipment supply chain or the implantable and wearable medical devices, tracking EHR access, and supply chain to improve current issues of cost and transparency. However, this technology can only be implemented by building on the foundational technology implemented in the single-use and localization quadrants.

The last quadrant is transformation, which, like substitution, is high in its degree of coordination and complexity to implement and high in its degree of novelty. This technology *fundamentally changes the way businesses are created and capture value* [12]. Industry leaders use these technologies as keystones that proactively organize, influence, and coordinate the spread of networks of communities, users, and organizations [12]. This technology can be used for various types of smart contracts, health data tracking, the medical device supply chain, and data exchange. Smart contracts have the power to transform the system as they automate payments once the negotiated conditions are met in addition to being completely transparent and accessible. Health tracking data and data exchange will allow for better analytics to occur that will be easier for patients

to understand and allow for bigger data to be more easily manipulated for better understanding. The items under transformation are long-term goals, whereas single-use application and localization applications can be achieved in the near future.

Example of an Organ Transport Database and Transpiration Tracking

There is no lack of blockchain utilization in the health care and supply chain. One major area of interest for blockchain technology that has not been thoroughly investigated in health care is organ transplant supply chain management. The distributed blockchain ledger will record any individual who has come in contact with the organ as well as their location. Individual transactions are added into a block and cannot be altered. This system to record transportation information will not only help track organs but also distribute them to the appropriate locations within an allotted amount of time. All users who have access to a cryptographic key are able to track each step of their transportation. This ability is because each network user has their own replicated copy that allows for accountability and inhibits information tampering. Unlike in other industries, if 1 step of the supply chain process is delayed or halted, a recipient may lose their opportunity to be transplanted because organ transportation is highly time sensitive. Implementing blockchain technology and adding time alerts may remove the opportunity for error from the process.

Overall, the lack of novelty and complexity makes supply chain management in organ transplant a viable and feasible solution. Organ transport tracking via blockchain is feasible because it is easily adopted owing to its low novelty and complexity, as shown in Figure 5. Supply chain management of high-value items is common in health care and other industries. As discussed in the framework, using blockchain in supply chain management for high-value items is not a novel idea. Industries such as fishing and farming have been using this technology for years with great success. Implementing the technology in health care should be a relatively smooth transition as many flaws from lessons learned by the other industries have been discussed in this paper.

Startup technology companies have already begun experimenting with using blockchain in pharmaceutical supply chain tracking. Blockchain technologies have helped optimize the manufacturing, distribution, and dispensing of pharmaceutical products. They found that managing the supply chain *end-to-end* has decreased the amount of counterfeit medicines used for patients [32]. Another advantage of using blockchain technology in pharmaceuticals is that it can manage drug recalls easily by locating exactly who had access to the drug in question. The ability to use smart contracts also helps automate processes to reduce costs.

The complexity of developing the blockchain codes is simple and user friendly once implemented. Individuals provided with the key to enter the blockchain will be able to identify organs and their locations. This technology will increase the accountability of all involved personnel, from the procurement of the organ to the final delivery of the organ. Another factor in reducing the complexity of implementation of the supply chain management of transport organs is that this process is likely not part of the organization's EHR. This independency from EHRs and its vendor makes the adoption of this application simple.

As discussed throughout the paper, tracking transport organs is likely for fast adoption because of its common use in supply chain management, independence from the EHR, and advantages to the current infrastructure. The implementation of blockchain will facilitate communication in real time, identify any issues during transport, and allow for better efficiency in the use of limited resources (eg, hospital beds for admissions or time). The implementation of blockchain to supply chains in health care can help track each step of the transportation process from procurement of the organ to the surgery of the recipient.

Conclusions

The transformative power of blockchain has proven beneficial in non–health care industries. Blockchain has benefited non–health care industries with improved supply chains in terms of accountability, traceability, and transparency. It has proven its ability to improve outdated methods in other industries and can be applied to the health care industry as well. Application to health care will first be as single use in high-value items' supply chain, but we predict that it will eventually reach transformation and disrupt the industry through innovations such as smart contracts. Enabling blockchain to solve many complex issues that the health care industry encounters today will allow a transformation that will lead to improved and innovative methods for viewing the health care industry. Although we discussed many uses of blockchain in this paper, we do not suggest using blockchain in its current form. Depending on the needs of a company, the DLT can be adapted from traditional blockchain uses to suit its needs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Blockchain in other industries.
[DOCX File , 18 KB - jmir_v22i9e17423_app1.docx]

References

1. Krawiec R. Blockchain: Opportunities for Health Care. Deloitte US. 2016. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-blockchain-opportunities-for-health-care.pdf [accessed 2020-07-29]
2. Blockchain: The Chain of Trust and its Potential to Transform Healthcare – Our Point of View. Office of the National Coordinator for Health Information. 2016. URL: https://www.healthit.gov/sites/default/files/8-31-blockchain-ibm_ideation-challenge_aug8.pdf?source=post_page---------- [accessed 2020-08-05]
3. From Shore to Plate: Tracking Tuna on the Blockchain. Provenance: Every Product Has a Story. 2016. URL: https://www.provenance.org/tracking-tuna-on-the-blockchain [accessed 2020-07-20]
4. Case Studies. Provenance: Every Product Has a Story. 2018. URL: https://www.provenance.org/case-studies [accessed 2020-07-20]
5. Reinforcing Organic Certification in the Digital Age. Provenance: Every Product Has a Story. 2017. URL: https://www.provenance.org/case-studies/soil-association [accessed 2020-07-20]
6. Seam T. The Seam to Form Cotton Blockchain Consortium with IBM on Hyperledger Fabric. Google Docs. 2017. URL: https://docs.google.com/document/d/1xQecMjByN_gMT_SPyXS4MYsJJMK0nT9evvBwBSL4i0U/edit [accessed 2020-07-20]
7. Increasing Transparency in Fashion With Blockchain. Provenance: Every Product Has a Story. 2018. URL: https://www.provenance.org/case-studies/martine-jarlgaard [accessed 2020-07-20]
8. Biesheuvel T. Diamond Trade Turns to Technology That Keeps Bitcoin Anonymous. Bloomberg. 2018. URL: https://www.bloomberg.com/news/articles/2018-01-16/diamond-trade-turns-to-technology-that-keeps-bitcoin-anonymous [accessed 2020-07-20]
9. Roberts J. The Diamond Industry Is Obsessed With the Blockchain. Fortune. 2017. URL: https://fortune.com/2017/09/12/diamond-blockchain-everledger/ [accessed 2020-08-05]
10. Condliffe J. The World's Largest Shipping Company Trials Blockchain to Track Cargo. MIT Technology Review. 2017. URL: https://www.technologyreview.com/2017/03/06/153405/the-worlds-largest-shipping-company-trials-blockchain-to-track-cargo/ [accessed 2020-08-05]
11. Hackett R. IBM and Maersk Are Creating a New Blockchain Company. Fortune. 2018. URL: https://fortune.com/2018/01/16/ibm-blockchain-maersk-company/ [accessed 2020-08-05]
12. Iansiti M, Lakhani KR. The Truth About Blockchain. Harvard Business Review. 2017. URL: https://hbr.org/2017/01/the-truth-about-blockchain [accessed 2020-08-05]
13. Dimitrov DV. Blockchain applications for healthcare data management. Healthc Inform Res 2019 Jan;25(1):51-56 [FREE Full text] [doi: 10.4258/hir.2019.25.1.51] [Medline: 30788182]
14. Culver K. Blockchain Technologies: a Whitepaper Discussing How the Claims Process Can Be Improved. The Blockchain Lab. 2016. URL: https://blockchainlab.com/pdf/3-47-whitepaperblockchainforclaims_v10.pdf [accessed 2020-08-05]
15. Ekblaw A. A Case Study for Blockchain in Healthcare: 'MedRec' Prototype for Electronic Health Records and Medical Research Data. Office of the National Coordinator for Health Information. 2016. URL: https://www.healthit.gov/sites/default/files/5-56-onc_blockchainchallenge_mitwhitepaper.pdf [accessed 2020-08-05]
16. Linn L, Martha MD, Koo B. Blockchain For Health Data and Its Potential Use in Health IT and Health Care Related Research. Office of the National Coordinator for Health Information. 2016. URL: https://www.healthit.gov/sites/default/files/11-74-ablockchainforhealthcare.pdf [accessed 2020-08-05]
17. Peterson K. A Blockchain-Based Approach to Health Information Exchange Networks. Colleaga: Harnessing the Power of Collective Intelligence. 2016. URL: https://colleaga.org/sites/default/files/12-55-blockchain-based-approach-final.pdf [accessed 2020-08-05]
18. Vian K, Voto A, Haynes-Sanstead K. A Blockchain Profile for Medicaid Applicants and Recipients. Seattle Education. 2016. URL: https://seattleducation.com/category/a-blockchain-profile-for-medicaid-applicants-and-recipients-institute-for-the-future/ [accessed 2020-08-05]
19. Blockchain Supply Chain Solutions. IBM. 2018. URL: https://www.ibm.com/blockchain/supply-chain/ [accessed 2020-07-20]
20. Rogers A. Solve Genomics with the Blockchain? Why the Hell Not. Wired. 2018. URL: https://www.wired.com/story/solve-genomics-with-blockchain/ [accessed 2020-08-05]
21. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
22. Trust & Automation Between Companies. Chronicled. 2019. URL: https://www.chronicled.com [accessed 2020-07-20]
23. Ivan D. Moving Toward a Blockchain-Based Method for the Secure Storage of Patient Records. Office of the National Coordinator for Health Information. 2016. URL: https://www.healthit.gov/sites/default/files/9-16-drew_ivan_20160804_blockchain_for_healthcare_final.pdf [accessed 2020-07-20]
24. US Homeland Security Lists Blockchain as Covid-19 Critical Service. Ledger Insights. 2020. URL: https://www.ledgerinsights.com/us-homeland-security-lists-blockchain-as-covid-19-critical-service/ [accessed 2020-07-20]
25. Massie A, Desai NM, Montgomery RA, Singer AL, Segev DL. Improving distribution efficiency of hard-to-place deceased donor kidneys: predicting probability of discard or delay. Am J Transplant 2010 Jul;10(7):1613-1620 [FREE Full text] [doi: 10.1111/j.1600-6143.2010.03163.x] [Medline: 20642686]

26. Gerber D, Arrington CJ, Taranto SE, Baker T, Sung RS. DonorNet and the potential effects on organ utilization. Am J Transplant 2010 Apr;10(4 Pt 2):1081-1089 [FREE Full text] [doi: 10.1111/j.1600-6143.2010.03036.x] [Medline: 20420654]
27. Over Half a Million Hackable Pacemakers Can Now Be Fixed. Business News - Latest Headlines on CNN Business - CNN. 2017. URL: http://money.cnn.com/2017/08/30/technology/business/pacemakers-security-update [accessed 2020-07-20]
28. Marr B. A Complete Beginner's Guide To Blockchain. Forbes. 2017. URL: https://www.forbes.com/sites/bernardmarr/2017/01/24/a-complete-beginners-guide-to-blockchain/#47b8e3a6e607 [accessed 2020-08-05]
29. Hann DW. Someone Could Die: Blockchain in the Highest Trust Situations. Digital Oil and Gas. 2018. URL: http://digitaloilgas.com/someone-could-die-blockchain-in-the-highest-trust-situations/ [accessed 2020-07-20]
30. Garrity M. SAP Launches Blockchain Platform to Verify Prescription Drugs. Becker's Hospital Review - Healthcare News. 2019. URL: https://www.beckershospitalreview.com/supply-chain/sap-launches-blockchain-platform-to-verify-prescription-drugs.html?oly_enc_id=0128A5649590B8U [accessed 2020-07-20]
31. Wust K, Gervais A. Do you Need a Blockchain? In: Crypto Valley Conference on Blockchain Technology. 2018 Presented at: CVCBT'18; June 20-22, 2018; Zug, Switzerland. [doi: 10.1109/cvcbt.2018.00011]
32. Clauson KA, Breeden EA, Davidson C, Mackey TK. Leveraging blockchain technology to enhance supply chain management in healthcare. Blockchain Healthcar Today 2018 Mar 23;1:-. [doi: 10.30953/bhty.v1.20]

Abbreviations

DLT: distributed ledger technology
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
IoT: Internet of Things
IT: information technology
IV: intravenous
OPC: organ procurement coordinator
OPO: organ procurement organization
PHI: personal health information

Edited by G Eysenbach; submitted 11.12.19; peer-reviewed by A Hasselgren, JT te Gussinklo, T Ueno, W Zhang; comments to author 01.04.20; revised version received 27.05.20; accepted 14.06.20; published 17.09.20.

Please cite as:
Gaynor M, Tuttle-Newhall J, Parker J, Patel A, Tang C
Adoption of Blockchain in Health Care
J Med Internet Res 2020;22(9):e17423
URL: https://www.jmir.org/2020/9/e17423
doi:10.2196/17423
PMID:32940618

Viewpoint

A New Era of Epidemiology: Digital Epidemiology for Investigating the COVID-19 Outbreak in China

Zonglin He[1,2*], MBBS; Casper J P Zhang[3*], MPH, PhD; Jian Huang[4*], PhD, MPH; Jingyan Zhai[1], MBBS; Shuang Zhou[1], BNurs; Joyce Wai-Ting Chiu[2], MSc, BSc, MBBS; Jie Sheng[5], BA; Winghei Tsang[1], MBBS; Babatunde O Akinwunmi[6,7], MD, PhD, MMSc; Wai-Kit Ming[1], MD, MPH, PhD, MMSc, EMBA

[1]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[2]Faculty of Medicine, International School, Jinan University, Guangzhou, China

[3]School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[4]MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, St Mary's Campus, Imperial College London, London, United Kingdom

[5]College of Economics, Jinan University, Guangzhou, China

[6]Center for Genomic Medicine, Massachusetts General Hospital, Harvard University, Boston, MA, United States

[7]Pulmonary & Critical Care Medicine Unit, Asthma Research Center, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, MPH, PhD, MMSc, EMBA
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
601 Huangpu W Ave
Guangzhou,
China
Phone: 86 14715485116
Email: wkming@connect.hku.hk

Abstract

A novel pneumonia-like coronavirus disease (COVID-19) caused by a novel coronavirus named SARS-CoV-2 has swept across China and the world. Public health measures that were effective in previous infection outbreaks (eg, wearing a face mask, quarantining) were implemented in this outbreak. Available multidimensional social network data that take advantage of the recent rapid development of information and communication technologies allow for an exploration of disease spread and control via a modernized epidemiological approach. By using spatiotemporal data and real-time information, we can provide more accurate estimates of disease spread patterns related to human activities and enable more efficient responses to the outbreak. Two real cases during the COVID-19 outbreak demonstrated the application of emerging technologies and digital data in monitoring human movements related to disease spread. Although the ethical issues related to using digital epidemiology are still under debate, the cases reported in this article may enable the identification of more effective public health measures, as well as future applications of such digitally directed epidemiological approaches in controlling infectious disease outbreaks, which offer an alternative and modern outlook on addressing the long-standing challenges in population health.

(J Med Internet Res 2020;22(9):e21685) doi:10.2196/21685

KEYWORDS

digital epidemiology; COVID-19; risk; control; public health; epidemiology; China; outbreak; case study

Introduction

A pneumonia-like coronavirus disease (COVID-19) outbreak caused by a newly identified coronavirus, SARS-CoV-2, swept across China in early 2020. As of early June, 215 countries or regions have reported confirmed cases, with 6,799,713 confirmed cases and 397,388 deaths, and a case fatality rate over 5.84% worldwide [1]. With the increasing incidence of confirmed cases, corresponding spread control policies and emergency actions are taking place. Holiday travel related to

the Spring Festival in China has led to great difficulties in tracking suspected cases for outbreak control.

Conventional epidemiology dating back to the 1800s mainly relies on health-related data such as information gathered within health care systems, medical records, or insurance systems. Such data can only be collected and recorded from diagnosed or treated patients; therefore, it would be outdated and hinder the corresponding management efforts upon the abrupt outbreak of infectious diseases [2].

The public health measures that showed effectiveness in previous infection outbreaks (ie, mass use of face masks, social distancing, and home quarantine) were also implemented in the COVID-19 outbreak. Although the effectiveness of these public health measures in this outbreak is not clear, the availability of multidimensional media network data can provide an alternative outlook that takes advantage of the recent rapid development of information and communication technologies, allowing for better tracing and control of the disease spread. The quantity and dimensionality of data have substantially increased along with the continued development of technologies (eg, telecommunication), revolutionizing the way we communicate. Such technologies have shown great potential in terms of convenience and precision for the surveillance and modelling of infectious diseases such as influenza and severe acute respiratory syndrome, through extracting information from electronic health (eHealth), electronic payments, the internet, and social media [3,4]. This also brings epidemiology into a new era, that of so-called digital epidemiology [5], where digital data or data that were generated outside of the public health system are used, as proposed by some scholars [6]. Social media provides much of the data generated on the internet; by examining the search index or the texts posted, researchers can foresee the outbreak of an infectious disease. If certain keywords were searched for many times during a short period of time, this could indicate an infectious disease in the community; Google Flu Trends (Google Inc) makes use of this type of data [7-9]. Moreover, the spatiotemporal data related to individual behaviors can be extrapolated from the use of electronic payments, cellular service, or social media to study the distribution, incidence, and etiology of a disease, contributing to disease prediction and prevention [7,10-12]. However, some scholars in digital epidemiology have excessively used the internet, web-based systems, or network surveillance of media information, which may be biased and constrained by information overload, false reports, a lack of specificity of signals, and sensitivity to external forces [10].

Nowadays, advances in mobile applications have enabled users to perform daily activities on their mobile phones, including making electronic payments and checking social media. The data on each activity performed, including the location of the mobile user, were also stored (Figure 1). Generally, there are three types of electronic data streams in the field of epidemiology, namely medical encounter data (eg, electronic records of medical institutions), participatory syndrome data (eg, personal health data, data from the population), and nonhealth digital data (eg, data from internet search engines, social media, or mobile use) [13]. The everyday movements of individuals create a dynamic link that connects people, which can be used to study the geographical spread and sustained transmission of infectious diseases [5]. In the past, population movements were traditionally estimated using travel surveys, road networks, or small-scale GPS studies, which have long hindered efforts to understand these dynamics [5]. Diverse types of digital trace data may enhance exposure measurement and facilitate strong tests of specific routes of transmission [5]. These data sources, if used appropriately, can provide preliminary and timely information about disease outbreaks and related events around the world. Furthermore, these sources enable a reduced time between initial detection of an outbreak and formal recognition of an outbreak, thus allowing for a more expedited response to such public health threats [14]. Since the epidemic spread is related to location-specific human contact patterns [15,16], it is deemed that more accurate estimates of transmission routes and the number of infection cases can be achieved by using available big data derived from mobile phones and video surveillance. Here, we present two publicly reported cases of COVID-19 in China that demonstrated the significant role that digital data can have in modernizing epidemiological investigation, showing the potential of guiding public health measures accordingly (Figure 2).

Figure 1. An infographic illustrating the development of digital epidemiology and its application in controlling infectious disease epidemics. CDC: Centers for Disease Control.

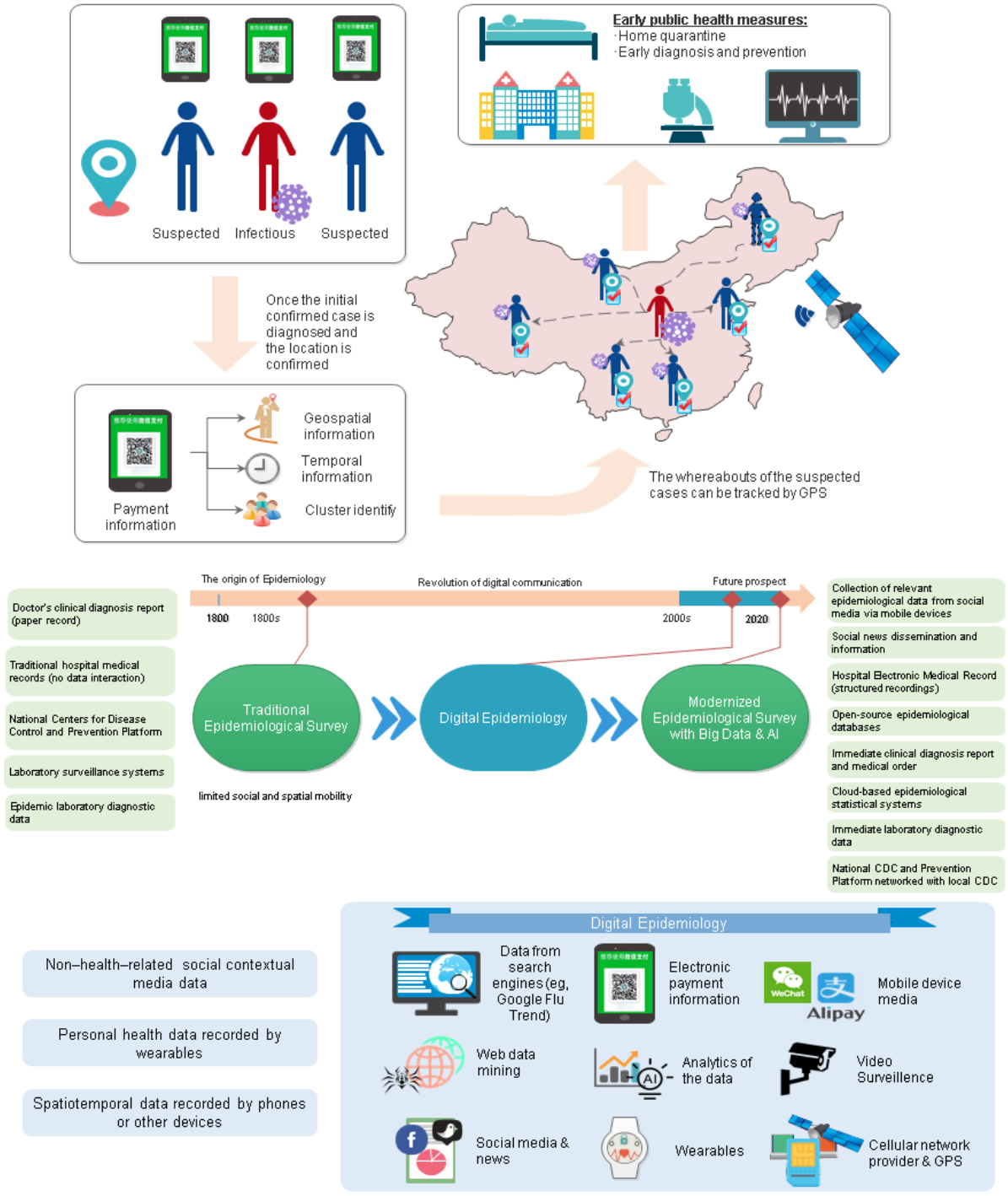
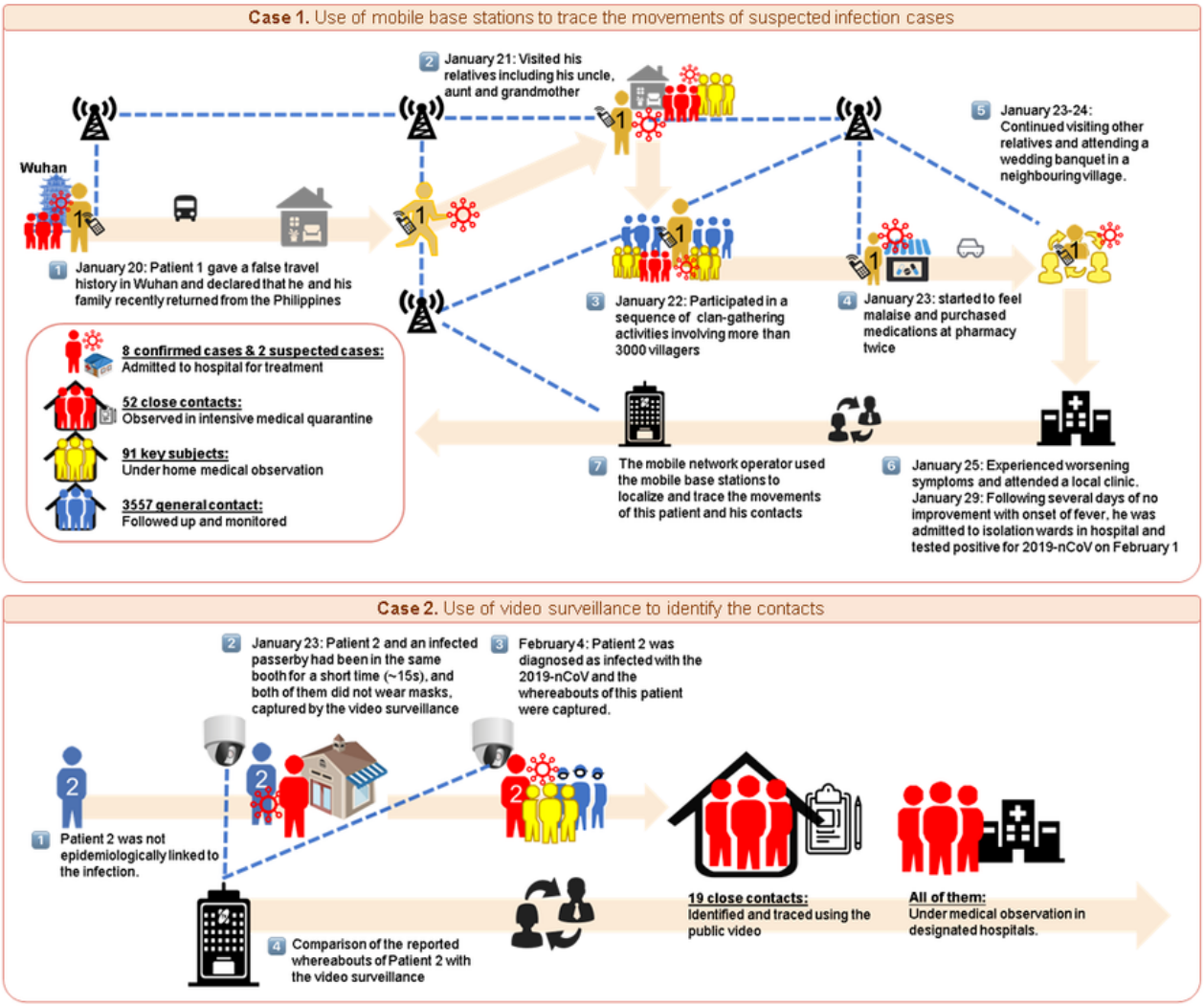


Figure 2. The application of digital epidemiology in the outbreak of COVID-19. Case 1: Use of mobile base stations to trace the movements of suspected infection cases. Case 2: Use of video surveillance to identify the contacts.



By using a phone carrier's mobile phone tracking system and scrutinizing the data transmission between different base stations under the authorization of the local government, 3557 people were identified as general contacts and 8 people were confirmed as having infections. Strict measures were then undertaken: 8 confirmed cases and 2 suspected cases were admitted to hospital to receive treatments, 52 close contacts were observed in intensive medical quarantine, 91 key subjects received home medical observation, and all 3557 general contacts were followed up and monitored.

Case 1

A male, a resident of Village A, City A, China, was diagnosed with COVID-19 on February 1, 2020 [17-19], after returning from Wuhan, where he ordinarily lives and works. To avoid unnecessary interruption to his schedule, he claimed that he and his family recently returned from the Philippines rather than Wuhan when they arrived at the village on January 20, without symptoms, prior to the lockdown of Wuhan (Figure 2). During the following days, he resided with his father and younger brother in the village and was involved in several activities. On January 21 and 22, he visited his relatives and attended a series of clan-gathering activities that more than 3000 people partook in. Starting on January 23, he felt malaise. He purchased medications at a pharmacy twice on January 23. Despite his symptoms, he continued visiting other relatives on the same day. On January 24, he attended a wedding banquet in a neighboring village. He experienced worsening symptoms on January 25 and decided to attend a local clinic. It was recommended that he undergo a home quarantine given his lack of fever. Following several days of no improvement and the onset of fever, he was admitted to an isolation ward in a hospital on January 29 and tested positive for COVID-19 on February 1.

Case 2

A 56 year old male, living in Town B, City B, China, was diagnosed as positive for COVID-19 on February 4, 2020, and was quarantined and received treatment in a designated medical institution. Through traditional epidemiological investigation methods, this patient was determined to have no history of residence or travel in the epidemic area and no exposure to wild

animals in the 14 days before the onset of symptoms. In addition, he had no acquaintances with confirmed cases in his local district. However, the activities of this patient were captured by video surveillance. After referring to the videos, it was determined that the patient spent a short period standing near a stranger at the same booth in a farmer's market at 7:47 AM on January 23. They were not wearing face masks. This stranger was in fact a confirmed case living in the same district.

Using video surveillance, the whereabouts of this patient were retrieved, which resulted in the identification of 19 subjects with close contact, who were then put under observation in designated hospitals to prevent further contamination.

Discussion

These two cases are examples of the successful application of emerging technology in monitoring people's movements during disease outbreaks, with the potential to offer near real-time estimation of disease-related activities and fast identification of potentially infected subjects. The surveillance work in both cases was led by the local governments, and the privacy of the subjects remained protected and personal information was not leaked; the information was only accessible by designated authorities within the local governments.

During the COVID-19 outbreak, there has been general agreement regarding the lack of readiness for such a viral outbreak. Although China's government introduced strict measures to restrict gathering and travel during the outbreak, the virus still spread due to its high infection rate, even during the incubation period. The outbreak could have been better controlled if better surveillance systems and high-end technologies were used to incorporate spatiotemporal movement data in models of the potential transmission patterns. The outbreak of COVID-19 has prompted a discussion on the incorporation of digital data in epidemiological research. The use of digital data can enhance traditional epidemic surveillance as well as digital epidemiology–directed applications, including incident infections, viral sequencing, improved infectious disease outbreak predictions, suspected contacts detection, early prevention and management, real-time numerical forecasting of pandemics, and evaluating the effectiveness of disease response strategies or interventions [13,20-24]. The use of spatiotemporal information generated by the daily usage of online communication tools, such as WeChat and Alipay, could play an important role in controlling the spread of this disease and others, if properly used. In China, a color - coded health code and travel card system was created. The system tracks where citizens have been during the last 14 days through phone carriers, whose system logs can determine whether a given citizen's phone connected to base stations in high-risk areas. Thus, the system will note which citizens have been to high-risk regions, and the provided code then dictates where citizens can go (ie, whether they should continue quarantining or are able to leave the house) [25,26].

With the rapid development of China's economy and the widespread adoption of cell phones, mobile payment systems have also developed rapidly. There are two mobile payment operators, Alipay and WeChat, which currently cover more than 90% of the domestic market in China, and they are leaders in the field of third-party payment. WeChat and Alipay are secure and convenient, and they have penetrated every aspect of people's lives (eg, transactions, online shopping, self-service, public transport, and personal finances) [27]. These payment systems also obtain multidimensional data from users, including payment information, GPS information, and social media information [27], which can be used to help monitor and control the spread of infectious diseases.

Moreover, the popularization of wearable devices has enhanced our ability to collect data regarding spatial and temporal aspects of human movements with higher precision [28], affording a much more detailed identification and stratification of social behaviors [29], complementing previous work based on large-scale surveys and self-reported information [5,30]. These data provide one of today's most exciting opportunities to study human mobility and its influence on disease dynamics [31].

Despite the merits of using such technologies and data, several concerns still remain. First, validation of real-world data should be considered because the extraction of meaningful data from social networks has always been challenging [13,22,32]. Second, although the cases discussed in this article used a novel stream of data, the investigation methods and strategies were still outmoded. Therefore, how such digital data can be more effectively used and analyzed, using analytic algorithms with scientific justification and statistical power, requires further exploration [33]. Third, the legal and ethical aspects of using digital data remain questionable. The use of digital data has been extensively debated worldwide. Some of the electronic traces that we leave behind as digital citizens are meant to be public, while others are not, resulting in ethical and legal challenges [34,35]. Regarding the ethics surrounding public health and digital epidemiology, there are the competing issues of protecting and promoting the health of populations and potentially causing individual harm as a result of collecting data from digital networks [35,36]. These two COVID-19 cases in China serve as a successful example of how digital data generated by companies and used by local governments can be used to mitigate the spread of COVID-19, by identifying people who have travelled to high-risk areas or tracing people who have contacted people with COVID-19. Indeed, such data should be covered by data-protection regulations, and privacy and confidentiality should be guaranteed, but there would have been no other way for the relevant authorities to obtain this data. In addition, the issue of privacy has been extensively discussed [37-41]. Fourth, false discrimination has been demonstrated in previous studies as a result of incorrect identification of internet users; thus, an improvement in this aspect is required. Fifth, multidimensional data such as the data extracted from electronic payments in China may not be available in other countries; thus, further exploration of local contexts is needed. Finally, issues related to data access, data sharing, user privacy, and data security still require attention, yet public health takes precedence in such situations. The two abovementioned cases serve as perfect examples of local governing bodies taking part in epidemiological research using digital data. Therefore, we hold an optimistic view on the further implementation of digital

epidemiology for disease outbreaks, especially following related achievements and experiences during the COVID-19 outbreak.

This article demonstrated the plausibility of using digital epidemiology to control and prevent infections, based on two real-life cases during the COVID-19 outbreak in China. Taking advantage of emerging information and communication technologies and accessible multidimensional spatiotemporal data for monitoring people's movements, this modernized epidemiological approach can help shed more light on the pattern of disease spread and contribute to identifying more effective public health measures to mitigate the negative impact of COVID-19. It can also be used to identify long-standing challenges in population health.

Authors' Contributions

WKM conceived the original idea. ZH, CJPZ, and JH developed the idea, collected the data, and generated the figures. ZH, CJPZ, JH, WKM, JZ, SZ, and JWTC drafted the manuscript. CJPZ, BA, and JH revised and edited the manuscript. All authors contributed to the development and writing of the paper.

Conflicts of Interest

None declared.

References

1. World Health Organization. WHO Director-General's opening remarks at the Mission briefing on COVID-19 - 12 March 2020. 2020 Mar 12. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-mission-briefing-on-covid-19---12-march-2020 [accessed 2020-03-15]
2. Lippi G, Cervellin G. Is digital epidemiology reliable?-insight from updated cancer statistics. Ann Transl Med 2019 Jan;7(1):15 [FREE Full text] [doi: 10.21037/atm.2018.11.55] [Medline: 30788362]
3. Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: Epidemiology in the era of big data. Epidemiology 2015 May;26(3):390-394 [FREE Full text] [doi: 10.1097/EDE.0000000000000274] [Medline: 25756221]
4. Fagherazzi G, Goetzinger C, Rashid M, Aguayo G, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
5. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the Importance of Digital Epidemiology. N Engl J Med 2013 Aug;369(5):401-404. [doi: 10.1056/nejmp1307752]
6. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
7. Bates M. Tracking Disease: Digital Epidemiology Offers New Promise in Predicting Outbreaks. IEEE Pulse 2017 Jan;8(1):18-22. [doi: 10.1109/mpul.2016.2627238]
8. Adawi M, Bragazzi N, Watad A, Sharif K, Amital H, Mahroum N. Discrepancies Between Classic and Digital Epidemiology in Searching for the Mayaro Virus: Preliminary Qualitative and Quantitative Analysis of Google Trends. JMIR Public Health Surveill 2017 Dec 01;3(4):e93 [FREE Full text] [doi: 10.2196/publichealth.9136] [Medline: 29196278]
9. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014 [FREE Full text] [doi: 10.1038/nature07634] [Medline: 19020500]
10. Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection — Harnessing the Web for Public Health Surveillance. N Engl J Med 2009 May 21;360(21):2153-2157. [doi: 10.1056/nejmp0900702]
11. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. Canadian Medical Association Journal 2009 Mar 12;180(8):829-831. [doi: 10.1503/cmaj.1090215]
12. Roche B, Gaillard B, Léger L, Pélagie-Moutenda R, Sochacki T, Cazelles B, et al. An ecological and digital epidemiology analysis on the role of human behavior on the 2014 Chikungunya outbreak in Martinique. Sci Rep 2017 Jul 20;7(1):5967 [FREE Full text] [doi: 10.1038/s41598-017-05957-y] [Medline: 28729711]
13. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. J Infect Dis 2016 Dec 01;214(suppl_4):S375-S379 [FREE Full text] [doi: 10.1093/infdis/jiw400] [Medline: 28830113]
14. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, et al. Global capacity for emerging infectious disease detection. Proc Natl Acad Sci U S A 2010 Dec 14;107(50):21701-21706 [FREE Full text] [doi: 10.1073/pnas.1006219107] [Medline: 21115835]
15. Anderson RM. Discussion: The Kermack-McKendrick epidemic threshold theorem. Bltn Mathcal Biology 1991 Mar;53(1-2):1. [doi: 10.1007/bf02464422]
16. Cao L, Li X, Wang B, Aihara K. Publisher's Note: Rendezvous effects in the diffusion process on bipartite metapopulation networks. Phys Rev E 2011 Nov 14;84(5):041936. [doi: 10.1103/physreve.84.059904]

17. Common against epidemic! More than 3,000 villagers in Jinjiang, Fujian province, were monitored due to fraud [Chinese]. CCTV News. 2020 Feb 05. URL: http://news.cctv.com/2020/02/05/ARTId8OoGN30uEp1JnLoUxDJ200205.shtml [accessed 2020-02-06]
18. A familial cluster case in Jinjiang, Fujian Province, China. Xinhuanet. 2020 Feb 02. URL: http://www.xinhuanet.com/2020-02/05/c_1125534821.htm?spm=C73544894212.P59511941341.0.0 [accessed 2020-02-06]
19. Quanzhou Public Security Bureau. The Municipal Public Security Bureau shall crack down on illegal and criminal activities related to the epidemic: Quanzhou Public Security Bureau; Feb 05, 2020.
20. van de Belt TH, van Stockum PT, Engelen LJLPG, Lancee J, Schrijver R, Rodríguez-Baño J, et al. Social media posts and online search behaviour as early-warning system for MRSA outbreaks. Antimicrob Resist Infect Control 2018 May 30;7(1):69 [FREE Full text] [doi: 10.1186/s13756-018-0359-4] [Medline: 29876100]
21. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One 2015 Oct 5;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
22. Wang L, Li X. Spatial epidemiology of networked metapopulation: an overview. Chin Sci Bull 2014 Jul 19;59(28):3511-3522 [FREE Full text] [doi: 10.1007/s11434-014-0499-8] [Medline: 32214746]
23. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. PLoS Negl Trop Dis 2017 Jan 13;11(1):e0005295 [FREE Full text] [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
24. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir Viruses 2014 May 23;8(3):309-316 [FREE Full text] [doi: 10.1111/irv.12226] [Medline: 24373466]
25. Yu A. Digital surveillance in post-coronavirus China: A feminist view on the price we pay. Gend Work Organ 2020 Jun 08:1 [FREE Full text] [doi: 10.1111/gwao.12471] [Medline: 32837009]
26. Xu J, Lin X, Pan X, Huang X, Chen Y, Zhou E. The trend of morbidity and mortality of Coronavirus disease 2019 under the first-level public health emergency response in China. Research Square 2020 May 22:1 [FREE Full text] [doi: 10.21203/rs.3.rs-26576/v1]
27. Li J, Wang J, Wangh S, Zhou Y. Mobile Payment With Alipay: An Application of Extended Technology Acceptance Model. IEEE Access 2019;7:50380-50387. [doi: 10.1109/access.2019.2902905]
28. Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton J, Vespignani A. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. PLoS ONE 2010 Jul 15;5(7):e11596 [FREE Full text] [doi: 10.1371/journal.pone.0011596]
29. Pentland A. Honest signals: how they shape our world. Cambridge, MA, USA: The MIT Press; 2010:1-208.
30. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008 Mar 25;5(3):e74 [FREE Full text] [doi: 10.1371/journal.pmed.0050074] [Medline: 18366252]
31. Brockmann D. Anomalous diffusion and the structure of human transportation networks. Eur Phys J Spec Top 2008 Apr;157(1):173-189. [doi: 10.1140/epjst/e2008-00640-0]
32. Luo W, Tay WP, Leng M. Identifying Infection Sources and Regions in Large Networks. IEEE Trans Signal Process 2013 Jun;61(11):2850-2865. [doi: 10.1109/tsp.2013.2256902]
33. Höhle M. A statistician's perspective on digital epidemiology. Life Sci Soc Policy 2017 Nov 24;13(1):17 [FREE Full text] [doi: 10.1186/s40504-017-0063-9] [Medline: 29177850]
34. Boyd D, Crawford K. Six Provocations for Big Data. SSRN Journal 2011 [FREE Full text] [doi: 10.2139/ssrn.1926431]
35. Denecke K. An ethical assessment model for digital disease detection technologies. Life Sci Soc Policy 2017 Sep 20;13(1):16 [FREE Full text] [doi: 10.1186/s40504-017-0062-x] [Medline: 28929347]
36. Faden R, Bernstein J, Shebaya S. Public health ethics. Stanford Encyclopedia of Philosophy. Stanford, CA, USA: Metaphysics Research Lab, Stanford University; 2020. URL: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=publichealth-ethics [accessed 2020-08-26]
37. Oliver N, Letouzé E, Sterly H, Delataille S, De NM, Lepri B. arXiv. Mobile phone data and COVID-19: Missing an opportunity? 2020 Mar 27:1 [FREE Full text]
38. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr 27;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]
39. Park S, Choi GJ, Ko H. Information Technology-Based Tracing Strategy in Response to COVID-19 in South Korea-Privacy Controversies. JAMA 2020 Apr 23;323(21):2129 [FREE Full text] [doi: 10.1001/jama.2020.6602] [Medline: 32324202]
40. Barbieri C, Darnis JP. Technology: An exit strategy for COVID-19. 2020 Mar. URL: https://www.iai.it/en/pubblicazioni/technology-exit-strategy-covid-19 [accessed 2020-08-26]
41. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]

Abbreviations

COVID-19: coronavirus disease

Edited by G Eysenbach; submitted 22.06.20; peer-reviewed by A Al-Hasan, Y He; comments to author 08.07.20; revised version received 23.07.20; accepted 11.08.20; published 17.09.20.

Please cite as:
He Z, Zhang CJP, Huang J, Zhai J, Zhou S, Chiu JWT, Sheng J, Tsang W, Akinwunmi BO, Ming WK
A New Era of Epidemiology: Digital Epidemiology for Investigating the COVID-19 Outbreak in China
J Med Internet Res 2020;22(9):e21685
URL: https://www.jmir.org/2020/9/e21685
doi:10.2196/21685
PMID:32805703

Viewpoint

Cybersecurity Risks in a Pandemic

Christina Meilee Williams[1*], BS; Rahul Chaturvedi[2*], BS; Krishnan Chakravarthy[3], MD, PhD

[1]Trinity College of Arts & Sciences, Duke University, Durham, NC, United States

[2]School of Medicine, University of California, San Diego, La Jolla, CA, United States

[3]Department of Anesthesiology, University of California, San Diego, La Jolla, CA, United States

[*]these authors contributed equally

Corresponding Author:
Rahul Chaturvedi, BS
School of Medicine
University of California, San Diego
9500 Gilman Drive
La Jolla, CA
United States
Phone: 1 6096512690
Email: rchaturv@health.ucsd.edu

Abstract

Cybersecurity threats are estimated to cost the world US $6 trillion a year by 2021, and the number of attacks has increased five-fold after COVID-19. Although there is substantial literature on the threats technological vulnerabilities have on the health care industry, less research exists on how pandemics like COVID-19 are opportunistic for cybercriminals. This paper outlines why cyberattacks have been particularly problematic during COVID-19 and ways that health care industries can better protect patient data. The Office for Civil Rights has loosened enforcement of the Health Insurance Portability and Accountability Act, which, although useful in using new platforms like Zoom, has also loosened physical and technical safeguards to cyberattacks. This is especially problematic given that 90% of health care providers had already encountered data breaches. Companies must implement well-defined software upgrade procedures, should use secure networks like virtual local area networks, and conduct regular penetration tests of their systems. By understanding factors that make individuals, health care organizations, and employers more susceptible to cyberattacks, we can better prepare for the next pandemic.

(J Med Internet Res 2020;22(9):e23692) doi:10.2196/23692

KEYWORDS

cybersecurity; pandemic; COVID-19; SARS-CoV-2; risk; privacy; hack; patient data

As society has become increasingly technology dependent, it has also become increasingly vulnerable to cybercrime. Cybersecurity threats are expected to cost the world US $6 trillion a year by 2021, doubling from US $3 trillion dollars in 2015 [1]. This is particularly concerning for the health care industry, as cyberattacks are the leading cause of health security breaches [2]. Since 2016, the health care industry has been the victim of more cybersecurity attacks than even the financial industry [3]. Although there is substantial literature on the threats technological vulnerabilities have on the health care industry, less research exists on how pandemics like COVID-19 are opportunistic for cybercriminals. In this paper, we provide a review of the literature on cybersecurity issues surrounding health care and discuss possible solutions to mitigate data breaches.

One of the primary reasons cybercriminals thrive during pandemics is because heightened emotional states like fear make victims more susceptible to falling for scams [4]. According to the World Health Organization (WHO), the number of cyberattacks launched has increased five-fold during the COVID-19 pandemic [5]. A similar phenomenon was seen in 2005 after Hurricane Katrina, where thousands of fraudulent websites appeared soliciting fake donations and offering false government relief [6]. Cybercriminals often pretend to be credited and trusted organizations like the WHO and, therefore, exploit individual feelings of vulnerability in the uncertain times of a pandemic.

Additionally, health care organizations become prime targets during health crises. The use of telemedicine has proven vital to helping many patients during pandemics such as the COVID-19 crisis, especially as traditional in-person visits have become increasingly inaccessible. For example, New York University saw a 4330% increase in nonurgent virtual visits after the outbreak of COVID-19 [7]. The Office for Civil Rights

has loosened enforcement of the Health Insurance Portability and Accountability Act (HIPAA), which, although useful in opening up new platforms for care like Zoom, Skype, and FaceTime, has loosened physical and technical safeguards to cyberattacks [2,8]. This is especially problematic given that 90% of health care providers had already encountered data breaches in the past with these safeguards [2]. There is also a significant positive correlation between workload and the probability a health care worker will open a phishing email, which is particularly problematic in that, during pandemics, workloads can be at an all-time high [9].

Another potential problem for health care systems is the outbreak of ransom-motivated attacks. For example, the University of California, San Francisco (UCSF) was hacked by the cybercrime group "Netwalker," who demanded payment in exchange for not releasing confidential information. Out of fear of the consequences of this information's release, UCSF paid the group US $1.14 million [10]. The same group also took over the Champaign Urbana Public Health District website. Similarly, the Hollywood Presbyterian Medical Center in Los Angeles paid US $17,000 to get a decryption key to regain access to their hospital system. Although they regained access, they lost 10 days of revenue and likely took a hit to their reputation [2]. Unfortunately, however, complying with the demands of the cybercriminal may in fact be the most cost-effective solution, as a successful cyberattack costs an average of US $3.7 million to recover from [2]. Additionally, failure to comply can pose a serious threat to patient safety.

Access to patient records is a gold mine for cybercriminals, as they often contain information like date of birth, insurance and health provider information, as well as genetic and health data—information that cannot be easily altered, unlike the case of a credit card being stolen [3]. This information is particularly lucrative for hackers because a patient's health information can be sold for 10-20 times more than the amount for credit card information or even their social security number on the dark web.

Leak of this information can also compromise the physician-patient relationship. For instance, electronic medical record breaches could make patients less likely to disclose more private aspects of their medical history, which has the potential to impact their quality of care [11]. Furthermore, the longer a health care provider's network is down, the longer those health care workers lack access to information critical to a patient's care, like comorbidities, blood type, and allergies, in times of crisis [3]. The cost both financially and in terms of reputation and patient safety can cripple already strained hospital operations.

One additional avenue of attack presents itself as a result of the increase in the number of health care workers working from home during a pandemic like COVID-19. In the attempt to transition employees to a work-from-home setup as quickly as possible, many employers fail to consider the potential security threats these new setups create. For instance, in the hospital or office, employees may be using secure internal computer systems and updated computers, but at home, the same employees could be using insecure or outdated devices that are more vulnerable to attack [4]. Although many hospitals opted to use the Zoom platform because they view it as HIPAA-compliant, easy for both providers and patients to use, and cost-effective with medical videoconferencing accounts costing only US $200 a month, hacking of Zoom meetings has been a significant threat. Services like Zoom currently do not offer end-to-end encryption, making it not truly HIPAA-compliant, even though the Department of Health and Human Services Office for Civil Rights has relaxed enforcement of HIPAA's privacy rule during the COVID-19 pandemic [12].

Although the issue of how to safely administer health care during a pandemic is a complex one, it is clear that increased awareness is needed concerning the potential cyberthreats that pandemics exacerbate. Awareness of these threats can help hospitals and their employees protect themselves and their patients from these vulnerabilities. For instance, being aware that hackers develop phishing scams containing buzzwords during a pandemic, like "WHO," "vaccine," or "donation," can be an essential step in reviewing and flagging such emails, thereby tightening security by the information technology (IT) departments. One technique that can be employed is to have hospital IT departments send out fake phishing emails to their employees and to require training for those who failed to report the phishing attempt [13]. At the very least, this process can raise awareness among employees about cybersecurity concerns. Companies should also have well-defined software upgrade procedures, should use secure networks like virtual local area networks, and conduct regular penetration tests of their systems [2]. Hospitals need to more closely monitor administrative privileges, as the majority of large scale attacks began with a compromised account like that of a third-party provider, as seen in the case of the Hancock Regional Hospital in January 2018 [3]. By monitoring the log activity of user accounts and revoking account access when no longer needed, and employing techniques such as multifactor authentication, hospitals can better protect their IT infrastructure [3].

By understanding the factors that make individuals, health care organizations, and employers more susceptible to cyberattacks, we can better prepare for the next pandemic.

Acknowledgments

The authors would like to thank the reviewers and the editorial team for their time and comments.

Authors' Contributions

CMW, RC, and KC all contributed, read, and approved the first and final version of the paper.

Conflicts of Interest

None declared.

References

1. The 2020 Official Annual Cybercrime Report. Herjavic Group. 2020. URL: https://tinyurl.com/y56trmgv
2. Kruse C, Frederick B, Jacobson T, Monticone D. Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Heal Care 2017 Feb 21;25(1):1-10. [doi: 10.3233/thc-161263]
3. Argaw ST, Troncoso-Pastoriza J, Lacey D, Florin M, Calcavecchia F, Anderson D, et al. Cybersecurity of hospitals: discussing the challenges and working towards mitigating the risks. BMC Med Inform Decis Mak 2020 Jul 03;20(1):146. [doi: 10.1186/s12911-020-01161-7] [Medline: 32620167]
4. Naidoo R. A multi-level influence model of COVID-19 themed cybercrime. Eur J Inf Syst 2020 May 31;29(3):306-321. [doi: 10.1080/0960085x.2020.1771222]
5. WHO reports fivefold increase in cyber attacks, urges vigilance. World Health Organization. 2020 Apr 23. URL: https://www.who.int/news-room/detail/23-04-2020-who-reports-fivefold-increase-in-cyber-attacks-urges-vigilance
6. Hurricane Katrina fraud. FBI. URL: https://www.fbi.gov/history/famous-cases/hurricane-katrina-fraud
7. Jalali M, Landman A, Gordon W. Telemedicine, privacy, and information security in the age of COVID-19. SSRN J 2020:A. [doi: 10.2139/ssrn.3646320]
8. Wirth A. Cyberinsights: COVID-19 and what it means for cybersecurity. Biomed Instrum Technol 2020;54(3):216-219. [doi: 10.2345/0899-8205-54.3.216] [Medline: 32442003]
9. Jalali MS, Bruckes M, Westmattelmann D, Schewe G. Why employees (still) click on phishing links: investigation in hospitals. J Med Internet Res 2020 Jan 23;22(1):e16775 [FREE Full text] [doi: 10.2196/16775] [Medline: 32012071]
10. Winder D. The University Of California pays $1 million ransom following cyber attack. Forbes. 2020 Jun 29. URL: https://www.forbes.com/sites/daveywinder/2020/06/29/the-university-of-california-pays-1-million-ransom-following-cyber-attack/#38c2dce18a8f [accessed 2020-07-16]
11. Ronquillo JG, Erik Winterholler J, Cwikla K, Szymanski R, Levy C. Health IT, hacking, and cybersecurity: national trends in data breaches of protected health information. JAMIA Open 2018 Jul;1(1):15-19 [FREE Full text] [doi: 10.1093/jamiaopen/ooy019] [Medline: 31984315]
12. Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency. U.S. Department of Health & Human Services. 2020. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html
13. Gordon WJ, Wright A, Glynn R, Kadakia J, Mazzone C, Leinbach E, et al. Evaluation of a mandatory phishing training program for high-risk employees at a US healthcare system. J Am Med Inform Assoc 2019 Jun 01;26(6):547-552 [FREE Full text] [doi: 10.1093/jamia/ocz005] [Medline: 30861069]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
IT: information technology
UCSF: University of California, San Francisco
WHO: World Health Organization

Edited by G Eysenbach; submitted 19.08.20; peer-reviewed by E Tang, S Sheth, J Li; accepted 07.09.20; published 17.09.20.

Please cite as:
Williams CM, Chaturvedi R, Chakravarthy K
Cybersecurity Risks in a Pandemic
J Med Internet Res 2020;22(9):e23692
URL: http://www.jmir.org/2020/9/e23692/
doi:10.2196/23692
PMID:32897869

Tutorial

Leveraging Interdisciplinary Teams to Develop and Implement Secure Websites for Behavioral Research: Applied Tutorial

Christie L Martin[1*], MN, MPH, RN, PHN, LHIT-HP; Eydie N Kramer-Kostecka[2*], PhD; Jennifer A Linde[3], PhD; Sarah Friend[1], MPH, RD; Vanessa R Zuroski[4], MEd; Jayne A Fulkerson[1], PhD

[1]School of Nursing, University of Minnesota, Minneapolis, MN, United States

[2]School of Kinesiology, University of Minnesota, Minneapolis, MN, United States

[3]School of Public Health, University of Minnesota, Minneapolis, MN, United States

[4]Office of Information Technology, University of Minnesota, Minneapolis, MN, United States

[*]these authors contributed equally

Corresponding Author:
Christie L Martin, MN, MPH, RN, PHN, LHIT-HP
School of Nursing
University of Minnesota
5-140 Weaver-Densford Hall
308 Harvard Street SE
Minneapolis, MN, 55455
United States
Phone: 1 612 624 9600
Email: mart1026@umn.edu

Abstract

Behavioral researchers are increasingly using interactive digital platforms, either as standalone or supplementary intervention tools, to facilitate positive changes in research participants' health habits. Research-oriented interactive websites optimally offer a variety of participatory mediums, such as blogs, user-driven content, or health activities. Owing to the multidirectional features of interactive websites, and a corresponding need to protect research participants' identity and data, it is paramount that researchers design ethical platforms that ensure privacy and minimize loss of anonymity and confidentiality. Authentication (ie, digital verification of one's identity) of interactive sites is one viable solution to these concerns. Although previous publications have addressed ethical requirements related to authenticated platforms, few applied guidelines in the literature facilitate adherence to ethical principles and legally compliant study protocols during all phases of research website creation (feasibility, design, implementation, and maintenance). Notably, to remain compliant with ethical standards and study protocols, behavioral researchers must collaborate with interdisciplinary teams to ensure that the authenticated site remains secure and usable in all stages of the project. In this tutorial, we present a case study conducted at a large research university. Through iterative and practical recommendations, we detail lessons learned from collaborations with the Institutional Review Board, legal experts, and information technology teams. Although the intricacies of our applied tutorial may require adaptations based on each institution's technological capacity, we are confident that the core takeaways are universal and thus useful to behavioral researchers creating ethically responsible and compliant interactive websites.

(J Med Internet Res 2020;22(9):e19217) doi:10.2196/19217

KEYWORDS

research ethics and compliance; website development; behavioral research; digital interventions; website authentication; website security

Background

In recent years, health behavior researchers have begun to use digital health promotion tools (eg, websites and mobile applications) as intervention platforms or supplemental tools due to the cost-effectiveness and availability of technology [1,2]. Many researchers are utilizing interactive web technologies with user-driven content and bidirectional participatory design features (eg, discussion boards, blogs, or gaming) to increase access to socially supportive digital systems that allow participants to network and engage with other participants online [3-7]. The interactive and collaborative

design elements of these sites have the potential to enhance socially supportive environments aimed at facilitating healthy behavior change. However, the multidirectional features supported by interactive sites also introduce a level of inherent risk to the research participant: loss of anonymity and confidentiality [5-16].

In the United States (US), research institutions that receive government funding for human subjects research must abide by the tenets of the Common Rule, which are ethical standards upheld by Institutional Review Boards (IRBs) [17,18]. Behavioral researchers must submit protocols to be reviewed by the IRB to ensure that research projects are compliant with federal and institutional ethical rules and regulations. These rules and regulations include such things as informing participants of research risks and benefits, ensuring voluntary consent to participation, documenting all study protocols, and safeguarding participant data. Behavioral researchers who utilize electronic tools and mediums (eg, electronic health records) must also abide by Health Insurance Portability and Accountability Act (HIPAA) guidelines—information privacy standards set forth by the US Department of Health and Human Services—to protect identifiable information. Adhering to these standards helps to minimize the threat of disclosing protected health information (PHI; anything from social security numbers to medical record numbers) or personally identifiable information (PII; non-health specific personal identifiers, like phone numbers, license plate numbers, pictures, email addresses, etc) [10,12,19-21]. To conduct ethically responsible and legally compliant behavioral research aimed at protecting human subjects, investigators often interface with ethics (ie, IRB) and compliance (ie, legal and compliance officers) support teams. Adhering to ethics and compliance standards are particularly important for behavioral researchers utilizing interactive websites as health promotion tools, as these researchers must carefully consider how to simultaneously limit open access to website materials and protect research participants' PHI/PII.

A viable solution to protecting participants' PHI/PII before they engage with research websites is to create authenticated logins. To facilitate this procedure, behavioral researchers can collaborate with information technology (IT) support teams. In human subjects research, authentication—digital verification of one's identity—involves the use of a de-identified username and password. Unfortunately, however, best practice ethics and compliance guidelines that clearly outline practical recommendations for when or how to implement digital authentication processes are currently lacking in the literature. In 2019, the Association of Internet Researchers' ethics committee released the *Internet Research: Ethical Guidelines 3.0*. These guidelines highlight the importance of utilizing closed-access digital platforms to enhance security, yet their overarching considerations for protecting research participants' anonymity and privacy are quite broad and do not include authentication specifications [22,23]. Ideally, institution-level guidelines would include tutorials for researchers in every stage of project development. These tutorials often do not exist, yet they can be challenging to locate and are not consolidated if available. A previous search of the literature illustrates that there are only a few academic institutions and professional associations that provide written conceptual frameworks for ethically responsible and compliant internet-based research [6]. Further, these frameworks are not applied in scientific settings and do not articulate exactly how to create an ethical and compliant authenticated website [6].

Purpose

This tutorial provides a case study describing the development of our research team's interactive health promotion website to address the gap in practical guidelines. Below we detail lessons learned through a solution-focused framework, complete with specific examples from our institution and our research study *denoted in italics*. Our research team, comprised solely of behavioral researchers, set out to create an authenticated website with a blog component to engage research participants assigned to the intervention arm of a randomized controlled trial. We intended to enhance our primary in-person intervention activities by providing an opportunity for research participants to blog about their experiences and engage with others in the intervention arm of the trial by posting informal text entries or uploading pictures and videos. This opportunity needed to be limited to only intervention participants to prevent the threat of cross-contamination (ie, access to intervention content by control participants). It was necessary to develop an authenticated website to create a secure platform that prohibited posts from being visible to participants in the control group or the general public.

Due to the limited technical expertise of our research team (ie, our core research team did not include IT nor data privacy experts), it was critical to get assistance to ensure compliance with IRB and legal recommendations at every stage of website development. Notably, during the feasibility and design phases, our team was presented with a variety of unanticipated security and privacy challenges. As a result, our workflow was dynamic and flexible throughout all phases (feasibility, design, implementation, and maintenance) of the project; we made iterative adjustments to ensure that our activities were both ethical and compliant. We present information in our tutorial to provide behavioral researchers with a roadmap for unforeseen issues that may arise when creating and disseminating a secure research website. Of importance, we discovered that leveraging cross-sector collaboration with an interdisciplinary team was a crucial strategy to adhere to all ethics and compliance standards relevant to our authenticated platform.

Although online resources for creating websites do exist at our large research-based university, specific instructions for authentication and institutional data privacy and security policies are not readily available. We discovered that these instructions also vary by department. For example, at our institution, some departments or schools that conduct research involving community members—rather than with patients in clinical settings or de-identified electronic health record data—may or may not be required to follow the same ethical guidelines for managing PHI/PII that are prescribed by HIPAA. To navigate our institution's security and privacy rules and regulations, we needed to engage with IRB and legal experts in the formative stages before engaging with IT. The IT team then provided tools

and expertise to assist us in building a secure website that met research ethics and compliance standards. Initial conversations led to several months of interdisciplinary collaborations that were vital to understanding specific safeguards for protecting PHI/PII to execute our website project successfully. The goal of this tutorial is to share insights gleaned from these interdisciplinary collaborations. Through our case study, we provide specific considerations and recommendations unique to each phase of our project: feasibility, design, implementation, and maintenance. Lessons learned will be useful to health behavior researchers interested in creating ethically responsible and compliant authenticated websites. Research ethics and compliance guidelines differ by country and, thus, researchers outside of the US may need to amend the following tutorial. However, our tutorial highlights key issues of consideration that apply to website development in any behavioral health research context.

Feasibility Phase

Determine Your Website Purpose & Target Audience: Is Authentication Warranted?

The first and most important consideration when conceptualizing a website for your research project is to determine whether or not electronic authentication is needed. It is, therefore, necessary to understand the overall purpose of your website within the context of your research project's goals. As a general rule, websites that utilize interactive features require more stringent ethical stipulations than websites that provide static site content (ie, "read-only" materials). Authentication is warranted if the goal is to engage research participants in activities that would be covered by institutional confidentiality protections, such as using a social component (eg, blogging, gaming) for a research intervention. Authentication is also warranted to ensure confidentiality and data security if the goal is to observe or collect data from participants rather than provide them with resources. If it is not immediately apparent to your team that authentication is warranted, you must then consider how the website will engage your target audience and determine whether or not the website content should be accessible to the public. Authentication is warranted if your website includes an interactive component and has the potential to display PHI/PII. *This was the case with our website's interactive blog that allowed research participants to post text and/or pictures, as text could potentially contain identifiable content, and images could include visually identifiable features.* Notably, authentication is also recommended for static websites if the program is accessible to certain research participants at different times, *as was the case with our research design that provided website access to different cohorts at varying time points.*

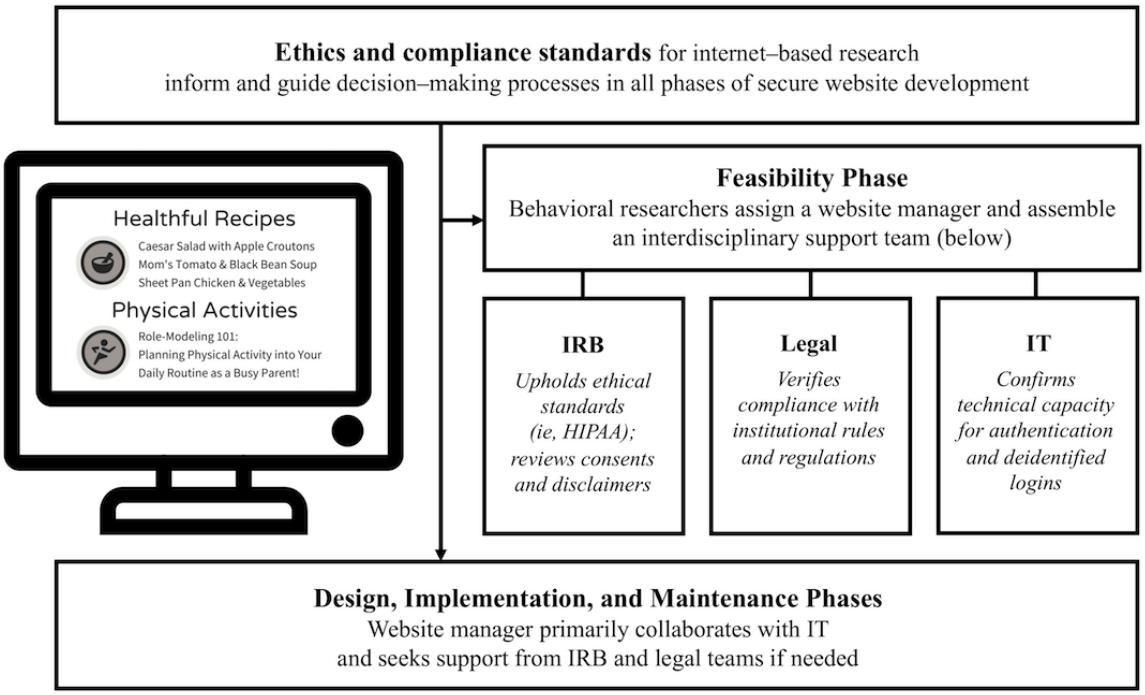
Determine Your Institution's Technical Capacity for Authentication: Is Authentication Feasible?

The next consideration when conceptualizing your website is to work with IT professionals to assess your institution's technical capacity for authentication and to determine whether ethical and compliant website creation is feasible. First, you will need to gain access to an appropriate web content management system (CMS), like Drupal or WordPress, that you can integrate with authentication systems [24]. *At our institution, Drupal was the only CMS that could be authenticated.* Once you have confirmed that authentication is feasible, you will need to verify if your institution has the technical capacity to create de-identified usernames, which preserves your research participants' identity and ensures that data privacy is maintained during the authentication process.

Determine Your Interdisciplinary Team: Is Authentication Supported?

The final consideration when conceptualizing your website is to determine your institution's ability to support the project throughout all stages of development, from feasibility and design to maintenance and completion. It is crucial to collaborate with an interdisciplinary team from the outset to mitigate any issues that may arise. Support personnel should include a website manager, as well as ethics (ie, IRB), compliance (ie, legal), and technology (ie, IT) experts. *At our university, online tutorials or written guidelines for ethical and compliant website development did not exist, so we reached out to the IRB and legal teams early in the process after reaching out to IT to confirm technical capacity.* See Figure 1 for a workflow representing interdisciplinary collaboration.

Figure 1. Workflow representing interdisciplinary collaboration required for secure website development in a university research setting. IRB: Institutional Review Board; IT: information technology.



Website Manager

After you have evaluated the needs of your target audience and technical capacity for authentication, you should appoint a website manager(s). *For example, we chose someone with previous website design and management experience to fill this role.* The website manager(s) should then seek out existing website regulations and ethics and compliance guidelines and become familiar with their institution's current technical safeguards to protect privacy and anonymity. This person or team will be responsible for understanding authentication procedures as well as determining the content, layout, aesthetics, and navigation tools to be utilized during the initial design phase. All decisions should be made in tandem with recommendations from IT experts, as they are most knowledgeable about your institution's authentication processes. *In our case, these individuals also guided additional institutional requirements (ie, mandatory branding, simultaneously web and mobile-friendly interfaces, accessibility features for those with visual impairments) and other web-specific tips (ie, reusable HTML scripts) to help expedite the development process.*

Institutional Review Board

The IRB can guide you through the creation of informed consent documents to make sure your website upholds federal and institutional ethical standards to protect human subjects. If your website is not in and of itself an intervention but rather supplements your research, you may want to consider "opt-in/opt-out" disclaimers that allow research participants to choose whether or not they receive access to the website. Regardless of the consent processes that you utilize, make sure to outline the risks and benefits of engaging with your website in your consent form; the presence of this information will be verified and approved by the IRB before study initiation as part of standardized informed consent procedures. *For example, we made it clear to our research participants that to minimize perceived risks, they could access the website and download materials with a de-identified login without having to participate in the interactive blog component.*

Legal Experts

In addition to the necessary contact with the IRB before starting human subjects research, it is essential to seek guidance from legal experts to align with institutional rules and regulations. Legal experts can suggest the appropriate informed consent processes to protect your research participants' anonymity and can suggest additional risk reduction strategies that you should outline in research protocols to align with best practice. Additionally, it is essential to seek advice from research compliance teams that understand institution-specific rules and regulations. *For example, we specifically reached out to internet security analysts, health information compliance officers, and lawyers familiar with HIPAA compliance. At our institution, health information compliance officers are affiliated with the Research Compliance Office, which oversees university-wide ethical and regulatory standards.*

Information Technology Experts

During the feasibility phase, your IT team can inform you of existing data security processes and tools to protect PHI/PII. *For example, we specifically reached out to expert website developers within the Office of Information Technology to better understand the available IT tools.* All technical security measures taken must be in alignment with institutional ethics and compliance standards, as these may differ depending on location [6,7,21,25-27]. After IT specialists have confirmed technical capacity and your IRB and legal teams have laid the foundation for creating an ethical and compliant website, you

will then primarily seek guidance from IT specialists during subsequent website development phases. Due to your close partnership with IT, we also recommend that you review ethical recommendations created by the Association of Internet Researchers [23] in conjunction with IT personnel before authenticating or launching your website.

Design Phase

Understand Existing Authentication Processes

Once you have determined that your website project is indeed feasible, and you have assigned your primary contact(s) for website development processes, you will need to carefully consider how to uphold ethics and compliance standards during the design phase. First, it is crucial to understand the details of existing authentication procedures, which make your website content accessible only to those with preassigned de-identified usernames and passwords. *Of note, we were interested in assigning usernames and passwords to intervention participants, cohort by cohort over time.* More and more institutions are utilizing two-factor authentication (ie, Duo), whereby a one-time password is sent to a second electronic device as another method of verification [28]. Two-factor authentication is an additional level of security used to prevent identity theft and other online fraud (eg, phishing and malware) [9,29]. This added security is particularly crucial for behavioral researchers who have access to PHI/PII and need to protect it, yet the extra security complicates the login process. *As a result, our study staff engaged with study participants by 1) notifying them that they would receive a predetermined username and password via email, 2) emailing them detailed tutorials with images and text describing how to change their passwords (for security purposes) and to log into the website, and 3) allotting additional time and resources during in-person intervention sessions for technical troubleshooting.*

Select the Correct Web Content Management System

When selecting a CMS, it is important to note that not all CMSs are created equally in terms of capacity to meet ethics and compliance standards for human subjects research. Therefore, selecting the correct web CMS is crucial. *For example, at our institution, four CMSs were available at the institutional level. Two were part of the G Suite for Education (ie, Google Blogger and Google Sites), and two were different university-managed configurations of the CMS Drupal (ie, Drupal Lite and Drupal Enterprise). At the time of web development, Drupal Enterprise was the only CMS that allowed for authentication. While multiple systems were available at an institutional level, our department had its own IT team and communications team available for support, so we needed to interface with several individuals to select the system that met ethics and compliance standards—Drupal Enterprise.* In addition to confirming the capacity for password-protected logins, we recommend that you confirm with the support team that your selected CMS will enable you to review and approve, or reject, all content posted by research participants before being visible to others. This extra security measure may prevent inappropriate exposure of PHI/PII to participants and hackers alike.

Choose a Suitable Website Host

Once you have determined the appropriate CMS, you will have to decide which department within your larger institution can host your authenticated website. *For our research team, the academic department with whom we were associated could not host authenticated websites for research, so we reached out to our administrative umbrella organization that oversees multiple academic departments within the university. The Drupal support team affiliated with this organization agreed to host our website.* Of note, once you have selected an adequate host, it is vital to determine the type and breadth of support you will receive from your host, as the turnaround time will directly impact the length of the development phase. If assistance is not timely or adequate, you will need to consider this when determining the appropriate team member to manage your website. *In our case, the Drupal support team provided prompt technical assistance throughout our project.*

Select the Appropriate Internet Account

To create de-identified logins for your research participants, you will need to set up individual internet accounts. Similar to CMS and website hosts, internet account types vary considerably. *For example, at our institution, there are a variety of internet accounts (ie, sponsored accounts; proxy accounts; departmental/organizational accounts), each with differing levels of access and stipulations for expiration.* To determine which account is best, you must first consider whether your research participants already have an affiliation with your organization or whether they will need a separate account that is HIPAA compliant. *For example, guest accounts at our institution typically use the first part of a participant's email, which potentially could contain PHI/PII and thus should not be used.* You will then need to determine how long your participants will need to access your website and make sure that the account can be discontinued at an appropriate time. Next, you should consider the level of access needed. *Most internet accounts at our institution are created with access to other services, such as the G Suite, which provides access to a variety of Google Cloud computing tools like Gmail, Docs, Drive, and Calendar. For our project, we only needed to create usernames for our research participants, not an additional email account, so departmental accounts were best suited for our website project because we could manually "opt-out" of G Suite. These usernames were then linked to Drupal Enterprise on the backend, granting login access to each user.* Of note, regardless of the type of internet accounts you ultimately choose, it is essential to avoid using identifiable information, as it could be visible to your IT help staff if and when they assist you with creating or changing usernames and passwords.

Implementation Phase

Develop and Beta Test

After designing your website, you will need to develop and beta test it before granting access to your research participants. You should test the site and the desired features to ensure that they meet your project's goals. *For example, during this phase, we tested the authentication process with temporary usernames and made sure that all website content was accessible from*

multiple devices (eg, desktops, laptops, tablets, mobile phones). We also tested all hyperlinks and made sure that the blog posts were only visible to research participants once our web manager approved them.

Create Internet Accounts/Deidentified Logins

After beta testing and documenting participant consent, choose the appropriate internet account type that will allow you to create individualized, de-identified logins for all participants. Then set a reasonable expiration date (eg, the length of the intervention) and make sure to link the de-identified usernames to your CMS on the backend. To facilitate the process, you may want to consider a one-time password reset for all newly created accounts before sharing the account information with research participants. *This extra step worked well for us, given that our institution required that all initial passwords be changed within 24 hours; a second password reset was completed at the research participants' leisure. For our website project, we created pseudonyms for our research participants and stored all internet account information on a secure server, accessible only to a limited subset of our research team who had IRB approval.*

Maintenance Phase

Maintain Your Website

Once you have gone live with your site, make sure you visit the site regularly. *In our case, we developed separate pages of content in alignment with our monthly in-person sessions because we wanted research participants to access the online material on a month-to-month basis. Therefore, regular biweekly website maintenance was crucial to align with our research project's overall goals.* Indeed, ensuring that your website is properly maintained is not only necessary for overall functionality but also to secure the site from unintended breaches of confidentiality. *Frequent monitoring was especially crucial for our website project, given that our platform did not contain built-in alerts when someone posted on the blog. Although we approved all posts before they went live, unintended security breaches could have occurred if we did not conduct regular maintenance.*

Project Completion

Upon project completion, you will need to complete three critical steps to ensure data privacy. First, confirm the expiration of the de-identified logins. *At our institution, the IT department sent an email notifying the website manager of upcoming internet account expirations*. Second, at the end of the intervention, double-check to make sure all de-identified logins are disconnected from your CMS. These precautions prevent users from accessing the site after completion of the intervention. *Of note, we were able to disconnect users from Drupal without deleting blog posts, which served useful when analyzing usage after completion*. Finally, make sure that all participant information associated with de-identified usernames is stored on a secure server and only available to approved research personnel for a predetermined length of time denoted in your research protocol. *Per our IRB and grant funder's rules and regulations, we were instructed to keep digital copies of all participant data on a secure server for seven years after completion of the randomized controlled trial.* These three steps are vital to adhere to ethics and compliance standards, particularly if your compliance office ever audits you.

Discussion

Takeaways for Authenticated Website Development and Implementation

Despite the increasing trend of developing digital platforms to deliver or support health behavior research, practical or applied ethics and compliance guidelines for research-based website development are lacking in the literature, and institution-specific tutorials are often not readily available. As depicted in our case study, designing and implementing an authenticated digital platform that aligns with best practices for human subjects research requires time and sufficient resources. Therefore, we present this case study protocol with suggested guidelines as a roadmap for research teams. Although we linearly present our considerations, it is essential to note that website development is an iterative process, and you may need to adjust your timelines throughout all stages of design and implementation. We also highly recommend that you consider the above recommendations well in advance of going live with your website, and that you collaborate closely with IRB, legal, and IT experts early in the process. These early partnerships are vital to the success of creating an authenticated website that abides by ethics and compliance standards.

In summary, lessons learned from our applied case study highlight three overarching considerations future researchers must be aware of before launching health behavior research-based authenticated websites. First, behavioral researchers must be mindful that when they ask participants to be cocreators or sharers of intervention materials through blogging or uploading content to the website, it is their responsibility to protect participants' anonymity and privacy through the provision of an authenticated website. Second, when striving to create authenticated websites that align with ethics and compliance standards, behavioral researchers must expect the unexpected. Due to unanticipated challenges, we needed extra time to learn the ins and outs of ethically responsible and compliant website authentication outlined above. While our paper has filled a gap in the literature by providing researchers with an applied tutorial for developing an authenticated website, we acknowledge the novelty of the field of digital health research and know that our proposed best practice recommendations will likely morph with changes in technology-based systems. We also acknowledge that authentication processes, technological capacities, and availability of ethics (ie, IRB), compliance (ie, legal), and technical (ie, IT) support teams may differ from institution to institution, so we present these recommendations as an initial template for best practices. Third, behavioral researchers should note that although programming considerations and design phases may differ by project and institution, the foundation of all website-oriented decision making must be grounded in ethical principles. Thus, to conduct ethical and compliant research, interdisciplinary teams must rigidly prioritize the protection of

participants' privacy and confidentiality in all phases of website design and implementation.

Future Considerations

This paper provides behavioral researchers with an applied tutorial for secure website development and implementation. Project takeaways and lessons learned from our interdisciplinary collaborations may assist future researchers in designing secure websites that could be used as a standalone intervention or as an intervention supplement. We recommend that colleagues consider developing individualized protocols for their respective universities or institutions since there are few published evidence-based guidelines for ethical and compliant website development in behavioral research. Moreover, authentication processes and technologies differ from institution to institution, and researchers must consider this variance. Publishing applied tutorials or guidelines will advance the science of ethically responsible and compliant internet-based behavioral research, as the dissemination of institutional guidelines will provide opportunities for knowledge to be shared across sectors.

Furthermore, we acknowledge that although ensuring security is vital to technology-based behavioral research, researchers may not always be the most knowledgeable when it comes to creating secure and authenticated websites. Behavioral researchers should leverage interdisciplinary collaborations to be compliant with ethical standards and institutional rules and regulations. To ensure project success, we promote interdisciplinary team science and encourage behavioral researchers to engage in a comprehensive dialogue with IRB, legal, and IT professionals throughout all phases of website development.

Acknowledgments

This project was supported by Award Number 1R01HL123699 from the National Heart, Lung, and Blood Institute, PI: Jayne A. Fulkerson, PhD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health (NIH). This study is registered with NIH ClinicalTrials.gov: NCT02973815.

Authors' Contributions

CM was the website designer and manager. VZ was the principal information technology support contact. CM and EKK wrote the manuscript. JL, SF, VZ, and JF reviewed and edited the tutorial. CM, EKK, and JF were primarily responsible for the final content. All coauthors read and approved the submitted tutorial.

Conflicts of Interest

None declared.

References

1. Ellis-Barton C. Ethical considerations in research participation virality. J Empir Res Hum Res Ethics 2016 Jul;11(3):281-285. [doi: 10.1177/1556264616661632] [Medline: 27534590]
2. Possemato K, Acosta MC, Fuentes J, Lantinga LJ, Marsch LA, Maisto SA, et al. A web-based self-management program for recent combat veterans with PTSD and substance misuse: Program development and veteran feedback. Cogn Behav Pract 2015 Aug 01;22(3):345-358 [FREE Full text] [doi: 10.1016/j.cbpra.2014.03.005] [Medline: 26120269]
3. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: Reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
4. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
5. Zaidan AA, Zaidan BB, Kadhem Z, Larbani M, Lakulu MB, Hashim M. Challenges, alternatives, and paths to sustainability: Better public health promotion using social networking pages as key tools. J Med Syst 2015 Feb;39(2):7. [doi: 10.1007/s10916-015-0201-y] [Medline: 25631841]
6. Hokke S, Hackworth NJ, Quin N, Bennetts SK, Win HY, Nicholson JM, et al. Ethical issues in using the internet to engage participants in family and child research: A scoping review. PLOS One 2018 Sep 27;13(9):e0204572 [FREE Full text] [doi: 10.1371/journal.pone.0204572] [Medline: 30261041]
7. Mahon PY. Internet research and ethics: Transformative issues in nursing education research. J Prof Nurs 2014 Mar;30(2):124-129. [doi: 10.1016/j.profnurs.2013.06.007] [Medline: 24720940]
8. Montgomery KC, Chester J, Milosevic T. Children's privacy in the big data era: Research opportunities. Pediatrics 2017 Nov;140(Suppl 2):S117-S121 [FREE Full text] [doi: 10.1542/peds.2016-1758O] [Medline: 29093045]
9. Pieters W, Hadžiosmanović D, Dechesne F. Security-by-experiment: Lessons from responsible deployment in cyberspace. Sci Eng Ethics 2016 Jun;22(3):831-850 [FREE Full text] [doi: 10.1007/s11948-015-9648-y] [Medline: 25896029]
10. Lamberg L. Confidentiality and privacy of electronic medical records: Psychiatrists explore risks of the information age. JAMA 2001 Jun 27;285(24):3075-3076. [doi: 10.1001/jama.285.24.3075] [Medline: 11427118]

11. Elovici Y, Fire M, Herzberg A, Shulman H. Ethical considerations when employing fake identities in online social networks for research. Sci Eng Ethics 2014 Dec;20(4):1027-1043. [doi: 10.1007/s11948-013-9473-0] [Medline: 24218141]
12. Ellett MLC, Lane L, Keffer J. Ethical and legal issues of conducting nursing research via the internet. J Prof Nurs 2004 Jan;20(1):68-74. [doi: 10.1016/j.profnurs.2003.12.005] [Medline: 15011195]
13. Demiris G. Consumer health informatics: Past, present, and future of a rapidly evolving domain. Yearb Med Inform 2016 May 20;25(S 01):S42-S47 [FREE Full text] [doi: 10.15265/IYS-2016-s005] [Medline: 27199196]
14. Bakken S, Marden S, Arteaga SS, Grossman L, Keselman A, Le P, et al. Behavioral interventions using consumer information technology as tools to advance health equity. Am J Public Health 2019 Jan;109(S1):S79-S85. [doi: 10.2105/AJPH.2018.304646] [Medline: 30699018]
15. Azer SA. Social media channels in health care research and rising ethical issues. AMA J Ethics 2017 Nov 01;19(11):1061-1069 [FREE Full text] [doi: 10.1001/journalofethics.2017.19.11.peer1-1711] [Medline: 29168677]
16. Pittenger DJ. Internet research: An opportunity to revisit classic ethical problems in behavioral research. Ethics Behav 2003;13(1):45-60. [doi: 10.1207/S15327019EB1301_08] [Medline: 14552304]
17. DeRenzo EG, Moss J, Singer EA. Implications of the Revised Common Rule for human participant research. Chest 2019 Feb;155(2):272-278 [FREE Full text] [doi: 10.1016/j.chest.2018.09.022] [Medline: 30312589]
18. Office for Human Research Protections. Federal policy for the protection of human subjects ('Common Rule'). 2016 Mar 18. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html [accessed 2020-09-12]
19. Centers for Disease Control and Prevention (CDC). HIPAA privacy rule and public health. Guidance from CDC and the U.S. Department of Health and Human Services. MMWR Suppl 2003 May 02;52:1-17, 19 [FREE Full text] [Medline: 12741579]
20. Demiris G. The diffusion of virtual communities in health care: Concepts and challenges. Patient Educ Couns 2006 Aug;62(2):178-188. [doi: 10.1016/j.pec.2005.10.003] [Medline: 16406472]
21. Nebeker C, Harlow J, Espinoza GR, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth 2017 Oct;8(4):266-276. [doi: 10.1080/23294515.2017.1403980] [Medline: 29125425]
22. Franzke AS, Bechmann A, Zimmer M, Ess C, the Association of Internet Researchers. Internet research: Ethical guidelines 3.0 Association of Internet Researchers. Association of Internet Researchers. 2020. URL: https://aoir.org/reports/ethics3.pdf [accessed 2020-09-12]
23. Ethics. Association of Internet Researchers. 2020. URL: https://aoir.org/ethics/ [accessed 2020-09-11]
24. Radulovski I. Enterprise Content Management System (DE-CMS). Drupal. 2017 Jul 30. URL: https://www.drupal.org/project/enterprise_cms [accessed 2020-09-12]
25. Martensen R. Institutional review boards, professionalism, and the internet. Sci Transl Med 2010 May 05;2(30):30cm15. [doi: 10.1126/scitranslmed.3000884] [Medline: 20445198]
26. Walther JB. Research ethics in Internet-enabled research: Human subjects issues and methodological myopia. Ethics Inf Technol 2002;4:205-216. [doi: 10.1023/a:1021368426115] [Medline: 15977362]
27. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
28. Two-Factor Authentication (2FA) from Duo. Duo Security. 2020. URL: https://duo.com/product/multi-factor-authentication-mfa/two-factor-authentication-2fa [accessed 2020-09-12]
29. Perkel J. Cybersecurity: How safe are your data? Nature 2010 Apr 29;464(7293):1260-1261. [doi: 10.1038/4641260a] [Medline: 20428136]

Abbreviations

CMS: Content Management System
HIPAA: Health Insurance Portability and Accountability Act
IRB: Institutional Review Board
IT: Information Technology
PHI: Personal Health Information
PII: Personally Identifiable Information

Edited by G Eysenbach; submitted 08.04.20; peer-reviewed by E Buchanan, S Hokke, Z Koporc; comments to author 12.06.20; revised version received 29.07.20; accepted 13.08.20; published 23.09.20.

Please cite as:
Martin CL, Kramer-Kostecka EN, Linde JA, Friend S, Zuroski VR, Fulkerson JA
Leveraging Interdisciplinary Teams to Develop and Implement Secure Websites for Behavioral Research: Applied Tutorial
J Med Internet Res 2020;22(9):e19217
URL: http://www.jmir.org/2020/9/e19217/
doi:10.2196/19217
PMID:32965234

Original Paper

Engagement With Motivational Interviewing and Cognitive Behavioral Therapy Components of a Web-Based Alcohol Intervention, Elicitation of Change Talk and Sustain Talk, and Impact on Drinking Outcomes: Secondary Data Analysis

Ajla Mujcic[1,2], MSc; Stuart Linke[3,4], PhD; Fiona Hamilton[4], PhD; Alexandria Phillips[5], MSc; Zarnie Khadjesari[6], PhD

[1]Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands

[2]Trimbos Institute, The Netherlands National Institute of Mental Health and Addiction, Utrecht, Netherlands

[3]Camden and Islington Mental Health Trust, London, United Kingdom

[4]eHealth Unit, Research Department of Primary Care and Population Health, University College London, London, United Kingdom

[5]Department of Psychology, King's College London, London, United Kingdom

[6]School of Health Sciences, University of East Anglia, Norwich Research Park, Norwich, United Kingdom

Corresponding Author:
Ajla Mujcic, MSc
Erasmus School of Social and Behavioural Sciences
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Rotterdam
Netherlands
Phone: 31 030 29 59 256
Email: amujcic@trimbos.nl

Abstract

Background: Down Your Drink (DYD) is a widely used unguided web-based alcohol moderation program for the general public based on cognitive behavioral therapy (CBT) and motivational interviewing (MI); it provides users with many opportunities to enter free-text responses.

Objective: The aim of this study was to assess participants' use of key CBT and MI components, the presence of change talk and sustain talk within their responses, and whether these data are associated with drinking outcomes after 3 months.

Methods: An exploratory secondary data analysis was conducted on data collected in 2008 from the definitive randomized trial of DYD (N=503). Past week alcohol use at baseline and 3-month follow-up were measured with the TOT-AL. Covariates included baseline alcohol use, age, gender, education level, and word count of the responses. Use of MI and CBT components and presence of change talk and sustain talk were coded by two independent coders (Cohen κ range 0.91-1). Linear model regressions on the subsample of active users (n=410) are presented along with a negative binomial regression.

Results: The most commonly used component was the listing of pros and cons of drinking. The number of listed high-risk situations was associated with lower alcohol use at 3-month follow-up (B_{adj} −2.15, 95% CI −3.92 to −0.38, *P*=.02). Findings on the effects of the percentage of change talk and the number of listed strategies to deal with high-risk situations were inconsistent.

Conclusions: An unguided web-based alcohol moderation program can elicit change talk and sustain talk. This secondary analysis suggests that the number of listed high-risk situations can predict alcohol use at 3-month follow-up. Other components show inconsistent findings and should be studied further.

(J Med Internet Res 2020;22(9):e17285) doi:10.2196/17285

KEYWORDS

eHealth; digital health; self-management; alcohol use; motivational interviewing; cognitive behavioural therapy; engagement

Introduction

Interventions aimed at reducing risky alcohol use are diverse and vary in many ways, including their mode of delivery (eg, in-person, bibliotherapy, digital), theoretical approach, and length (ranging from ultrabrief to extended interventions). This variation is also reflected in digital alcohol interventions [1]. Ultrabrief digital alcohol interventions are usually limited to self-monitoring exercises and personalized feedback (eg, decisional balance feedback [2]). Brief interventions can consist of self-monitoring exercises combined with personalized feedback and additional modules such as identification of high-risk situations, which help reduce alcohol use in specific situations (eg, the web-based personalized feedback program Drinktest [3]). On the other hand, extended interventions offer a digital form of intensive treatment, including multiple sessions (eg, the self-help alcohol intervention Balance [4]).

Several systematic reviews have shown that web-based alcohol interventions can be effective at reducing alcohol use in adult populations, finding small and moderate effect sizes [5-7]. Moderators of effectiveness have been studied and include length of intervention [1], level of guidance, setting, and integrated therapeutic principles [5]. Multiple studies have examined ways to increase engagement with alcohol and other health behavior interventions. Although these studies have conceptualized engagement in different ways, such as the received dose, adherence, degree of involvement over a longer period of time, or process of linked behaviors [8], engagement is mostly linked to frequency and length of use or to the use of specific components. However, few studies have examined users' engagement with specific components of web-based alcohol interventions in detail [1].

Furthermore, although many alcohol interventions are partly based on motivational interviewing (MI) [1], few studies have explored the presence of change talk and sustain talk, which are important components of this "collaborative, goal-oriented style of communication with particular attention to the language of change, designed to strengthen personal motivation for and commitment to a specific goal by eliciting and exploring the person's own reasons for change within an atmosphere of acceptance and compassion [9]." Change talk can be defined as language referring to movement toward change of a target behavior, including verbalizations of consideration, motivation, or commitment to change. Sustain talk is language referring to movement toward sustaining the target behavior and the status quo. In a meta-analysis from 2018 on MI processes, including 21 studies on alcohol use that mostly involved face-to-face treatment, it was found that a higher proportion of change talk was associated with reductions in risk behaviors [10]. One recent study looked at the presence of change talk and sustain talk in a web-based intervention; however, the intervention targeted physical activity, not alcohol use [11]. To the best of our knowledge, no study has explored the presence of change talk and sustain talk in a web-based alcohol intervention.

The Down Your Drink (DYD) website is a web-based alcohol intervention aimed at the general population [12]. DYD has some distinct features. First, it is targeted at opportunistic electronic help (e-help)–seekers who are not enrolled in an alcohol treatment pathway and who differ from dependent help seekers [13]. Another important feature of DYD is that it is one of the first web-based extended alcohol interventions based on MI techniques and cognitive behavioral therapy (CBT). Other examples are web-based self-help alcohol interventions from the Netherlands [14] and Norway [4]. DYD attempted to translate components of usual face-to-face treatment to a web-based, unguided setting and encourages self-monitoring of drinking behavior in the Drinking Episode Diary. Lastly, it offers many opportunities for free-text responses, which provides an opportunity to study the way web-based alcohol interventions are understood and used.

The original trial reported descriptive data on engagement with DYD, namely number of logins and number of pages viewed [15]. However, user data, that is, actual responses provided by participants to MI questions designed to strengthen personal motivation for and commitment to a specific goal, were not analyzed. Understanding how DYD was used and whether it encouraged people to think about changing their drinking is imperative to optimize future web-based alcohol programs. With this secondary analysis, we aim to answer the following research questions:

1. Does DYD elicit change talk and sustain talk?
2. Are the MI and CBT components of DYD used actively (ie, users responded to questions related to the component at least once)?
3. Do i) user responses indicating change talk and sustain talk and ii) users making use of the separate MI and CBT components have an impact on change in alcohol consumption (intervention effectiveness)?

Methods

Design

This paper reports on an exploratory secondary data analysis conducted on data from the definitive randomized trial of DYD (see the next section for further details). Ethical approval for the secondary analyses conducted in this study was granted by the University College London (UCL**)** Research Ethics Committee ("Engagement with internet-based alcohol moderation intervention 'Down Your Drink'" project ID: 3770/002).

DYD Trial

Data used for this study were collected during the definitive DYD trial from October 2007 to May 2009. This two-arm randomized controlled trial aimed to assess the effectiveness of DYD in reducing alcohol consumption and alcohol-related harm at 3- and 12-month follow-up. People aged 18 years or older who had internet access, scored 5 or more on the Alcohol Use Disorders Identification Test-Consumption (AUDIT-C), understood written English, and were willing to complete follow-up questionnaires could participate. A total of 2652 adults with scores of 5 or higher on the AUDIT-C were included [16]. During the pilot and main trial extension phases, another 5238 participants were included. The primary outcome measure of alcohol consumption in the past week was collected using

the TOT-AL [17]. The entire study was conducted on the internet. A more extensive description of the main trial procedures can be found in Murray et al [18] and Wallace et al [15].

Participants

For this study, we included the DYD trial participants who were randomized to the intervention group in 2008 (the only full calendar year of inclusion during the main trial) and who reported their alcohol use at 3-month follow-up. This ensured that participants who enrolled at various time periods during the year were all included. The sample provided a sufficient yet manageable number of participants for the analysis. In 2008, a total of 2543 participants were randomized; of those, 1271 were randomized to the intervention group. After removing participants who withdrew consent during the course of the study (n=41) and those who had not reported their alcohol use at 3-month follow-up (n=727), the final sample consisted of 503 participants. In addition to this full sample, we independently analyzed the active use sample, a subsample of 410 participants who actively engaged with the program at least once, defined as those who responded to at least one of the questions within the DYD program.

Intervention

DYD is a web-based alcohol intervention that is primarily based on two evidence-based approaches that are widely used in the psychological treatment of alcohol misuse, namely MI and CBT; the latter is a goal-oriented therapeutic approach that systematically addresses dysfunctional emotions, behaviors, and thought processes. For example, CBT includes components urging participants to set a goal, recognize high-risk situations, and articulate their attitudes concerning (moderating) their drinking behavior. Self-monitoring of drinking behavior (eg, amount, type, setting, cost) is facilitated in the Drinking Episode Diary. The Alcohol Units Counter is another self-monitoring exercise; however, it does not keep track of changes in drinking over time. DYD delivers the MI approach by presenting a series of questions prompting free-text responses. The effectiveness of MI depends on how people respond to the questions, that is, whether their responses suggest that they wish to reduce their drinking (ie, change talk) or not (ie, sustain talk). For a more extensive description of the intervention, see [12].

Measures

Measures from the DYD trial included in the current analyses are past week self-reported alcohol consumption in UK units (ie, 1 unit=8 grams of ethanol) at baseline and 3-month follow-up measured using the TOT-AL (total past week alcohol consumption) [17], age (continuous variable), gender (male/female), and education level (university degree or equivalent/A Levels or equivalent)/General Certificate of Secondary Education (GCSE) or equivalent/other qualifications/no qualifications). Ethnicity and relationship status were used to describe the sample but were not included as covariates in the model. This study included measures on the use of DYD components, which will be described further in the Qualitative Analysis section. To assess the effects of change talk and sustain talk, the percentage of change talk (change talk frequency over the sum of change talk frequency plus the sustain talk frequency) was included as recommended [19]. Lastly, the total number of words entered into the program was computed for each participant.

Qualitative Analysis

The coding scheme (see Multimedia Appendix 1) included the main MI components and CBT components, and it was informed by the Client Language EAsy Rating (CLEAR) coding system [19] and items from the Revised Cognitive Therapy Scale (CTS-R) [20]. CLEAR is a coding system that can be used to assess change talk and counter change talk (ie, sustain talk) in a participant's responses. It was originally developed for in-therapy client language. MI components that were coded for included the presence of change talk, the presence of sustain talk, listing of the pros and cons of drinking alcohol, and noting of values (what is most important and meaningful to oneself). For each of these components, participants were assigned either a 1 (present) or a 0 (not present) depending on their responses. Furthermore, we coded the number of times the following were uttered: change talk, sustain talk, and pros and cons of drinking. Note that the frequency of change talk included the uttered cons of drinking and all other change talk present in response to the other questions as evaluated using CLEAR. The frequency of sustain talk included the number of mentioned pros of drinking and all other sustain talk present in response to other questions within the program.

The CTS-R is a scale for measuring therapist competence in cognitive therapy; it lists several key components of cognitive therapy, which helped us identify the main CBT components. We coded the following CBT components: setting a start date for alcohol moderation, setting a goal for alcohol moderation, completing another part of the moderation plan (eg, noting someone who might help), noting alcohol use prior to DYD, noting high-risk situations for alcohol use (eg, places, people, emotions), noting strategies to deal with high-risk situations, exploring feelings of craving, exploring relapse prevention (eg, thinking back to circumstances around a previous relapse), making a relapse plan (eg, stating who to call in case of a lapse), exploring thoughts (about drinking), monitoring alcohol use (ie, type, frequency, and amount of drinking). For all these codes, participants were assigned either a 1 (present) or a 0 (not present) depending on their responses. Furthermore, we coded the number of times the following were mentioned: high-risk situations and strategies to handle high-risk situations.

The coding was completed by two coders (AM and AP); SL was consulted on any discrepancies. Each coder coded 50% of the sample independently, and 51 (10%) of the sample were coded by both. Interrater reliability was high for both the dichotomous variables, as addressed with Cohen κ (range 0.91-1, mean 0.97), and the continuous variables, as addressed with intraclass correlation (range 0.90-0.99, mean 0.97). Continuous variables included the number of times the following were present: change talk, sustain talk, pros, cons, high-risk situations, and ways to handle high-risk situations.

Quantitative Analysis

Descriptive statistics for each DYD use variable were computed to assess the presence of change talk and sustain talk within the responses and the use of CBT and MI components. All analyses described in this study are exploratory, as they were not preregistered. Multiple hierarchical regression models and analyses were used to assess the predictive value of the different components as well as of the change talk and sustain talk on alcohol use reduction at 3-month follow-up. In accordance with CLEAR guidelines, change talk and sustain talk were entered into the model as percentage change talk [19]. In the first step, a linear model was generated containing the baseline drinking level and the following covariates: number of words entered in all responses by the participant, age, gender, and education level. This model was then compared to a linear model containing all the variables of interest to assess the added explained variance of the full model. The distribution of the alcohol consumption outcome was highly skewed; therefore, two sensitivity analyses were conducted. The first sensitivity analysis used a linear model in which the alcohol consumption variables were log-transformed (after adding 1 unit per week). In the second sensitivity analysis, a negative binomial regression was performed, as the log transformation of drinking outcomes did not result in perfectly normally distributed data. A negative binomial regression as recommended by Atkins et al [21] was used to model the alcohol count data. Model estimates are presented for data from the subset of users who engaged with the program at least once (ie, the active use sample). All analyses were conducted in R version 3.5.1 (R Project) [22].

Results

Sample Characteristics

The majority of the sample was female (308/503, 61.2%) and aged 18 to 73 years (median 41, mean 40, SD 11.24) with a predominant ethnicity of White British (419/503, 83.3%). Within the total sample, 93 participants had registered accounts but did not engage with any of the interactive elements of the web-based program, creating a subgroup (410/503, 81.5%) who responded to at least one question. Baseline alcohol consumption in the past 7 days was 54 units on average (median 45.4, range 0-322.1, SD 37.6) in the total sample, and 55.5 units (median 47.8, range 0-322.1, SD 36.0) in the active use sample. Past week alcohol consumption at 3-month follow-up was 37.2 units on average (median 30.6, range 0-284.6, SD 31.2) in the total sample and 37.4 units (median 30.2, range 0-284.6, SD 31.4) in the subsample. The characteristics of both the total sample and the active use sample are displayed in Table 1; there were no notable differences.

Table 1. Demographic characteristics of the participant sample (N=503).

Baseline characteristic	Full sample (N=503), n (%)[a]	Active use sample[b] (n=410), n (%)[a]
Age (years)		
18-34	165 (32.9)	125 (30.6)
35-44	150 (29.9)	121 (29.6)
45-54	138 (27.5)	123 (30.1)
55-73	49 (9.8)	40 (9.8)
Not specified	1 (0.2)	1 (0.2)
Gender		
Female	308 (61.2)	258 (62.8)
Education level		
University degree or equivalent	289 (57.4)	245 (59.6)
A Levels or equivalent	88 (17.5)	68 (16.5)
GCSE[c] or equivalent	73 (14.5)	56 (13.6)
Other qualifications	39 (7.8)	32 (7.8)
No qualifications	14 (2.8)	10 (2.4)
Relationship status		
Married/long term relationship	316 (62.8)	258 (62.8)
Unmarried/divorced	187 (37.2)	152 (37.1)
Ethnicity		
White British	419 (83.3)	344 (83.9)
White other	46 (9.1)	36 (8.8)
White Irish	25 (5.0)	17 (4.1)
Mixed	4 (0.8)	4 (1.0)
Asian/Asian British	3 (0.6)	3 (0.7)
Black African/Black Caribbean/Black British	2 (0.4)	2 (0.5)
Other	4 (0.8)	4 (1.0)

[a]Due to rounding, percentages may not total 100.

[b]Subsample of participants who responded to at least one question in the Down Your Drink web-based alcohol intervention.

[c]GCSE: General Certificate of Secondary Education.

Change Talk and Sustain Talk

On average, participants entered almost three times the number of segments classified as change talk (mean 12.3, SD 9.79) than as sustain talk (mean 4.1, SD 2.00) (see Table 2). The percentage of change talk was 61.5% on average (median 70%, range 0% to 100%). The majority of participants had entered at least one segment that was classified as change talk (341/410, 83.2%). Change talk encompassed the mention of benefits of quitting/moderating drinking, mention of disadvantages of current drinking behavior, recognition of problematic drinking behavior or needing help, and explicating the intent to change drinking behavior. Noting other people's desire for the participant to quit drinking was not classified as change talk. Examples:

> *I want to be in control of everything I do instead of putting things off because I want to drink alcohol.*
>
> *I need help to stop drinking.*
>
> *I know I am drinking too much and this number of units does not surprise me.*

Sustain talk encompassed naming benefits of current drinking behavior, mentioning perceived disadvantages or fears about quitting/moderating drinking, and expressing a lack of self-confidence in changing current drinking behavior. Examples:

> *I feel positive about life and make plans when I have been drinking.*
>
> *I look forward to drinking.*
>
> *[…] But fear [losing] the good side of alcohol and how it makes me feel.*
>
> *Don't feel confident to change.*
>
> *I have no will power.*

Use of MI and CBT Components

Descriptive information on the use of the MI and CBT components is shown in Table 2. Listing the pros or cons of drinking was the component actively engaged with by the most participants (337/410, 82.2%). The components engaged with by the fewest participants were exploring cravings (16/410, 3.9%), exploring relapse prevention (24/410, 5.9%), and making a relapse plan (9/410, 2.2%).

Table 2. Use of motivational interviewing and cognitive behavioral therapy components by participants in the Down Your Drink web-based alcohol intervention.

Component	Active use sample (n=410)				Users who engaged in component at least once[a]	
	n (%)	Mean (SD)	Median	Range	Mean (SD)	Median
MI[b] components (n=352, 85.9%)						
Change talk[c]	341 (83.2)	10.3 (10.02)	8	0-90	12.3 (9.79)	9
Sustain talk[c]	329 (80.2)	3.3 (2.43)	3	0-13	4.1 (2.00)	4
Percentage change talk[d]	N/A[e]	61.5% (N/A)	70%	0%-100%	72.5% (N/A)	73.7%
Any listed pros or cons	337 (82.2)	N/A	N/A	N/A	N/A	N/A
Pros[c]	334 (81.5)	2.9 (1.98)	3	0-12	3.6 (1.49)	4
Cons[c]	327 (79.8)	6.8 (4.87)	7	0-24	8.4 (3.99)	7
What is important	311 (75.9)	N/A	N/A	N/A	N/A	N/A
CBT[f] components (n=288, 70.2%)						
Setting start date	132 (32.2)	N/A	N/A	N/A	N/A	N/A
Setting a drinking goal	129 (31.5)	N/A	N/A	N/A	N/A	N/A
Completing another part of moderation plan	136 (33.2)	N/A	N/A	N/A	N/A	N/A
Noting alcohol use before DYD	251 (61.1)	N/A	N/A	N/A	N/A	N/A
Noting high-risk situations[c]	113 (27.6)	1.1 (3.04)	0	0-21	4.1 (4.60)	2
Noting strategies[c]	103 (25.1)	2.7 (7.11)	0	0-48	10.9 (10.67)	7
Exploring cravings	16 (3.9)	N/A	N/A	N/A	N/A	N/A
Exploring relapse prevention	24 (5.9)	N/A	N/A	N/A	N/A	N/A
Making a relapse plan	9 (2.2)	N/A	N/A	N/A	N/A	N/A
Exploring thoughts about drinking	49 (12.0)	N/A	N/A	N/A	N/A	N/A
Monitoring drinking[c]	141 (34.4)	12.0 (47.30)	0	0-654	34.8 (75.73)	9

[a]Responses by subset of participants who entered at least one response to a question corresponding with the component (eg, for high-risk situations, a subset of 113 participants entered at least one high-risk situation).

[b]MI: motivational interviewing.

[c]Mean and median refer to the number of statements corresponding to the component.

[d]Mean, median, and range are presented for percentage change talk.

[e]N/A: not applicable.

[f]CBT: cognitive behavioral therapy.

MI Components

Participants responded to questions about the pros and cons of drinking by listing their own perceived pros and cons of drinking. Pros often referred to the experience of alcohol drinking as enjoyable; they also stated that it aided their relaxation and confidence in social situations. Examples:

It makes me more confident and talkative.

It takes the edge off things.

I enjoy the taste.

Cons of drinking most often centered on worries about alcohol drinking affecting the participants' health condition, lowered mood after alcohol drinking, and worries about drinking removing inhibitions, which sometimes led to regrettable behavior. Other cons focused on more practical issues, such as the cost of drinking. Examples:

It's bloody expensive!

Feeling low and depressed the next day and having no motivation

Concerns about health effects

Out of control

Say/do things I regret

A specific part of the DYD program encouraged participants to think about what is most important to them. The responses differed; most participants gave to-the-point answers, whereas others elaborated extensively on the role alcohol played in their lives. Examples:

Good health, friends, my family.

The most important things in my life are my children.

My job and how drinking affect that

My children! […] I want to be a good role model for them and never make them feel worried about me when they are older. I want to be able to have just one drink and not feel as though I can't stop. I'd like to enjoy alcohol socially without feeling ashamed of myself the next day. I want to lose weight, feel good about myself every day and be as healthy as possible for my children and future grandchildren. […]

CBT Components

Some CBT components were rarely engaged with (eg, making a relapse plan and exploring cravings); however, those that were used showed a large variation in responses, with some participants adding large amounts of detail to their responses and others noting only keywords. Most responses were related to alcohol drinking, but not all; for example, in the "goal setting" component, some participants related a broader life goal instead of a specific drinking goal. Examples of goal setting nonspecific to drinking are:

To repay my debts and be good at my job

To keep friends with people, to make new friends and to be respected

Lose weight, be better Mum, get pregnant

Drinking goals also varied in their specificity. Some participants set a clear maximum number of drinks, and others stated a general goal of drinking less. Examples:

To reduce my drinking and the habits that surround it.

To put a stop to the binge drinking sessions

Limit myself to 2 large glasses in the evening and have x2 alcohol free nights

Stated high-risk situations varied from negative feelings to times of day, social situations, and events. Peer pressure was also frequently mentioned. Responses showed an overall good comprehension of the questions, although some answers were unspecific. Examples:

Anger, loneliness, despair

Boredom in the evenings

Going on holiday at New Year will also be a time of temptation

Seeing friends. Alcohol just makes the conversation flow easier. This is probably the hardest situation.

Getting another [because] everyone else is

Social situation

Strategies to cope with high-risk situations could either be selected from a list or noted in free-text responses. Free-text responses were generally unspecific, aimed at distraction or avoidance, and did not account for any difficulties that might arise from the coping strategies. However, some responses seemed to have incorporated strategies that were suggested in the DYD program. Examples:

Don't buy it

Read or watch a film

Doing activities

Drinking slowly and make glass last longer

Use of DYD and Drinking Outcomes

Model estimates are presented for data from the subset of users who engaged with the program at least once, namely the active use sample (n=410). Lower alcohol use at 3 months, when controlled for age, gender, education level, alcohol use at baseline, and number of words, was associated with a greater percentage of change talk (B_{adj} −0.17, 95% CI −0.32 to −0.02, *P*=.03) and a higher number of listed high-risk situations (B_{adj} −2.15, 95% CI −3.92 to −0.38, *P*=.02). In the unadjusted models, listing any high-risk situations (B_{unadj} −8.31, 95% CI −14.88 to −1.74, *P*=.01) and the number of listed strategies to deal with high-risk situations (B_{unadj} −8.31, 95% CI −14.88 to −1.74, *P*=.01) also showed significant associations with alcohol use at 3-month follow-up but not when adjusted for all other components (adjusted R^2 0.38). The complete results are shown in Table 3.

Table 3. Linear model estimates for the subsample of active users (n=410).

Model and variable	Unadjusted[a]			Adjusted[b]		
	B	95% CI	*P* value	B	95% CI	*P* value
Null model[c]						
Covariates						
Baseline alcohol use	N/A[d]	N/A	N/A	0.54	0.47 to 0.62	<.001[e]
Gender (male)	N/A	N/A	N/A	−6.01	−11.45 to −0.56	.03[e]
Education (A level)	N/A	N/A	N/A	4.25	−2.75 to 11.24	.23
Education (O level)	N/A	N/A	N/A	4.52	−2.94 to 11.99	.23
Education (other)	N/A	N/A	N/A	2.22	−7.34 to 11.79	.65
Education (no qualification)	N/A	N/A	N/A	−4.70	−20.70 to 11.30	.56
Age	N/A	N/A	N/A	0.27	0.03 to 0.51	.03[e]
Number of words	N/A	N/A	N/A	0.00	−0.01 to 0.02	.53
Full model[f]						
MI[g] components						
Percentage change talk	−0.11	−0.20 to −0.01	.02[e]	−0.17	−0.32 to −0.02	.03[e]
Any pros or cons listed	−1.17	−8.02 to 5.68	.74	0.34	−13.07 to 13.76	.96
Number of pros	1.31	−0.05 to 2.67	.06	1.60	−0.46 to 3.67	.13
Number of cons	−0.17	−0.83 to 0.50	.62	−0,14	−1.16 to 0.88	.78
What is important	−0.36	−6.64 to 5.91	.91	5.52	−4.23 to 15.27	.27
CBT[h] components						
Setting a start date	−1.41	−7.59 to 4.78	.65	17.50	−4.81 to 39.82	.12
Setting a drinking goal	−2.15	−8.39 to 4.09	.50	2.88	−16.05 to 21.81	.77
Completing another part of the moderation plan	−3.22	−9.38 to 2.94	.30	−13.32	−36.97 to 10.34	.27
Noting alcohol use before DYD[i]	−4.40	−9.83 to 1.03	.11	−5.00	−11.53 to 1.53	.13
Any high-risk situations	−8.31	−14.88 to −1.74	.01[e]	−11.86	−24.22 to 0.49	.06
Number of high-risk situations	−1.82	−3.12 to −0.52	.01[e]	−2.15	−3.92 to −0.38	.02[e]
Any strategies	−1.21	−8.22 to 5.80	.73	5.51	−5.08 to 16.58	.30
Number of strategies	0.01	−0.47 to 0.49	.01[e]	0.64	−0.09 to 1.37	.09
Exploring cravings	2.90	−11.19 to 16.98	.69	3.71	−11.74 to 19.17	.64
Exploring relapse prevention	−1.65	−14.71 to 11.41	.80	−0.24	−15.42 to 14.95	.98
Making a relapse plan	3.52	−14.92 to 21.97	.71	5.05	−15.03 to 25.13	.62
Exploring thoughts about drinking	−0.41	−9.16 to 8.34	.93	−4.00	−15.09 to 7.09	.48
Any monitoring of drinking	2.66	−3.00 to 8.32	.36	1.77	−4.38 to 7.93	.57
Number of times drinking was monitored	0.00	−0.05 to 0.06	.87	−0.01	−0.07 to 0.04	.62

[a]Unadjusted coefficients are based on a series of models in which alcohol use at 3-month follow-up is regressed based on baseline alcohol use, covariates, and each single intervention component.

[b]Adjusted coefficients are based on a model in which alcohol use at 3-month follow-up is regressed based on baseline alcohol use, covariates, and all intervention components.

[c]For the null model only containing the covariates, the adjusted R^2 value is 0.36.

[d]N/A: not applicable.

[e]*P* value <.05.

[f]For the full model, the adjusted R^2 value is 0.38.

[g]MI: motivational interviewing.

[h]CBT: cognitive behavioral therapy.

[i]DYD: Down Your Drink.

Sensitivity Analyses

When comparing the estimates from the linear model without log transformation (Table 3) with the model estimates including log transformations of alcohol use variables (adjusted R^2 0.20, see Table S2 in Multimedia Appendix 2), and the model estimates resulting from a negative binomial regression (McFadden R^2 0.31, see Table S3 in Multimedia Appendix 3), a higher number of listed high-risk situations predicted lower alcohol use at 3-month follow-up in both the model with log-transformation (B_{adj} −0.10, 95% CI −0.16 to −0.04, *P*=.001) and the negative binomial model (B_{adj} −0.07, 95% CI −0.12 to −0.03, *P*=.001). However, a higher number of listed strategies also predicted higher alcohol use at 3-month follow-up in both the log-transformed model (B_{adj} 0.04, 95% CI 0.01 to 0.06, *P*=.002) and the negative binomial model (B_{adj} 0.03, 95% CI 0.01 to 0.05, *P*=.01). No evidence was found in either of the sensitivity analyses for the effect of the percentage of change talk (log-transformed: B_{adj} 0.00, 95% CI −0.01 to 0.00, *P*=.07; negative binomial: B_{adj} 0.00, 95% CI –0.01 to 0.00, *P*=.22). The findings for these two components are therefore inconsistent. For all other components, none of the models showed an effect.

Discussion

Principal Findings

Participants were found to actively use both the MI and CBT components of the DYD website, with MI components used by more participants. The CBT components pertaining to relapse prevention and exploration of cravings were rarely used by participants. Change talk and sustain talk were elicited by the most participants (341/410, 83.2%, and 329/410, 80.2%, respectively); generally, more instances of change talk (median 8) than sustain talk (median 3) were reported, although the between-person variance was large (SD 10.02 and 2.43, respectively). One explanation for the more frequent use of MI components than of CBT components is that the former are presented at the beginning of the program. Participants were free to choose how to move through the program; however, the ordering may still have contributed to the more frequent use of the MI components. A significant finding was that a higher number of listed high-risk situations robustly predicted a greater reduction in alcohol use at 3-month follow-up (*P*=.02).

Analyses were repeated with log-transformed measures of alcohol use at baseline and 3-month follow-up and by applying negative binomial regression. All the models showed an effect of the number of listed high-risk situations; however, there were also some inconsistent findings. There were differences between the models in the effects of the percentage of change talk and the number of listed strategies to deal with high-risk situations. The number of strategies to deal with high-risk situations predicted higher alcohol use in the log-transformed and negative binomial models but not in the linear model without log transformation. The percentage of change talk was only found to have an effect on reduction of alcohol use in the linear model without log transformation. All other components showed null findings.

The lack of effects of the percentage of change talk found in the sensitivity models may be due to a lack of evidence rather than the absence of a true effect. A meta-analysis by Magill [10] summarized 58 MI process studies, including 21 on face-to-face alcohol interventions. A higher proportion of change talk was found to be related to the reduction of all risky behaviors, including reduction of alcohol use. It is noteworthy that the latest guidance on MI practice suggests removing the pros and cons decisional balance exercise [9], as it can have the undesired effect of encouraging sustain talk [23].

A higher number of noted strategies was associated with a slight increase in alcohol use at 3-month follow-up according to the sensitivity analyses. A possible reason for this seemingly paradoxical finding is that participants who are having more severe alcohol problems may work more extensively on the program. For this group, a web-based alcohol intervention may offer insufficient support to actually reduce their alcohol use, thereby possibly leading to increased alcohol use. As only some of the participants were asked to fill in the complete AUDIT questionnaire in the original trial, this hypothesis could not be tested in this study. However, a recent individual patient data meta-analysis (19 trials) revealed no difference in the effectiveness of internet interventions between binge drinkers and non–binge drinkers, nor any difference in effectiveness related to the amount of alcohol consumption at baseline (heavy drinkers vs nonheavy drinkers) [5].

Another component that is often considered an “active ingredient” of brief alcohol interventions is self-monitoring of alcohol use [24]. Self-monitoring (ie, entry of drinks into the Drinking Episodes Diary), was actively used at least once by only 141/410 (34.4%) of the active participants. Also, for participants who did use it, the amount of times they reported their drinking was very skewed: half of them reported their drinking a maximum of 9 times (mean 34.8, SD 75.73). Among DYD users, there was also a lack of active engagement with relapse prevention exercises and exercises on craving. Craving was previously found to be an important predictor of relapse [25].

Future research should focus on testing the roles of components of interest in encouraging alcohol reduction using a preregistered analysis plan while considering the influence of the ordering of components on their use and subsequent effects.

Strengths and Limitations

This study made use of a large sample of active users of a web-based alcohol intervention program. These users were e-help–seekers who were not currently enrolled in a treatment pathway. Uniquely, in this study, we were able to analyze a large number of free-text responses, thus obtaining insight into how well the participants understood the questions and whether key MI and CBT components were actively used. The free-text responses also enabled the assessment of key MI mechanisms of change talk and sustain talk, which may influence drinking outcomes [26]. Whether the program encouraged change talk or softened sustain talk (counter change talk) over time could not be assessed in this study. The presence of these key MI mechanisms and the use of the separate components were assessed independently by two coders with high interrater reliability. Change talk and sustain talk were coded using the CLEAR coding system [19], which was developed for coding of in-therapy client language. To account for the amount of total activity within the program as a possible confounder, the total number of words was included as a covariate within the model. These analyses were exploratory and were not preregistered; several sensitivity analyses are therefore presented. It is possible that significance of the effects of some components may be detected when using a larger sample, as we included many parameters without any effects in the full model. The generalizability of the study is limited because the sample only consisted of participants whose alcohol use at 3-month follow-up was known; this complete case analysis limits the generalizability of the results to nonresponders. Furthermore, the DYD intervention was closely modeled on the face-to-face MI/CBT approach used in therapeutic settings, and it required engagement and reflection across many different exercises. DYD was only accessible on a computer (ie, not compatible with smartphones, which were less prolific in 2007 when DYD was first launched). The current results are therefore only generalizable to similar extensive web-based MI/CBT interventions. More recent web-based interventions are responsive websites or apps, which tend to include a small number of “active” behavior change components that can be used easily and quickly, and are not intended to elicit change talk or sustain talk.

Conclusions

A web-based alcohol intervention was able to elicit both change talk and sustain talk. A higher number of listed high-risk situations can predict lower alcohol use at 3-month follow-up. Other components show inconsistent findings and should be studied further using a preregistered analysis plan. This study points to components of the DYD website that may constitute effective internet alcohol moderation programs and thus complies with the high research priority of studying specific components of web-based interventions.

Acknowledgments

This study was funded by the Erasmus Trust Fund Foundation (the Netherlands), Foundation “De Drie Lichten” (the Netherlands), and Foundation “Jo Kolk Studiefonds” (the Netherlands). The funders had no role in the design, analysis, or interpretation of the data nor in the writing and publication of the scientific paper or the results. We are grateful to Elizabeth Murray for initiating the collaboration between the authors and Richard McGregor (Codeface Ltd) for extracting the data needed for the analyses. The authors acknowledge the contributions of Elizabeth Murray, Paul Wallace, Jim McCambridge, Ian R White, Simon G Thompson, Eleftheria Kalaitzaki, and Christine Godfrey to the original DYD trial.

Conflicts of Interest

Authors SL and ZK were involved in the development of the DYD intervention described in this study.

Multimedia Appendix 1
Coding scheme.
[DOC File , 69 KB - jmir_v22i9e17285_app1.doc]

Multimedia Appendix 2
Supplementary Table S2: Linear model estimates with log-transformed alcohol use variables at baseline (independent variable) and 3-month follow-up (dependent variable) for the active use sample (n=410).
[DOC File , 70 KB - jmir_v22i9e17285_app2.doc]

Multimedia Appendix 3
Negative binomial model estimates for the active use sample.
[DOCX File , 16 KB - jmir_v22i9e17285_app3.docx]

References

1. Sundström C, Blankers M, Khadjesari Z. Computer-Based Interventions for Problematic Alcohol Use: a Review of Systematic Reviews. Int J Behav Med 2017 Oct;24(5):646-658 [FREE Full text] [doi: 10.1007/s12529-016-9601-8] [Medline: 27757844]

2. Collins SE, Kirouac M, Lewis MA, Witkiewitz K, Carey KB. Randomized controlled trial of web-based decisional balance feedback and personalized normative feedback for college drinkers. J Stud Alcohol Drugs 2014 Nov;75(6):982-992 [FREE Full text] [doi: 10.15288/jsad.2014.75.982] [Medline: 25343656]
3. Boon B, Risselada A, Huiberts A, Riper H, Smit F. Curbing alcohol use in male adults through computer generated personalized advice: randomized controlled trial. J Med Internet Res 2011;13(2):e43 [FREE Full text] [doi: 10.2196/jmir.1695] [Medline: 21719412]
4. Brendryen H, Lund IO, Johansen AB, Riksheim M, Nesvåg S, Duckert F. Balance—a pragmatic randomized controlled trial of an online intensive self-help alcohol intervention. Addiction 2014 Feb;109(2):218-226. [doi: 10.1111/add.12383] [Medline: 24134709]
5. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: An individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714 [FREE Full text] [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
6. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9:CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
7. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011 Feb;106(2):267-282. [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]
8. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
9. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. New York, NY: Guilford Press; 2012.
10. Magill M, Apodaca TR, Borsari B, Gaume J, Hoadley A, Gordon REF, et al. A meta-analysis of motivational interviewing process: Technical, relational, and conditional process models of change. J Consult Clin Psychol 2018 Feb;86(2):140-157 [FREE Full text] [doi: 10.1037/ccp0000250] [Medline: 29265832]
11. Nurmi J, Knittle K, Ginchev T, Khattak F, Helf C, Zwickl P, et al. Engaging Users in the Behavior Change Process With Digitalized Motivational Interviewing and Gamification: Development and Feasibility Testing of the Precious App. JMIR Mhealth Uhealth 2020 Jan 30;8(1):e12884 [FREE Full text] [doi: 10.2196/12884] [Medline: 32003750]
12. Linke S, McCambridge J, Khadjesari Z, Wallace P, Murray E. Development of a psychologically enhanced interactive online intervention for hazardous drinking. Alcohol Alcohol 2008;43(6):669-674 [FREE Full text] [doi: 10.1093/alcalc/agn066] [Medline: 18693217]
13. Khadjesari Z, Stevenson F, Godfrey C, Murray E. Negotiating the 'grey area between normal social drinking and being a smelly tramp': a qualitative study of people searching for help online to reduce their drinking. Health Expect 2015 Dec;18(6):2011-2020 [FREE Full text] [doi: 10.1111/hex.12351] [Medline: 25676536]
14. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2008 Feb;103(2):218-227. [doi: 10.1111/j.1360-0443.2007.02063.x] [Medline: 18199300]
15. Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, et al. On-line randomized controlled trial of an internet based psychologically enhanced intervention for people with hazardous alcohol consumption. PLoS One 2011 Mar 09;6(3):e14740 [FREE Full text] [doi: 10.1371/journal.pone.0014740] [Medline: 21408060]
16. Khadjesari Z, White IR, McCambridge J, Marston L, Wallace P, Godfrey C, et al. Validation of the AUDIT-C in adults seeking help with their drinking online. Addict Sci Clin Pract 2017 Jan 04;12(1):2 [FREE Full text] [doi: 10.1186/s13722-016-0066-5] [Medline: 28049515]
17. Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Godfrey C, et al. Test-retest reliability of an online measure of past week alcohol consumption (the TOT-AL), and comparison with face-to-face interview. Addict Behav 2009 Apr;34(4):337-342. [doi: 10.1016/j.addbeh.2008.11.010] [Medline: 19097705]
18. Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, Godfrey C, et al. The DYD-RCT protocol: an on-line randomised controlled trial of an interactive computer-based intervention compared with a standard information website to reduce alcohol consumption among hazardous drinkers. BMC Public Health 2007 Oct 26;7:306 [FREE Full text] [doi: 10.1186/1471-2458-7-306] [Medline: 17963483]
19. Glynn LH, Moyers TB. Manual for the Client Language EAsy Rating (CLEAR) Coding System: Formerly Motivational Interviewing Skill Code (MISC) 1.1. Albuquerque, NM: Center on Alcoholism, Substance Abuse, and Addictions; 2012.
20. James IA, Blackburn IM, Reichelt FK. Manual of the Revised Cognitive Therapy Scale (CTS-R). Newcastle, UK: Newcastle Cognitive and Behavioural Therapies Centre; Dec 2001.
21. Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. Psychol Addict Behav 2013 Mar;27(1):166-177 [FREE Full text] [doi: 10.1037/a0029508] [Medline: 22905895]
22. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. URL: http://www.R-project.org [accessed 2020-08-19]

23. Miller WR, Rollnick S. Ten things that motivational interviewing is not. Behav Cogn Psychother 2009 Mar;37(2):129-140. [doi: 10.1017/S1352465809005128] [Medline: 19364414]
24. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
25. Witkiewitz K, Bowen S, Douglas H, Hsu SH. Mindfulness-based relapse prevention for substance craving. Addict Behav 2013 Feb;38(2):1563-1571 [FREE Full text] [doi: 10.1016/j.addbeh.2012.04.001] [Medline: 22534451]
26. Apodaca TR, Borsari B, Jackson KM, Magill M, Longabaugh R, Mastroleo NR, et al. Sustain talk predicts poorer outcomes among mandated college student drinkers receiving a brief motivational intervention. Psychol Addict Behav 2014 Sep;28(3):631-638 [FREE Full text] [doi: 10.1037/a0037296] [Medline: 25222170]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Consumption
CBT: cognitive behavioral therapy
CLEAR: Client Language EAsy Rating coding system
CTS-R: Revised Cognitive Therapy Scale
DYD: Down Your Drink
e-help: electronic help
GCSE: General Certificate of Secondary Education
MI: motivational interviewing

Edited by G Eysenbach; submitted 03.12.19; peer-reviewed by B Borsari, P Dutey-Magni, C Garnett; comments to author 03.02.20; revised version received 16.04.20; accepted 03.06.20; published 01.09.20.

Please cite as:
Mujcic A, Linke S, Hamilton F, Phillips A, Khadjesari Z
Engagement With Motivational Interviewing and Cognitive Behavioral Therapy Components of a Web-Based Alcohol Intervention, Elicitation of Change Talk and Sustain Talk, and Impact on Drinking Outcomes: Secondary Data Analysis
J Med Internet Res 2020;22(9):e17285
URL: https://www.jmir.org/2020/9/e17285
doi:10.2196/17285
PMID:32870162

Original Paper

User Experience and Effects of an Individually Tailored Transdiagnostic Internet-Based and Mobile-Supported Intervention for Anxiety Disorders: Mixed-Methods Study

Kiona K Weisel[1], MSc; Anna-Carlotta Zarski[1], PhD; Thomas Berger[2], PhD; Tobias Krieger[2], PhD; Christian T Moser[2], MSc; Michael P Schaub[3], PhD; Dennis Görlich[4], PhD; Matthias Berking[1], PhD; David D Ebert[5], PhD

[1]Department of Clinical Psychology and Psychotherapy, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

[2]Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland

[3]Swiss Research Institute of Public Health and Addiction ISGF, Associated to the University of Zurich, Zurich, Switzerland

[4]Institute of Biostatistics and Clinical Research, Westfälische Wilhelms-Universität Münster, Münster, Germany

[5]Clinical, Neuro- & Development Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Kiona K Weisel, MSc
Department of Clinical Psychology and Psychotherapy
Friedrich-Alexander University Erlangen-Nürnberg
Nägelsbachstraße 25a
Erlangen
Germany
Phone: 49 (0)9131 ext 8567570
Email: kiona.weisel@fau.de

Abstract

Background: Internet interventions have been shown to be effective in treating anxiety disorders. Most interventions to date focus on single disorders and disregard potential comorbidities.

Objective: The aim of this mixed-methods study was to investigate feasibility, user experience, and effects of a newly developed individually tailored transdiagnostic guided internet intervention for anxiety disorders.

Methods: This study is an uncontrolled, within-group, baseline, postintervention pilot trial with an embedded qualitative and quantitative process and effect evaluation. In total, 49 adults with anxiety disorders (generalized anxiety disorder n=20, social phobia n=19, agoraphobia without panic n=12, panic with agoraphobia n=6, panic without agoraphobia n=4, subclinical depression n=41) received access to the 7-session intervention. We examined motivation and expectations, intervention use, user experience, impact, and modification requests. Qualitative data were assessed using semistructured interviews and analyzed by qualitative content analysis. Quantitative outcomes included symptom severity of anxiety and depression (Hamilton Anxiety Rating Scale [HAM-A], Quick Item Inventory of Depressive Symptomatology clinician rating [QIDS-C]), diagnostic status in clinical interviews (Mini International Neuropsychiatric Interview [MINI]), and web-based self-reports (Generalized Anxiety Disorder–7 [GAD-7], Center for Epidemiological Studies Depression Scale [CES-D], Beck Anxiety Inventory [BAI], Panic and Agoraphobia Scale [PAS], Social Phobia Scale [SPS], Patient Health Questionnaire–9 [PHQ-9]) at baseline and postassessment. Quantitative data was analyzed by comparing within-group means expressed as Cohen *d*.

Results: Anxiety symptom severity (HAM-A *d*=1.19) and depressive symptoms (QIDS-C *d*=0.42) improved significantly, and 54% (21/39) no longer were diagnosed as having any anxiety disorder. The main positive effects were the general improvement of disease burden and attentiveness to feelings and risk situations while the main negative effects experienced were lack of change in disease burden and symptom deterioration. The most prevalent reasons for participation were the advantages of online treatment, symptom burden, and openness toward online treatment. Helpful factors included support, psychoeducation and practicing strategies in daily life; the main hindering factors were too little individualization and being overwhelmed by the content and pace.

Conclusions: The intervention was found to be feasible and results show preliminary data indicating potential efficacy for improving anxiety and depression. The next step should be the evaluation within a randomized controlled trial. Concerning

intervention development, it was found that future interventions should emphasize individualization even more in order to further improve the fit to individual characteristics, preferences, and needs.

(J Med Internet Res 2020;22(9):e16450) doi:10.2196/16450

KEYWORDS

transdiagnostic; anxiety; depression; tailored; internet intervention

Introduction

Internet interventions can be effective means of treating mental health problems such as anxiety disorders [1-3]. Anxiety disorders are highly prevalent [4], and individuals suffering from anxiety disorders tend to experience significant impairment in quality of life and a decreased sense of well-being and occupational and family satisfaction [5,6]. Anxiety disorders have also been found to be highly comorbid and act as risk factors for developing other anxiety disorders (comorbidity rate of patients with generalized anxiety disorder in the past 12 months and any other anxiety disorder: 55.9%) or major depressive disorder (comorbidity of generalized anxiety disorder and major depressive disorder: 59.1%) [7-10].

The fact that the majority of individuals who suffer from a mental disorder do not receive treatment is a demanding public health issue [11]. One primary reason for nontreatment seeking behavior apart from structural barriers such as treatment availability is attitudinal barriers including preference for self-reliance, low perceived treatment need, poor mental health literacy, and fear of stigmatization [12-14]. Internet interventions offer many advantages that could help bridge this treatment gap. Meta-analytic evidence has found internet interventions to be efficacious with medium to large effect sizes for the treatment of anxiety disorders [3,15-17].

As anxiety disorders are often comorbid with other anxiety disorders and depression [7-9], there are advantages to treating all comorbid disorders within one transdiagnostic treatment protocol. Transdiagnostic treatment for anxiety disorders and depression can be applied to a broad range of patients regardless of their primary diagnosis as they are designed to target common underlying factors and also provide a variety of treatment [18,19]. Research investigating internet-based transdiagnostic treatment protocols for different anxiety disorders and comorbid depression has found these type of interventions to be efficacious [20-22].

Recently, there have been attempts to further individualize treatments according to the symptom profile and preferences of patients, which is also referred to as individual tailoring [23-25]. Beyond being able to address comorbidities, the main advantage is that patient preferences are considered in the treatment protocol, which could increase treatment motivation and therefore adherence and ideally also improve the outcome [26]. One trial found that effects in an internet-based intervention for depression were more pronounced when individual tailoring was applied compared with standardized treatment indicating the potential of individual tailoring [23]. To the best of our knowledge, no meta-analytic review exists on the sole effect of tailoring; however, meta-analytic evidence proposes that transdiagnostic and individual tailored approaches are promising when dealing with comorbidity with medium to large effect sizes for anxiety (g=0.82) and depression: (g=0.79) [27].

Some general aspects that remain unknown in internet interventions are (1) why individuals choose to participate, (2) how such interventions work including helpful and hindering factors [17,28,29], and (3) subjective impact including negative effects [30]. One way to explore these themes is through interviews with participants and qualitative data analysis.

The aim of this pilot feasibility study is to investigate a newly developed individually tailored transdiagnostic guided internet intervention for anxiety disorders with and without comorbid subclinical depression and explore feasibility. Qualitative and quantitative data and methods will be used to understand user experience focusing on motivation for participation and initial expectations, intervention use, and helpful and hindering factors. The impact of the intervention is explored through qualitative interviews as well as self-report and clinician-rated diagnostics on symptom severity, occurrence of clinical diagnoses, and positive and negative training effects. Finally, suggestions for intervention improvement and development will be derived.

Methods

The paper describes the findings of a pilot feasibility study for a randomized controlled trial that was registered in the German Clinical Trials Register [DRKS00012656] and received ethical approval from Friedrich-Alexander University Erlangen-Nürnberg (144-16 B).

Recruitment

Participants were recruited via German health insurance companies, a study webpage, and open recruitment strategies such as social media and Google Ads for a primary trial on the prevention of depression and anxiety [31]. Individuals with a clinical diagnosis of a major depressive disorder in the screening process were referred to another trial [32]. If they did not fulfill the criteria of the prevention trial due to a clinical diagnosis of an anxiety disorder and did not have a clinical diagnosis of major depressive disorder, they were referred to this study.

Assessment of Eligibility

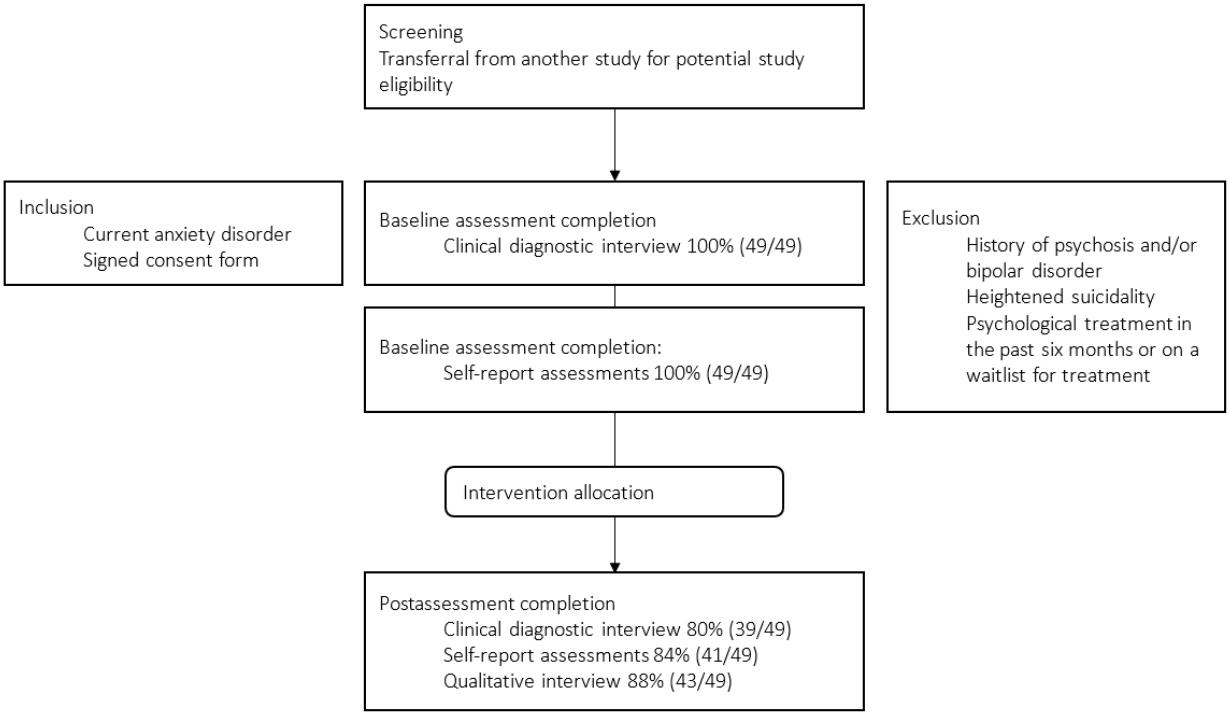
Participants were eligible to participate in the study if they fulfilled the following inclusion criteria of having a current diagnosis of an anxiety disorder (generalized anxiety disorder, panic disorder, agoraphobia, social phobia) assessed in the diagnostic interview based on the Mini International Neuropsychiatric Interview (MINI) [33] and signed informed consent without any of the following exclusion criteria: (1) history of psychosis, (2) bipolar disorder, (3) psychological

treatment in the past 6 months, (4) currently on a waiting list for psychological treatment, (5) heightened suicidality, (6) having a current, or past 6 months, episode of a major depressive disorder. To increase internal validity, we decided not to include individuals with a major depressive disorder and redirected them to a different trial [32] as we assumed they might have other characteristics and needs.

Study Design

The study has an uncontrolled, within-group, baseline, postintervention design with an embedded qualitative and quantitative process evaluation. All participants (n=49) included in this study received access to the individually tailored transdiagnostic guided treatment for anxiety disorders. Clinical diagnostic interviews on diagnostic status and symptom severity of anxiety disorders and major depressive disorder were conducted at baseline and postassessment 8 weeks after intervention allocation. The participants also completed web-based self-report assessments on anxiety and depressive symptom severity and a question on treatment motivation and guidance preference at baseline and postassessment. A semistructured qualitative interview was conducted at postassessment. Figure 1 displays the study flow.

Figure 1. Study flow.



Intervention

The internet intervention comprises 7 sessions plus one booster session. The content includes psychoeducation; methods to reduce incongruence between personal values, needs, and behavior; behavioral activation; exposure; and problem solving. The intervention is mainly text-based with additional elements such as short educational videos and audio files. To promote the transfer of acquired skills into daily life, participants could opt to receive short messages to their phone throughout the day (sent through an app or a messaging service) with motivational sentences or mini tasks referred to as Tiny Tasks. For more information on the reported intervention, see the published study protocol of the primary prevention trial [31]. See Textbox 1 for an overview of the intervention sessions.

Textbox 1. Session overview and elective modules.

1. Behavioral activation: satisfying needs and goals
- Introduction to the training and core elements
- Strengthening motivation and personal goal setting
- Understanding the relationship between personal needs and values and identifying discrepancies
- Planning activities to strengthen core values

2. Behavioral activation: overcoming difficulties and pleasant activity scheduling
- Overcoming difficulties of behavioral activation
- Understanding the nature of avoidant behaviors
- Planning mood-enhancing activities

3. Psychoeducation
- Psychoeducational information on depression and anxiety including etiological and maintaining factors
- Identifying individual symptomatology and course of development

4. Cognitive restructuring
- Introduction to the causal relationship between cognitions and emotions
- Application of a thought record
- Identifying automatic negative thoughts and practicing cognitive flexibility

5. Exposure I or Problem solving I
- Practicing problem solving by distinguishing between solvable and unsolvable problems and applying a 6-step problem-solving plan
- Practicing exposition to fear-inducing situations based on a personal fear hierarchy

6. Exposure II or Problem solving II

7. Plan for the future
- Recap of the training
- Plan for the future and relapse prevention

8. Booster session
- Reflecting on goal attainment
- Further planning of the future

Elective modules: rumination and worries, acceptance, relaxation, reducing alcohol, self-worth, perfectionism, appreciation and gratitude, sleep

Individual Tailoring

During the intervention development phase, there was an emphasis on individual tailoring which manifested itself through (1) tailoring to core clinical characteristics (any anxiety disorder or depression); (2) receiving optional Tiny Tasks; (3) choosing elective modules on various psychological topics such as acceptance, relaxation, or reducing alcohol based on interest, preference, and needs; (4) personal goal setting with monitoring of advancement in achieving goals and making adjustments throughout the intervention; (5) receiving personalized guidance by eCoaches who also monitored individual intervention use through adherence. The presented content in the intervention is triggered by patient input.

Guidance and Adherence Monitoring

After completion of a session, patients receive written content-focused feedback by an eCoach. eCoaches are supervised psychologists or psychotherapists (in training) who provide manualized text-based feedback and monitor for adherence and potential crises throughout the intervention. In case of noncompliance to the intervention, eCoaches send reminder messages to encourage session completion. Patients are also sent automatic weekly email reminders by the platform in case of nonadherence.

Assessments and Data Management

Qualitative Data

The interview manual was developed in collaboration with clinical experts and comprises 7 open questions including reasons for participation and expectations, training experience

including helpful and hindering factors, impact of treatment, and modification requests. All participants were asked to participate in the qualitative interviews regardless of their actual treatment progress, intervention adherence, or session completion rates. Table 1 gives an overview of the topical domains and interview questions translated into English. The interviews were recorded and the content was transcribed verbatim.

Table 1. Overview of interview questions and domains.

Code	Domain	Question
Q1	Motivation for participation	Why did you participate in the online training?
Q2	Fulfilled expectations	Which expectations toward the training were fulfilled?
Q3	Unfulfilled expectations	Which expectations toward the training were not fulfilled?
Q4	Impact of online training	How has your disease burden changed by using the online training?
Q5	Helpful training event	What part of the training was particularly helpful in improving your psychological well-being?
Q6	Hindering training element	What would you have needed in addition from the training to help improve your psychological well-being?
Q7	Negative effects	Which elements of the training had no or negative effects on your psychological well-being?

Quantitative Assessments

Quantitative assessments took place during screening to complete study inclusion, at baseline before intervention access, and after intervention completion (8 weeks after baseline). Assessments comprised diagnostic interviews conducted by clinicians via telephone and web-based self-report assessments. The clinicians were blind to the fact that there was no control group. Figure 1 displays the study flow. The web-based assessments included measures of anxiety symptom severity, depression symptom severity, a question on treatment motivation, and guidance preference.

Anxiety disorders and major depressive disorder were assessed by an adaption of the MINI 5.0 [34]. Severity of anxiety symptoms was assessed by the Hamilton Anxiety Rating Scale (HAM-A; 14 items; α_{T1}=.76) [35,36] and the Quick Item Inventory of Depressive Symptomatology clinician rating (QIDS-C; 16 Items; α_{T1}=.64) [37,38] via telephone by diagnostic raters.

Generalized anxiety disorder and symptom severity were measured by the Generalized Anxiety Disorder 7 (GAD-7; 7 items; α_{T1}=.82) [39,40]. The Beck Anxiety Inventory (BAI; 21 items; α_{T1}=.91) was used to assess clinical anxiety [41]. Panic and agoraphobia symptoms were assessed by the Panic and Agoraphobia Scale (PAS; 13 items; α_{T1}=.89) [42]. The Social Phobia Scale assessed social anxiety and pertains to fears of scrutiny during observations by others (SPS; 20 items; α_{T1}=.93) [43,44].

Depressive symptoms were also assessed by the Patient Health Questionnaire (PHQ-9; 9 items; α_{T1}=.73) [45] and the Center for Epidemiological Studies Depression Scale (CES-D; 20 items; α_{T1}=.68) [46,47].

Motivation to receive online treatment as well as guidance preference were assessed by prompting participants to choose from a set of predefined answers in the web-based assessment. For an overview of all possible answers, see Multimedia Appendix 1.

Study Adherence

Adherence to study completion was monitored. To standardize the study adherence procedure, a systematic adherence protocol was instated. After noncompletion of an assessment or the diagnostic interview, participants were sent reminder emails after 7, 14, 21, and 28 days and text messages after 14, 21, and 28 days; reminder calls took place after 21 days. The text messages contained different motivational approaches to appeal to different mindsets such as helping others by providing data, having received the training in exchange for completing assessments, furthering scientific evidence, and supporting individuals of the study management team in completing their scientific degrees.

Quantitative Data Analysis

Feasibility of the intervention was assessed by exploring changes in the diagnostic status of any anxiety disorder or major depressive disorder and symptom improvement of anxiety and depression. Pre-post data were compared with paired *t* tests and expressed by Cohen *d* and the 95% confidence interval [48,49]. We also reported the mean percentage of symptom improvement per assessment scale. As this is a pilot feasibility trial exploring data and not testing hypotheses, we decided not to implement any strategies to estimate missing data and instead used completer data only. Due to the explorative nature of the trial, we also did not control the global significance level for the multiple testing problem. We also report baseline differences as median comparisons between self-report assessment completers and noncompleters investigated by the Mann-Whitney *U* test.

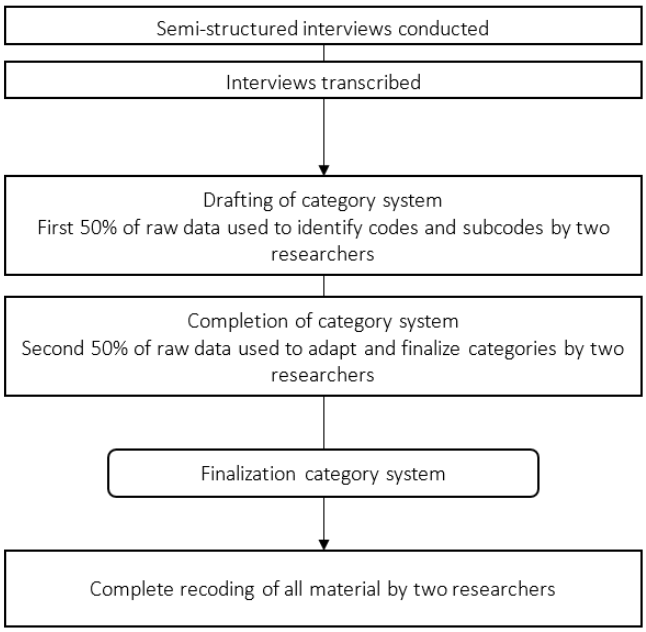
Qualitative Data Analysis

The recorded interviews were transcribed verbatim according to a predefined transcription guide. Content analysis and coding rules followed recommendations by Mayring [50]. The software program MAXQDA version 18.0.0 (VERBI GmbH) was used to analyze the qualitative data. Adherence to standards for reporting qualitative research was ensured by following the consolidated criteria for reporting qualitative research (COREQ) [51,52].

Taking an inductive approach, codes were developed by two researchers (KKW, LK) who used 50% of the raw data mindful of the topical domains. Codes were discussed until agreement was reached for a preliminary category system. Following this, the other 50% of raw data were analyzed by identifying codes and sorting them into the existing category system. If the codes did not match the existing system, the system was adapted by adjusting codes or creating new ones. The two researchers then discussed the coding and finalized the categorical system.

After finalization of the categorical system, two researchers (KKW, MNC) independently coded 10% of the data and analyzed their ratings in MAXQDA to determine an interrater agreement which is reported in the form of Cohen kappa (threshold was set to 10%). Figure 2 depicts the process of the qualitative data analysis.

Figure 2. Study flow qualitative data analysis.



Results

Baseline Characteristics

In total, 49 participants were included in the study, of which the majority was female (38/49, 78%). Participants were aged 40 years on average; the youngest participant aged 22 years and the oldest 68 years. Apart from two participants who resided in Switzerland and Austria, all others (47/49, 96%) had their residence in Germany. More than half of the participants (27/49, 55%) lived in cities with less than 100,000 inhabitants; 45% (22/49) of participants stated they lived in a city with more than 100,000 inhabitants. At baseline, all participants (49/49, 100%) included in the study had at least one anxiety disorder, while 18% (9/49) had at least two anxiety disorders and 6% (3/49) had three anxiety disorders. On average, participants had heightened symptoms of anxiety with a mean value of 21.29 (SD 7.79) on the HAM-A, 10.31 (SD 4.11) on the GAD-7, 38.35 (SD 10.96) on the BAI, 10.20 (SD 8.58) on the PAS, and 20.51 (SD 14.97) on the SPS. They also showed heightened symptom severity of depression with an average value of 8.92 (SD 4.41) on the QIDS-C, 21.71 (SD 6.58) on the CES-D, and 11.04 (SD 4.31) on the PHQ-9.

Of all participants, 43% (21/49) had no prior experience with psychotherapy and 57% (28/49) had some type of experience with psychotherapy. Of the 28 individuals with prior treatment experience, 32% (9/28) rated their experience as very helpful, 61% (17/28) found it somewhat helpful, and 7% (2/28) did not find their treatment to have been helpful. Considering prior experience with health-related trainings, 73% (36/49) claimed to have some and 27% (13/49) did not. See Table 2 for an overview.

Table 2. Baseline characteristics (n=49).

Characteristics	Value
Sociodemographic characteristics	
Age in years, mean (SD)	40.45 (12.9)
Gender, female, n (%)	38 (78)
Country of residence, n (%)	
Germany	47 (96)
Switzerland	1 (2)
Austria	1 (2)
Number of inhabitants, n (%)	
Less than 5000	8 (16)
5000-10,000	5 (10)
10,000-20,000	6 (12)
20,000-50,000	3 (6)
50,000-100,000	5 (10)
100,000-500,000	8 (16)
More than 500,000	14 (29)
Ethnicity, n (%)	
White	45 (92)
Other	4 (8)
Education, n (%)	
8 years of schooling	1 (2)
10 years of schooling	5 (10)
Abitur or 3 to 3.5 year traineeship	19 (39)
Bachelor or equivalent	8 (16)
Masters or equivalent	15 (31)
Doctorate degree	1 (2)
Previous psychological treatment, n (%)	
Yes	28 (57)
No	21 (43)
For those who experienced psychological treatment (n=28): How helpful was it? n (%)	
Not helpful	2 (7)
Somewhat helpful	17 (61)
Very helpful	9 (32)
Previous experience with health-related trainings, n (%)	
Yes	36 (73)
No	13 (27)
Symptom severity, mean (SD)	
HAM-A[a] anxiety	21.29 (7.79)
GAD-7[b] anxiety	10.31 (4.11)
BAI[c] anxiety	38.35 (10.96)
PAS[d] anxiety	10.20 (8.58)
SPS[e] anxiety	20.51 (14.97)

Characteristics	Value
QIDS-C[f] depression	8.92 (4.41)
CES-D[g] depression	21.71 (6.58)
PHQ-9[h] depression	11.04 (4.31)

[a]HAM-A: Hamilton Anxiety Rating Scale.

[b]GAD-7: Generalized Anxiety Disorder–7 item.

[c]BAI: Beck Anxiety Inventory.

[d]PAS: Panic and Agoraphobia Scale.

[e]SPS: Social Phobia Scale.

[f]QIDS-C: Quick Item Inventory of Depressive Symptomatology.

[g]CES-D: Center for Epidemiological Studies Depression Scale.

[h]PHQ-9: Patient Health Questionnaire–9 item.

Assessment Completion

In total, 43 (43/49, 88%) qualitative interviews were conducted. The 6 individuals who did not complete the interview dropped out of the study, 2 of which informed the study team they did not want to continue and the other 4 could no longer be reached. The interview duration ranged from 1 minute 51 seconds to 8 minutes 21 seconds. The interrater agreement of codes in 10% of the interview data was 81% between raters (KKW, MNC).

As the quantitative analysis was based on assessment completer data and we had an assessment dropout of 16% (8/49), we additionally investigated baseline differences of symptom severity and intervention adherence rates of quantitative assessment completers and noncompleters (did not complete the web-based assessment at postintervention). Depression and anxiety symptom severity did not differ significantly between those who dropped out and those who completed the assessments. However, completers had significantly lower anxiety levels on the BAI (median 33) compared with noncompleters (median 44.50, U=80.5, z=–2.27, P=.02). For a complete overview of baseline differences, see Multimedia Appendix 1.

Impact

Considering changes in diagnostic status at postassessment, of the individuals who completed the diagnostic interview, 46% (18/39) still had at least one anxiety disorder while 54% (21/39) no longer were diagnosed as having any anxiety disorder. There was a significant improvement of anxiety symptoms assessed by the HAM-A from 20.71 (SD 7.87, n=42) at baseline to 12.76 at postassessment (SD 9.18, n=42, T=7.0, df=41, P<.001, d=1.19, 95% CI 0.73 to 1.66), which translates to a mean symptom improvement of 38%. There was also a significant improvement of symptoms of depression assessed by the QIDS-C [38], from 8.4 (SD 4.13, n=42) to 6.38 at postassessment (SD 4.83, n=42, T=2.51, df=41, P=.016, d=0.42, 95% CI 0.01 to 0.86), which is a mean symptom of improvement of 24%.

All scales assessing anxiety symptom severity and depression symptom severity showed a significant improvement from baseline to postassessment. We also observed an improvement on the PAS [42] from baseline (mean 9.66 [SD 8.59], n=41) to postassessment (mean 8.12 [SD 7.06], n=41), which was not significant (T=1.86, df=40, P=.07, d=0.28, 95% CI –0.16 to 0.71). For a full overview see Tables 3 and 4.

Table 3. Completer baseline and postintervention data.

Questionnaire and assessment point	n[a]	mean (SD)	T score	df[b]	*P* value	*d*	95% CI	Mean symptom improvement (%)
HAM-A[c]			7.0	41	<.001	1.19	0.73 to 1.66	38.39
T1	42	20.71 (7.87)						
T2	42	12.76 (9.18)						
GAD-7[d]			4.94	40	<.001	0.75	0.31 to 1.20	29.85
T1	41	10.05 (4.15)						
T2	41	7.05 (3.97)						
BAI[e]			3.91	40	<.001	0.58	0.14 to 1.02	11.18
T1	41	36.93 (10.54)						
T2	41	32.80 (9.05)						
PAS[f]			1.86	40	.07	0.28	–0.16 to 0.71	15.94
T1	41	9.66 (8.59)						
T2	41	8.12 (7.06)						
SPS[g]			2.32	40	.03	0.35	–0.09 to 0.79	18.68
T1	41	18.95 (13.73)						
T2	41	15.41 (12.80)						
QIDS-C[h]			2.51	41	.02	0.42	–0.01 to 0.86	24.05
T1	42	8.40 (4.13)						
T2	42	6.38 (4.83)						
CES-D[i]			4.3	40	<.001	0.74	0.30 to 1.19	18.2
T1	41	21.59 (6.42)						
T2	41	17.66 (7.53)						
PHQ-9[j]			5.54	40	<.001	0.99	0.53 to 1.45	33.15
T1	41	10.83 (3.92)						
T2	41	7.24 (4.83)						

[a]completers only.

[b]df: degree of freedom.

[c]HAM-A: Hamilton Anxiety Rating Scale.

[d]GAD-7: Generalized Anxiety Disorder–7 item.

[e]BAI: Beck Anxiety Inventory.

[f]PAS: Panic and Agoraphobia Scale.

[g]SPS: Social Phobia Scale.

[h]QIDS-C: Quick Item Inventory of Depressive Symptomatology.

[i]CES-D: Center for Epidemiological Studies Depression Scale.

[j]PHQ-9: Patient Health Questionnaire–9 item.

Table 4. Diagnostic status depression and anxiety.

Clinical disorder assessed by the MINI[a] and assessment point	Individuals with clinical diagnoses of valid responses n/N (%)
Generalized anxiety disorder	
T1	20/49 (41)
T2	6/39 (15)
Social phobia	
T1	19/49 (39)
T2	8/39 (21)
Agoraphobia without panic	
T1	12/49 (24)
T2	10/39 (26)
Panic with agoraphobia	
T1	6/49 (12)
T2	2/39 (5)
Panic without agoraphobia	
T1	4/49 (8)
T2	0/39 (0)
Major depressive disorder	
T1	0/49 (0)
T2	3/42 (7)
Any anxiety disorder	
T1	49/49 (100)
T2	18/39 (46)
Two or more anxiety disorders	
T1	9/49 (18)
T2	7/39 (18)
Three or more anxiety disorder	
T1	3/49 (6)
T2	1/39 (3)
Subclinical depression [b] CES-D [c] ≥16	
T1	41/49 (84)
T2	20/41 (49)

[a]MINI: Mini International Neuropsychiatric Interview.

[b]Subclinical depression subgroup (CES-D ≥16) assessed by Center for Epidemiological Studies Depression Scale.

[c]CES-D: Center for Epidemiological Studies Depression Scale.

Any type of positive training effect was mentioned in 84% (36/43) of interviews while any type of negative training effect was identified in 30% (13/43). Only 2 interviews with negative effects had no mention of any positive effects; in one of these interviews, training discontinuation was mentioned.

Positive training effects stated were improvement of disease burden (n=26) including general improvement of disease burden (n=12), feeling of increased performance (n=3), improvement of depressive symptoms (n=2), less rumination (n=2), improvement of psychosomatic pain (n=1), reduction of suicidal and self-injurious thoughts (n=1), fewer panic attacks (n=1), less tension (n=1), more calmness (n=1), reduction of alcohol consumption (n=1), and improved sleep quality (n=1); attentiveness to feelings and risk situations (n=24); confrontation with one's situation (n=20) including acceptance of oneself and others (n=8), focus on important areas of life (n=4), improvement of self-worth (n=2), knowing that one's situation can change (n=2), preoccupation with oneself (n=1), proud of one's achievements (n=1), and excited for future changes (n=1); insights and suggestions (n=12); more awareness for positivity and increased gratitude (n=6); and helpful entry to psychological treatment (n=1). Satisfaction with the online treatment was

categorized as online treatment helpful (n=5), fulfilled expectations (n=6), excited about online treatment (n=6), and online treatment not helpful (n=4).

Negative training effects entailed lack of change in disease burden (n=11); symptom deterioration (n=9) including increased hopelessness (n=5), increased rumination (n=2), social withdrawal due to tension (n=1), general symptom deterioration (n=1); and training discontinuation (n=1).

Intervention Use

In total, 98% (48/49) of participants completed the first session, and 65% (32/49) completed all 7 sessions. The booster session was completed by 55% (27/49). On average, participants took on average 9.44 (SD 3.78) weeks (range 4-18 weeks) to complete the intervention. See Table 5 for an overview.

Table 5. Completion rates.

Completion	Participants who completed session, n (%)	Participants who did not complete session, n (%)
Session 1	48 (98)	1 (2)
Session 2	44 (90)	5 (10)
Session 3	42 (86)	7 (14)
Session 4	40 (82)	9 (18)
Session 5	38 (78)	11 (22)
Session 6	33 (67)	16 (33)
Session 7	32 (65)	17 (35)
Booster session	27 (55)	22 (45)

In the noncompleter group, adherence intervention completion ranged from 0 to 3 sessions, while in the completer group, the average completion rate was mean 7.15 (SD 1.42) sessions.

Of the 38 participants who completed the fifth training session, 58% (22/38) chose to practice exposure to fear-inducing situations, while the other 42% (16/38) chose to practice problem-solving skills.

Regarding Tiny Tasks, only 10% (5/48) of individuals opted to not receive them, while 58% (28/48) chose to receive the light version with 3 daily reminders and motivational tasks and 31% (15/48) opted for the intense version with 5 messages per day.

Motivation and Expectations

When asked in the online-based baseline assessment which type of guidance participants would like to receive, 74% (36/49) stated they would like guidance and feedback on completed training session, no one said they did not want guidance, and 27% (13/49) claimed they had no preference concerning guidance.

In the baseline assessment we asked the participants to select reasons, from predefined categories, why they wanted to participate in the online training. Almost all participants (47/49, 96%) selected the answer “I want to learn to cope with my complaints autonomously.” Approximately 67% (33/49) claimed that they found online training appealing and 31% (15/49) said that waiting times for psychotherapy are too long. For a complete overview see Multimedia Appendix 1.

There were 12 reasons associated with training motivation identified in the qualitative interviews: advantages of online treatment (n=38) including active self-help (n=29), time and place independent flexible use (n=4), anonymity and to not have to conduct face-to-face conversations (n=4), and something beyond self-help (n=1); symptom burden (n=29) including symptoms of anxiety and depression (n=18), not able to deal with one’s situation autonomously (n=5), unhappy with current life situation (n=2), sleep problems (n=2), loneliness (n=1), and feeling of putting burden on family (n=1); openness toward online treatment (n=12); desire for improvement (n=9); no expectations toward the online treatment (n=7); stressful life event (n=4); desire to better understand situation (n=3); negative psychotherapy experience (n=3); no face-to-face psychotherapy possible (n=3); heightened expectation of improvement by participation (n=3); interest in psychology and mental health (n=2); and positive experience with self-help (n=1). For a complete overview see Multimedia Appendix 1.

Helpful and Hindering Factors

In total, 16 helpful factors and 10 hindering factors were identified. Of all interviews, 98% (42/43) had some mention of helpful factors and 74% (32/43) had some mention of hindering factors.

Helpful factors encompassed psychoeducation (n=14); support (n=20) including support by an eCoach (n=13), reminder emails (n=4), app notifications (n=2), and diagnostic interview (n=1); practice strategies in daily life (n=9); the structure of the program (n=8); relatable stories of testimonials (n=7); practicing thought protocol (n=7); planning activities (n=7); write down problems (n=7); confrontation with personal needs and values (n=7); elective modules (n=6); focus on personal situation (n=5); individual tailoring (n=4); neutral perspective on situations (n=3); problem solving (n=3); concrete instructions (n=2); and strategy collection (n=1). When individuals mentioned change but also stated that it could not directly be traced back to the training, it was categorized as other reasons for the change in disease burden (n=5).

Hindering factors comprised too little individualization of intervention (n=29) including too standardized (n=12), online treatment not sufficient (n=9), no feedback to specific inquiries (n=5), and too little personal contact (n=3); being overwhelmed

by the content and pace (n=15); usability issues (n=8) including limited functionality of the app (n=6) and limited usability of weekly activity plan (n=2); difficulties doing exercises (n=7); motivational difficulties (n=6); difficulties to plan (n=3); not open to training elements (n=1); needs beyond the scope of the training (n=1); and stress (n=6).

Modification Requests

There were 10 types of modification requests made, which were the desire for more intense support and more individualized feedback (n=10); longer treatment duration or more time to complete a module (n=9); exchange options with other participants (n=3); have limits of online treatment stressed (n=3); first aid plan (n=2); clearer structure of the activity plan (n=2); more printable content (n=2); more support to enhance motivation (n=2); more alternatives after having tried exercises (n=1); and be able to share content with friends and family (n=1).

Discussion

Principal Findings

The aim of this pilot feasibility study was to investigate an individually tailored transdiagnostic guided internet intervention for anxiety disorders with and without comorbid subclinical depression with an embedded qualitative and quantitative process evaluation.

Overall, the intervention was found to be feasible and results indicate potential efficacy concerning the improvement in anxiety and depression symptom severity. Moderate to large within effect sizes were found for anxiety and moderate effects for depression severity on all assessment scales apart from the PAS. Another finding in favor of the potential of the intervention is that while all individuals had an anxiety disorder at baseline, we found that the overall rate of anxiety disorders decreased by more than half. Positive effects stated by participants were general improvement of disease burden including improvement of anxiety and depression symptom severity and feeling equipped to deal with risk situations in the future. These findings are in line with previous work indicating that internet interventions are effective in treating anxiety disorders [1-3] and different anxiety disorders and comorbid depressive symptoms can be addressed transdiagnostically in one intervention [27,53-55].

Adherence to the intervention was found to be satisfactory with 65% of participants completing the intervention as intended, which lies just below some findings of adherence of internet-based guided self-help for anxiety disorders with 75% in a tailored group and 70.5% in a standardized treatment group [25]. The most prevalent reasons for participation found were advantages of the online treatment, symptom burden, and a general openness toward the online treatment. Participants found the most helpful factors to be the support provided, the psychoeducation, and being taught and encouraged to practice strategies in their daily routines.

Certain factors were perceived as hindering, such as the treatment not being individualized enough, at time they felt overwhelmed by the content and pace, and there were some usability issues. Some individuals struggled with motivation, regularly practicing, and integrating exercises into their daily lives, and others perceived stress outside of the intervention to be negative toward change. Many individuals also expressed the wish to have more contact with their eCoach going beyond the written messages and would have liked to have a personal conversation with their eCoach.

We also identified some negative effects. Although the quantitative data clearly showed an average improvement of anxiety and depression, the qualitative data revealed some negative effects such as an experienced lack of change in disease burden and symptom deterioration. Some individuals had heightened expectations of improvement prior to the treatment, which might be linked to greater disappointment and hopelessness if the treatment was perceived as ineffective. The negative effects were similar to what has been found in previous work, such as the occurrence of symptom deterioration or the emergence of novel symptoms [56]. When considering these negative effects, it is important to keep in mind that the individuals often mentioned negative and positive effects in one interview. This finding indicates that although many individuals improved in many different areas of life, it is possible that they did not improve in all the areas they would have liked to, it was not as effective as they expected, or they were not able or willing to put in the work to achieve the change they would have liked to experience. Although the number of negative effects mentioned was much less pronounced than positive treatment impact, this finding indicates the importance of exploring the use of methods beyond quantitative data such as qualitative data, as it can provide a more nuanced insight into user experience.

Strengths and Limitations

This study has the following strengths: the combination of qualitative and quantitative methods; having a total of 43 independent voices included in the qualitative data analysis; conducting clinical interviews to assess diagnostic status; the high standard of conducting the qualitative analysis; the following of current recommendations in the field; and the high interrater reliability between coders.

This study also has some limitations. As this was an uncontrolled pilot feasibility study with an intervention group and no control group, there was an explorative analysis of only within pre-post data and there was no actual hypothesis testing; also, we did not apply any techniques to estimate missing data. The already small sample size was further decimated by study dropouts. Although we completed some statistical analyses of the quantitative data, the results should be interpreted with caution as the findings are based on a very small, self-selected, completer data sample. This should be kept in mind when regarding the quantitative findings. After investigating baseline symptom severity differences between assessment completers and noncompleters, we saw that noncompleters had higher symptom severity of anxiety on the BAI on average; therefore, it is possible that the effects based on completers only are an overestimation. Furthermore, individuals with a major depressive disorder at baseline were excluded from the study during the screening process to increase internal study validity.

As participants self-selected to participate in the study, which is a typical occurrence in studies conducted in general population samples, this was an investigation of a highly selective population of individuals with an anxiety disorder, and findings from the qualitative data might not be generalizable to other populations. Also, the qualitative interview was completed after the intervention phase; therefore, questions on motivation and expectations were assessed in a retrospective manner and might be biased due to memory errors and time lapse. Last, it is possible that attrition rates are higher than they would be in a natural setting as the specificities of the study design such as diagnostic interviews might positively influence adherence.

Learnings and Future Research

Our learnings for future studies when targeting individuals with anxiety disorders are that it is important to emphasize the self-help aspect of internet interventions, replace clinical and disorder-specific terms, highlight the flexible use options, and provide detailed information on what individuals can expect.

Although the intervention was heavily individualized and contained many elements of individual tailoring, participants still preferred to have the individualization options increased. First, future research should focus on the dose-response rate posing the question how flexible a treatment can be while still producing a positive outcome. Second, future research should investigate whether further individualization also increases initial acceptance and willingness to participate, which might increase the effects on a population level. Individual tailored interventions for anxiety and depression should also be systematically compared with standardized disorder specific treatment in one study.

Although not all individuals experienced negative effects, it still seems especially relevant to address negative effects due to their possible impact. If an individual who suffers from a mental disorder and has taken the step to seek treatment has a negative treatment experience, it could cause training discontinuation, detraction from seeking further psychological treatment, and chronification of symptomology, which is why it is important to intervene in a timely manner. One first step to address this would be to manage realistic expectations before and throughout the intervention, clarify possible limits of an online treatment, and engage or refer individuals after an intervention. To tackle issues of motivation and difficulties integrating exercises into daily life, it could be helpful to involve smartphones more often into the treatment protocol as they could function as an extension of the treatment into individuals' private lives and are practical to deliver reminders. Future interventions should create a well-rounded support and guidance system that includes guidance adaptable in intensity combined with adherence and symptom monitoring.

There might also be other indicators of differential treatment outcome such as symptom severity, previous experience with mental health treatment, fear of stigma, and expectation of improvement, factors that could explain who benefits from internet interventions and who does not. For this reason, it is important to not only understand mediators of treatment but also investigate moderators of treatment and the combination of both. We also believe qualitative research should be used to understand why some individuals do not respond to treatment.

Concerning future qualitative research, it would be interesting to use machine learning techniques to analyze data. This could be done by analyzing content and words used by patients (eg, how someone speaks about themselves and their progress) as well as features of speech such as coherence, intonation, amplitude, pitch, and timbre. In addition to using these features to investigate user experience and impact, they might also be useful as an additional outcome assessment.

Conclusion

The investigated individually tailored transdiagnostic guided internet intervention seems to be feasible and indicated potential to reduce anxiety and depression severity. The results suggest that when targeting individuals for this type of treatment, it can be helpful to emphasize the active self-help components in addition to the advantages of online treatments. The content should contain psychoeducation, emphasize practicing strategies in daily life, and be complemented by a support system that entails some type of guidance as well as adherence and symptom monitoring. Once thresholds for low adherence or heightened symptoms are crossed, mechanisms should be set in place to either adapt the intervention or guide individuals to further treatment. Further individualization of interventions should be explored to best adapt to patients' characteristics, needs, and preferences.

Acknowledgments

We would like to thank Laura Kursawe, Mária Németh-Czóka, Aylin Gögsen, and Hanna Stüber who helped in the collection and preparation of the data. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 634757. The funder was not involved in the study design, data collection, outcome assessments, or writing of the publications.

Conflicts of Interest

DDE and MB are stakeholders of the GET.ON Institute for Online Health Trainings [57] which aims to transfer scientific knowledge related to this research into routine health care. The foundation of such an institute which disseminates research findings and products developed within research projects was the primary aim of the European Union for funding the associated research project (EU EFRE; ZW6-80119999, CCI 2007DE161PR001).

Multimedia Appendix 1
Supplementary tables.
[PDF File (Adobe PDF File), 257 KB - jmir_v22i9e16450_app1.pdf]

References

1. Richards D, Richardson T, Timulak L, McElvaney J. The efficacy of internet-delivered treatment for generalized anxiety disorder: a systematic review and meta-analysis. Internet Interv 2015 Sep;2(3):272-282. [doi: 10.1016/j.invent.2015.07.003]
2. Kampmann IL, Emmelkamp PMG, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord 2016 Aug;42:71-84. [doi: 10.1016/j.janxdis.2016.06.007] [Medline: 27376634]
3. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
4. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
5. Loebach Wetherell J, Thorp SR, Patterson TL, Golshan S, Jeste DV, Gatz M. Quality of life in geriatric generalized anxiety disorder: a preliminary investigation. J Psychiatr Res 2004 May;38(3):305-312. [doi: 10.1016/j.jpsychires.2003.09.003]
6. Stein MB, Heimberg RG. Well-being and life satisfaction in generalized anxiety disorder: comparison to major depressive disorder in a community sample. J Affect Disord 2004 Apr;79(1-3):161-166. [doi: 10.1016/s0165-0327(02)00457-3]
7. Carter RM, Wittchen H, Pfister H, Kessler RC. One-year prevalence of subthreshold and threshold DSM-IV generalized anxiety disorder in a nationally representative sample. Depress Anxiety 2001;13(2):78-88. [doi: 10.1002/da.1020]
8. Bromberger J, Schott L, Kravitz H, Joffe H. Risk factors for major depression during midlife among a community sample of women with and without prior major depression: are they the same or different? Psychol Med 2015 Jun;45(8):1653-1664 [FREE Full text] [doi: 10.1017/S0033291714002773] [Medline: 25417760]
9. Hölzel L, Härter M, Reese C, Kriston L. Risk factors for chronic depression: a systematic review. J Affect Disord 2011 Mar;129(1-3):1-13. [doi: 10.1016/j.jad.2010.03.025]
10. Jacobi F, Wittchen H, Holting C, Höfler M, Pfister H, Müller N, et al. Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). Psychol Med 2004 May 21;34(4):597-611. [doi: 10.1017/S0033291703001399] [Medline: 15099415]
11. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
12. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
13. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
14. Coles ME, Coleman SL. Barriers to treatment seeking for anxiety disorders: initial data on the role of mental health literacy. Depress Anxiety 2010 Jan;27(1):63-71. [doi: 10.1002/da.20620]
15. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015;351:h5627 [FREE Full text] [Medline: 26559241]
16. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565. [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
17. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. Eur Psychol 2018 Apr;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
18. Barlow DH, Allen LB, Choate ML. Toward a unified treatment for emotional disorders—republished article. Behav Ther 2016 Nov;47(6):838-853. [doi: 10.1016/j.beth.2016.11.005] [Medline: 27993336]
19. Craske MG. Transdiagnostic treatment for anxiety and depression. Depress Anxiety 2012 Sep;29(9):749-753. [doi: 10.1002/da.21992] [Medline: 22949272]
20. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided Internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
21. Schröder J, Jelinek L, Moritz S. A randomized controlled trial of a transdiagnostic Internet intervention for individuals with panic and phobias: one size fits all. J Behav Ther Exp Psychiatry 2017 Mar;54:17-24. [doi: 10.1016/j.jbtep.2016.05.002] [Medline: 27227651]

22. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
23. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012 May;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]
24. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. Cogn Behav Ther 2011 Mar;40(1):57-64. [doi: 10.1080/16506073.2010.529457] [Medline: 21337215]
25. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
26. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthan E, et al. Individually-tailored, Internet-based treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2011 Jan;49(1):18-24. [doi: 10.1016/j.brat.2010.10.002]
27. Păsărelu CR, Andersson G, Bergman NL, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
28. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
29. Richards D, Dowling M, O'Brien E, Viganò N, Timulak L. Significant events in an Internet-delivered (Space from Depression) intervention for depression. Couns Psychother Res 2017 Sep 13;18(1):35-48. [doi: 10.1002/capr.12142]
30. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Interv 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
31. Weisel KK, Zarski A, Berger T, Krieger T, Schaub MP, Moser CT, et al. Efficacy and cost-effectiveness of guided and unguided internet- and mobile-based indicated transdiagnostic prevention of depression and anxiety (ICare Prevent): a three-armed randomized controlled trial in four European countries. Internet Interv 2018 Apr:1. [doi: 10.1016/j.invent.2018.04.002]
32. Weisel KK, Zarski A, Berger T, Schaub MP, Krieger T, Moser CT, et al. Transdiagnostic tailored internet- and mobile-based guided treatment for major depressive disorder and comorbid anxiety: study protocol of a randomized controlled trial. Front Psychiatry 2018;9:274 [FREE Full text] [doi: 10.3389/fpsyt.2018.00274] [Medline: 30022954]
33. Sheehan D, Janvas J, Baker R, Sheehan KH, Knapp E, Sheehan M. Mini International Neuropsychiatric Interview (MINI) English Version 5.0.0. Tampa: University of South Florida; 2006.
34. Sheehan DV, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34 [FREE Full text] [Medline: 9881538]
35. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [Medline: 13638508]
36. Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, et al. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). Depress Anxiety 2001;13(4):166-178. [Medline: 11413563]
37. Reilly TJ, MacGillivray SA, Reid IC, Cameron IM. Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: a systematic review and meta-analysis. J Psychiatr Res 2015 Jan;60:132-140. [doi: 10.1016/j.jpsychires.2014.09.008] [Medline: 25300442]
38. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biological Psychiatry 2003 Sep;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8]
39. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
40. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
41. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893-897. [doi: 10.1037/0022-006x.56.6.893]
42. Bandelow B. Assessing the efficacy of treatments for panic disorder and agoraphobia. II. The Panic and Agoraphobia Scale. Int Clin Psychopharmacol 1995;10(2):73-82. [doi: 10.1097/00004850-199506000-00003]
43. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [Medline: 9670605]
44. Rabung S, Jaeger U, Streeck U, Leichsenring F. Psychometrische Überprüfung der Social Phobia Scale (SPS) und der Social Interaction Anxiety Scale (SIAS) im stationären Setting. Diagnostica 2006 Jul;52(3):143-153. [doi: 10.1026/0012-1924.52.3.143]

45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
46. Lehr D, Hillert A, Schmitz E, Sosnowsky N. Screening depressiver Störungen mittels Allgemeiner Depressions-Skala (ADS-K) und State-Trait Depressions Scales (STDS-T). Diagnostica 2008 Apr;54(2):61-70. [doi: 10.1026/0012-1924.54.2.61]
47. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun;1(3):385-401.
48. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002;7(1):105-125. [doi: 10.1037/1082-989X.7.1.105]
49. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. Organ Res Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
50. Mayring P. Qualitative Inhaltsanalyse. Handb Qual Forsch der Psychol 2010:613. [doi: 10.1007/978-3-531-92052-8_42] [Medline: 355418924]
51. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
52. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
53. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
54. Reinholt N, Krogh J. Efficacy of transdiagnostic cognitive behaviour therapy for anxiety disorders: a systematic review and meta-analysis of published outcome studies. Cogn Behav Ther 2014;43(3):171-184. [doi: 10.1080/16506073.2014.897367] [Medline: 24646219]
55. Newby JM, Mewton L, Andrews G. Transdiagnostic versus disorder-specific internet-delivered cognitive behaviour therapy for anxiety and depression in primary care. J Anxiety Disord 2017 Mar;46:25-34. [doi: 10.1016/j.janxdis.2016.06.002] [Medline: 27396841]
56. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
57. GET.ON Institute for Online Health Trainings. URL: www.geton-institut.de [accessed 2020-08-30]

Abbreviations

BAI: Beck Anxiety Inventory
CES-D: Center for Epidemiological Studies Depression Scale
COREQ: consolidated criteria for reporting qualitative research
GAD: Generalized Anxiety Disorder–7 item
HAM-A: Hamilton Anxiety Rating Scale
MINI: Mini International Neuropsychiatric Interview
PAS: Panic and Agoraphobia Scale
PHQ-9: Patient Health Questionnaire–9 item
QIDS-C: Quick Item Inventory of Depressive Symptomatology clinician rating
SPS: Social Phobia Scale

Edited by G Eysenbach; submitted 01.10.19; peer-reviewed by K Tomoko, P Romero-Sanchiz, C Santos; comments to author 30.12.19; revised version received 20.02.20; accepted 22.06.20; published 16.09.20.

Please cite as:
Weisel KK, Zarski AC, Berger T, Krieger T, Moser CT, Schaub MP, Görlich D, Berking M, Ebert DD
User Experience and Effects of an Individually Tailored Transdiagnostic Internet-Based and Mobile-Supported Intervention for Anxiety Disorders: Mixed-Methods Study
J Med Internet Res 2020;22(9):e16450
URL: http://www.jmir.org/2020/9/e16450/
doi:10.2196/16450
PMID:32936085

Original Paper

Designing a Web-Based Psychological Intervention for Patients With Myocardial Infarction With Nonobstructive Coronary Arteries: User-Centered Design Approach

Sophia Monica Humphries[1], MSc; Elisabet Rondung[2], PhD; Fredrika Norlund[1], PhD; Örjan Sundin[2], PhD; Per Tornvall[3], PhD; Claes Held[4], PhD; Jonas Spaak[5], PhD; Patrik Lyngå[3], PhD; Erik M G Olsson[1], PhD

[1]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[2]Department of Psychology and Social Work, Mid Sweden University, Östersund, Sweden

[3]Department of Clinical Science and Education, Karolinska Institute, Stockholm, Sweden

[4]Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden

[5]Department of Clinical Sciences, Danderyd Hospital, Karolinska Institute, Stockholm, Sweden

Corresponding Author:
Sophia Monica Humphries, MSc
Department of Women's and Children's Health
Uppsala University
Akademiska Hospital
Uppsala, 75185
Sweden
Phone: 46 760341186
Email: sophia.humphries@kbh.uu.se

Abstract

Background: The involvement of patient research partners (PRPs) in research aims to safeguard the needs of patient groups and produce new interventions that are developed based on patient input. Myocardial infarction with nonobstructive coronary arteries (MINOCA), unlike acute myocardial infarction (MI) with obstructive coronary arteries, is presented with no significant obstructive coronary artery disease. Patients with this diagnosis are a subset of those diagnosed with traditional MI and often need more psychological support, something that is presently not established in the current treatment scheme in Swedish health care or elsewhere, to our knowledge. An internet-delivered intervention might offer patients with MINOCA the opportunity to access a psychological treatment that is tailored to their specific needs after MINOCA and could therefore supplement the existing medical care in an easily accessible format.

Objective: This paper aims to describe the development of a therapist-guided, internet-delivered psychological intervention designed specifically for patients with MINOCA.

Methods: The study used a participatory design that involved 7 PRPs diagnosed with MINOCA who collaborated with a team consisting of researchers, cardiologists, and psychologists. Intervention content was developed iteratively and presented to the PRPs across several prototypes, each continually adjusted and redesigned according to the feedback received. The intervention and experience of it were discussed by PRPs in a final meeting and then presented to a panel of 2 clinical psychologists and a cardiologist for further input.

Results: The outcome of the collaboration between PRPs and the research group produced a web-based psychological 9-step program focusing on stress, worry, and valued action. The input from PRPs contributed substantially to the therapy content, homework tasks, interactive activities, multimedia, and design presentation.

Conclusions: Working with PRPs to develop an intervention for people with MINOCA produced a web-based intervention that can be further evaluated with the goal of offering a new psychological treatment option to a patient group currently without one. Direct contribution from PRPs enabled us to obtain relevant, insightful, and valuable feedback that was put towards the overall design and content of the intervention.

(J Med Internet Res 2020;22(9):e19066) doi:10.2196/19066

KEYWORDS

web-based intervention; iCBT; myocardial infarction; nonobstructive coronary arteries; patient involvement; psychological treatment; MINOCA; takotsubo cardiomyopathy

Introduction

Background

Myocardial infarction with nonobstructive coronary arteries (MINOCA) comprises a subset of patients with myocardial infarction (MI) [1-3]. Depending on the defining criteria of nonobstructed arteries, prevalence lies somewhere between 1% and 14% of all patients with MI [4-6]. Around 50% of these patients are further classified as having takotsubo cardiomyopathy (TTC) and therefore constitute a large proportion of the literature regarding patients with MINOCA. MINOCA is briefly defined as an acute MI that presents on coronary angiography with no significant obstructive coronary artery disease (CAD). The distinction between TTC and other types of MINOCA can be challenging, but TTC is a condition more distinctly characterized by a sudden temporary weakening of the muscular portion of the heart. The remaining patients are often grouped into various subcategories, such as plaque and nonplaque mechanisms, or even "undefined" [7]. A detailed description of diagnostic procedures was recently published [4].

Diagnosis and prognosis separate patients with MINOCA from patients with MI due to obstructive CAD [8], as do patient characteristics showing that MINOCA occurs predominantly in women and at a younger age than MI due to obstructive CAD. Around 30% of all patients with MINOCA have had a previous diagnosis of psychiatric illness [9], and more than half of patients with MINOCA report some kind of emotional or physical stressor prior to the hospital admission compared with matched patients with CAD, who report a stressor in less than 15% of cases [10]. The risk of experiencing psychological distress, such as symptoms of anxiety or depression, after MI due to obstructive CAD is already reportedly quite high compared with controls [11], and patients with MINOCA report levels even higher [10].

For patients who have experienced MINOCA, the time following the acute event can be just as daunting and stressful as the event itself. Many report a low quality of life owing to this aftermath [12,13], and this coupled with the unclear etiology of the diagnosis can make the time following the acute event increasingly open to anxiety and depression. Qualitative interviews with patients with TTC have explored the experience of living with these problems, as well as the impact on long-term stress. A major finding has been that most, if not all, have described their lives as being lived under constantly stressful circumstances [14] and that they feel limited in their lives regarding future health or long-term recovery [15]. A new study into stress associated with TTC reported that more than half of the patients in the study were still on part-time sick leave 6 months post-onset of the diagnosis [16]. Patients also experienced more vulnerability to stress after the event, which consequently led to self-reported sleep disturbances, memory loss, and difficulties concentrating. Thus, the implications of MINOCA appear to encompass a lot more than the acute event itself.

Patients in this diagnostic group have reported that it can be challenging to comprehend having a diagnosis of MI, as well as being unfamiliar with the term MINOCA, or MI with "normal" arteries [15]. Health care staff often give vague answers or explanations of the cause, which in turn may trigger more anxiety [17]. The insufficient knowledge surrounding a diagnosis of TTC or MINOCA is a challenge in cardiac rehabilitation. Finding ways to manage the stress and psychological impact of the event, in addition to addressing any pre-existing problems that might have been present, are also important factors that should be targeted in cardiac rehabilitation for patients after MINOCA [9].

Cognitive behavioral therapy (CBT) is a longstanding and supporting therapy for treatment of mood and anxiety disorders, among many others. Guidelines from the National Institute for Health and Care Excellence (the so-called NICE guidelines) recommend CBT in the referral advice and treatment for subthreshold symptoms and mild to moderate common mental health disorders [18]. Support for using face-to-face CBT to treat symptoms of anxiety and depression already holds some benefits when compared with standard care, both in patients discharged after a coronary event [19] and with other somatic illnesses, such as chronic obstructive pulmonary disease [20]. According to a meta-analysis, psychological treatments produce lower mortality rates and reoccurrences in cardiac patients, although this effect has mainly been demonstrated in men [21]. CBT has successfully been converted to be provided online in the treatment of many mental health disorders. Studies have shown that using internet-based CBT (iCBT) to treat cases of mild anxiety and depression are just as effective as face-to-face therapy and hold many benefits over using traditional forms of CBT [22]. For those without easy access to mental health treatment, iCBT is an acceptable mode of treatment, even at long-term follow-up [23]. In the case of diverse groups, such as cardiac patients and stress-related disorders, iCBT appears to be better at lowering stress-related outcomes than its active or waitlist control counterparts [24-26]. With regard to patient acceptability, a recent meta-analysis showed that iCBT for the treatment of depressive and anxiety disorders is highly effective in clinical practice and that patients reported high levels of satisfaction [27]. On the contrary, some studies have reported that patient acceptability of iCBT is lower when compared with face-to-face psychological therapy [28,29]. However, none have explored this in patients with a MINOCA diagnosis. Since only a handful of small studies have explored patient acceptability of iCBT in a patient population experiencing cardiac-related illness, there is arguably an increasing need for such initiatives [30,31]. Many studies also lack qualitative methods and detailed methodological reporting, as well as investigation of patient opinions of online therapy. This is especially the case for treatment of anxiety and depression among those with cardiac-related illnesses [32]. The lack of existing and

conclusive research findings into the psychological effects of MINOCA means that intervention development could benefit from involving patients in the process. Patient and public involvement (PPI) in the development of a new intervention is based on the argument that it brings out a product that is relevant to the population [33] and gathers new insights [34], which in turn should help to map potential problems before they arise in a trial setting.

There are different ways to include patients in the process of the early development of an intervention study. According to guidelines, when the patient group is underserved or the intervention is novel, particularly for a unique patient group, it is of great importance to involve patients directly by using a patient-centered approach in both the intervention and the study development [35]. Using PPI to improve future internet-based intervention studies involves patients in the development stage of the intervention rather than just in the piloting and feasibility stages and beyond [35]. The current literature on patients with MINOCA is fairly limited, particularly with regard to their psychological needs, and we know of no psychological treatment specific to patients with MINOCA. In the process of applying a user-centered design, we also aim to lower the likelihood of attrition to the final intervention with a program that is better tailored to the needs of patients with MINOCA. Therefore, the uniqueness of the patient group, the limited knowledge about their needs, and the nonexistent psychological treatment options are together strong arguments supporting the development and testing of an iCBT treatment through PPI methods.

Aims and Objectives

One aim of this paper is to outline how PPI has contributed to the development of an internet-based intervention for treating symptoms of psychological distress experienced by patients after MINOCA. In this sense, we argue for why it is important to work *with*, rather than just *for*, patients with MINOCA to avoid making assumptions about the intervention design or without justifiable inclusion of certain methods and content.

Another aim is to present the procedure in which we designed and built the intervention by sharing the development process and strategies used throughout. This includes how we used patient feedback throughout each step of the intervention, the suggestions for changes during developing the intervention, and how those modifications were incorporated into the intervention (or considered for future improvement). Challenges will also be discussed in relation to developing a complex web-based intervention with collaboration from patients.

Methods

Patient Participatory Methods and Recruitment

Patients were involved at the collaboration level described in PPI protocols [36], reflecting their ongoing and shared level of contribution to the intervention development, and they are referred to as patient research partners (PRPs). PRPs were considered active members of the wider research group; they were involved throughout the entirety of the development and participated in consultation and testing of developed material.

This study recruited participants from Södersjukhuset Hospital in Stockholm using convenience sampling methods. The study initially began in May 2018 with 3 PRPs who took part in the first focus group discussion. The remaining PRPs were recruited gradually over a 6- to 8-month period from the study start, leading to a total of 7 PRPs, although attendance of in-person meetings varied depending on participant availability for the given date and time.

Participants

A total of 7 adults (5 female) participated, all with a diagnosis of MINOCA within 5 years prior to the start of the study and who lived near or were able to commute to Stockholm and were able to read and understand Swedish. Of the 7 PRPs participating in the study, 2 had confirmed TTC and the remaining had unspecified diagnoses of MINOCA. The median age of the group was 58 years and the time since first diagnosis to study inclusion ranged from 2 weeks to 5 years. Informed consent was obtained prior to the formal development process. All agreed to participate voluntarily, to speak about their experience and the psychological impact of MINOCA on their health, and to actively test the intervention as it was developed while providing feedback and insight. Ethics approval was granted by the Regional Ethical Review Board in Stockholm (Dnr 2018/1434-31/1 and 2018/2406). Approval was obtained for the use of the image in this paper by the participants.

Project Management

The research team consisted of 4 clinical psychologists (coauthors ER, FN, ÖS, and EMGO), 3 cardiologists (coauthors CH, PT, and JS), 1 cardiac nurse (PL), and 1 PhD student with a background in psychology (SMH). The clinical background of the group was thus within the fields of psychology and cardiology. In addition to these members, web developers were also involved in the digitalization process and in implementing design features on the web platform. The research team decided on a set of requirements to guide the development a priori and kept these in focus throughout the development (see Textbox 1). These criteria were mainly for the team's own use and have not been formally evaluated but will be reflected in the planned feasibility and randomized controlled trial (ClinicalTrials.gov NCT04178434). The core research team worked according to a schedule decided at the start of the development process and the intervention was finalized towards the end of 2019. Throughout this process, web-based development group meetings took place weekly and included the research members involved in creating the intervention. These often included discussions over content, exercises, and visual components. The creation of the digital intervention content began January 2019 and was split across 5 prototype phases, originally planned as 4 but extended due to time constraints and participant feedback. All written components of the treatment were initially created in Word documents (Microsoft Corp) and then transferred to be accessed via the Uppsala University Psychosocial Care Programme (U-CARE) portal, an online platform that enables delivery of psychological interventions over the web (used, for example, in the U-CARE heart study [37]). A more detailed description of the process is described in the upcoming sections.

Textbox 1. A priori intervention requirements.

1. The intervention should be intended to suit the majority of patients with myocardial infarction with nonobstructive coronary arteries (MINOCA) who have psychological distress.
2. The intervention should aim at reducing the psychological suffering that is common among these patients.
3. The intervention should be accessible, preferably an eHealth intervention.
4. The intervention should, at least to some extent, be flexible and possible to adopt to new technological developments.
5. After being developed, the intervention should be a low-cost intervention (less costly than a face-to-face intervention).
6. The intervention should be possible to implement and available to most patients with MINOCA in Sweden early after the cardiac event (implementation is not part of the project, but implementation should be kept in mind).
7. The intervention should not be longer than 3 months, preferably shorter.
8. The intervention should be possible to evaluate in randomized trials.
9. The intervention should be mapped or described in a detailed way, making it possible to replicate. (This means that all interventions should be linked to a behavior change mechanism that in turn is related to desired outcomes.)
10. The intervention should be developed in collaboration with potential end users (ie, patients with MINOCA).

Based on several focus group discussions (FGDs) with patients with MINOCA, scoping searches of the literature, and input from members of the core study group with experience working with patients with MINOCA or patients with associated symptomatology of mental disorders, the treatment targets were established. These helped inform the basis of the intervention, in combination with the prespecified requirements previously mentioned.

The early-stage FGDs focused on personal experiences and in turn provided in-depth material for a needs assessment, something that is recommended as the first step when designing an intervention, according to intervention mapping approaches [38]. This needs assessment worked to analyze the problems to identify what areas of change the intervention should target in patients with MINOCA. Furthermore, with the strong evidence base for CBT in mind, it was deemed appropriate to apply the fundamental and theoretical model of CBT for treating stress and anxiety in patients with MINOCA.

Intervention Development Procedure

The research group worked from a development plan that used a stepwise approach throughout the continued development of the intervention. The process of involving PRPs and prototyping took inspiration from the concept of user-centered design (UCD). The UCD approach views end users as pivotal to the design process of product development and is one way to avoid a one-directional relationship between the developers and the product, which often occurs when creating evidence-based treatments without end user involvement in the design process [39].

Early-Stage Focus Group Discussions

The first instances of PRP involvement included several of the PRPs who met with 3 of the researchers (PL, ER, and FN) to discuss their experiences of MINOCA in a broad manner. A total of 4 group-style discussions took place during a period of 6 months, and the number of PRPs attending each discussion varied depending on their availability and the number of participants recruited at that point in the study. The idea of these discussions was to gather as much information as possible about (1) the problems or psychological suffering encountered as a result of the MINOCA; (2) the type of help patients would like to receive if offered to them in a digital format; (3) the format, length, features, and possible supplementary material best suited for the intervention; and (4) their opinions on relevant outcomes and specific psychological evaluation questionnaires.

Development and Presentation of Material

Following the FGDs, the process of building the intervention began with creating the treatment content. This was done by the psychologists (coauthors ER and FN) and was developed over a period of roughly 4 to 6 weeks per prototype. Content was then uploaded online and added to a demo study in the U-CARE portal. In the initial stage, PRPs were given user log-ins that enabled them to access the same account throughout the testing. A period of approximately 10 to 14 days was allotted to the PRPs to work through the newly developed content during each testing stage. The idea was to allow enough time to actually complete the homework tasks that accompanied the content, read through the text, and look through the media content. Navigating the portal website and sending instant messages were also features the PRPs were encouraged to test. Shortly after working through the material and testing the features of the portal, PRPs were contacted by the psychologist assigned to them (ER or FN). Brief one-on-one unstructured telephone interviews were conducted that focused on PRPs' feedback on the language, content, exercises, and experience of the portal, as well as anything else they could think of. At times it was difficult for the participants to be specific enough to advise concrete changes. This was expected, and researchers did their best to ask follow-up questions. The team then used this feedback to rebuild and redesign content, features, and other requirements for the specific content in question. This process was repeated for all stages of the intervention until completion. See Table 1 for an overview. This overall approach made the process an iterative one, as the constant redesign and evaluation of the intervention allowed for adjustments and continuous improvements before a final treatment manual was established.

Table 1. Process stages of the development of the intervention.

Step	Type	Content	Exercises	Attendees
Panel discussion	FGD[a] with patient research partners group	Open discussion, information gathering	None	2 researchers, 3 patient representatives
Panel discussion	FGD with patient research partners group	Open discussion, identification of problems	Asked to go through some questionnaires, including the CAQ[b]	3 researchers, 3 patient representatives
Panel discussion	FGD with patient research partners group	Open discussion, thoughts and feedback of material, and ideas for iCBT[c] intervention	Were presented with a brochure	3 researchers, 3 patient representatives
Panel discussion	FGD with patient participatory group	Open discussion, some testing of relevant material, feedback used to gauge usefulness	Reviewed material from an existing online intervention, tested a homework exercise on fear after a cardiac event, reviewed a video interview with a cardiologist talking about MINOCA[d]	2 researchers, 2 patient representatives
Portal introduction	Phone call with patient representatives	Instructions/introduction to online portal	Logging in to the portal, testing of user account	N/A[e]
Prototype 1	Online iCBT testing	Introduction and "Fear after MINOCA" content	Describe own experience of MINOCA	7 patient representatives invited, of which 6 logged in
Feedback	Telephone feedback interviews	Feedback via phone	N/A	6 patient representatives
Prototype 2	Online iCBT testing	Stressors and stress reactions	Listing own stressors and self-observation of stress reactions	6 patient representatives
Feedback	Telephone/mail feedback interviews	Feedback via phone	N/A	6 patient representatives
Prototype 3	Online iCBT testing	The importance of consequences, recovery, and relaxation	Trying alternative behaviors and relaxation exercises	6 patient representatives
Feedback	Telephone feedback interviews	Feedback via phone	N/A	6 patient representatives
Prototype 4	Online iCBT testing	Values	Formulating values and planning activities accordingly	6 patient representatives
Feedback	Telephone feedback interviews	Feedback via phone	N/A	6 patient representatives
Prototype 5	Online iCBT testing	Summary, evolution, and maintenance	Formulating a plan for relapse prevention and future development	4 patient representatives
PRP[f] panel meeting and interviews	Open-group interview with semistructured guide	Overall feedback on the intervention and the process of being part of a patient panel	N/A	2 researchers, 4 patient representatives
Expert panel meeting	Discussion of the intervention from the perspectives of psychologists and cardiologists with previous knowledge/experience	Current version of the intervention with all 9 steps and material	Going through content in own time, online and through the PDF handout version	2 researchers, 2 external psychologists, and 1 external cardiologist

[a]FGD: focus group discussion.

[b]CAQ: Cardiac Anxiety Questionnaire.

[C]iCBT: internet-based cognitive behavioral therapy.

[d]MINOCA: myocardial infarction with nonobstructive coronary arteries.

[e]N/A: not applicable.

[f]PRP: patient research partner.

Final Panel Meeting With PRPs

Two of the researchers (EMGO and SMH) interviewed PRPs with the purpose of gaining insight on not only the final intervention (content, exercises, aesthetics, among others) but also the experience of being involved in the process as well as any extra remarks that may have otherwise been missed. The translated interview guide can be found in Multimedia Appendix 1. This was followed by a dinner gathering with the whole

research group and patient research partners together, allowing for discussion to continue more informally.

Expert Panel Meeting

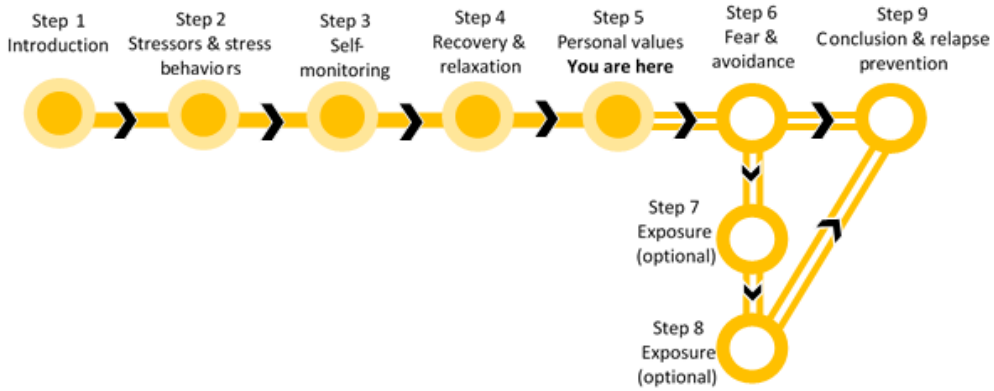
The final step in the development of the intervention included input from 3 experts who made up what we refer to as the expert panel. This consisted of external-only members, 2 psychologists and 1 cardiologist, and were consulted with the purpose of providing feedback from the perspectives of professionals working within their respective fields. The expert panel were sent the latest version of the full intervention manual in a written PDF format and provided with secure log-ins to the U-CARE portal as test users 1 week before the physical meeting. The panel met with 2 of the researchers from the core group (EMGO and SMH) to provide their feedback and comments on the intervention and content from the perspectives of the clinical psychology and cardiology fields.

Results

Final Intervention Design

The final intervention design resulted in a 9-step online psychological treatment for patients with MINOCA. These are outlined as (1) introduction, (2) stressors and stress behaviors, (3) stress-specific self-monitoring, (4) recovery and relaxation, (5) personal values, (6) fear and avoidance, (7) exposure, (8) exposure continued, and (9) conclusions and relapse prevention. The work-through rate was aimed at participants completing an average of 1 step per week, resulting in an estimated 2 to 3 hours of work per week, based on reports from the PRPs. Figure 1 shows the brief overview of the intervention as presented to the participants during treatment.

Figure 1. An overview of the steps in the intervention.



The first step includes an introduction, with an overview of the intervention and several video interviews. These were made specifically for the intervention to provide information that was lacking in the current care, as previously mentioned in the FGDs. These interviews focus on the physiology of MINOCA from a cardiologist's standpoint, a nurse talking about the acute phase of patient care, psychological reactions after MINOCA, and what it is like to experience MINOCA from a patient perspective. Stress was a commonly occurring topic in the FGDs, as well as being a well-known problem experienced by patients with MINOCA in the literature. Therefore, the next 2 steps cover stress-focused content and related exercises. Step 2 mainly focuses on psychoeducational aspects and on how to monitor stressors and stress situations in different contexts. Step 3 looks at how one can identify stress and its consequences. In Step 4, the importance of recovery and relaxation is introduced, including a small focus on sleep behavior. Step 5 encourages participants to focus on personal values relating to work, leisure time, health, and relationships. Step 6 introduces cardiac-related fear, worry, and safety behaviors. Step 7 and 8 are optional modules that focus mainly on exposure training, but participants are encouraged to work through them, particularly participants identifying high levels of cardiac-related anxiety in step 6. Step 9 is the last step of the program and begins with a summary and overview of the treatment. This step encourages maintenance of what has been learned during the treatment and further development beyond the end of treatment. Each step also contains exercises for participants to work through. Where the exercise requires written content or homework tasks, participants are given feedback from the psychologist, accessed from the online portal, a feature supported by research into digital CBT as facilitating homework completion [40]. Screenshots of the digital intervention can be seen in Multimedia Appendices 2 and 3.

Final Views of the Intervention

Findings were taken from the feedback provided at the final PRP meeting and based on the treatment version that the PRPs worked with up until that point.

Content

The final meeting with 4 members of the PRP group resulted in several small changes to the intervention. Participants expressed that it would be preferable if the heart-related fear steps, which were to be optional due to low activity in the testing, were instead mandatory additions to the first 5 steps, as the content was viewed as rather useful. With this in consideration, the first of the 3 cardiac-related fear steps was included as mandatory, while the 2 exposure steps were kept as optional. Implemented and attempted changes suggested by the PRPs in this meeting are presented later.

Platform Design

PRPs had some general agreements about the design features that should be included. They all agreed that the inclusion of pictures and images was necessary to make the intervention visually appealing and break up large blocks of text. They also welcomed the short video clips that featured certain scenarios that were used as examples during the treatment, and their positive feedback on the first clip even led to the inclusion of more clips throughout the treatment.

Terminology

PRPs expressed preference for the term MINOCA as opposed to something more general, such as "cardiac event," but expressed the importance of explaining the term in detail at the start of treatment. This was especially important because the PRPs did not all fall under the same category of MINOCA; some were classified as having TTC, for example. By using the broader term MINOCA, it ensures that the examples given in the text and exercises are less likely to be misunderstood as irrelevant or exclusive to those unfamiliar with the term. Similarly, nonmedical terms that are commonly used (such as "broken heart syndrome" instead of takotsubo cardiomyopathy) were not considered appropriate by the panel, and they expressed discontent with their use in most contexts. Figure 2 shows the final PRP feedback meeting.

Figure 2. The patient research partners meeting during the final feedback concerning the internet-based cognitive behavioral therapy intervention.

Expert Panel's View of the Intervention

Shortly after the final meeting with the PRPs, the expert panel was consulted for their views and advice on the intervention. Feedback revolved around the practical content, such as visual layout and design, as well as the treatment content. One suggestion from the expert panel was to include a visual aid to complement the progress of the participant as they work through the treatment steps. This was added to the intervention (Figure 1) and is changed with each additional step completed. Overall, the panel was enthusiastic and positive about the treatment program. Suggestions for change that were implemented upon receiving their feedback is also presented below (Table 2).

Table 2. Suggested changes made by patient research partners and the expert panel regarding the design and content of the intervention.

Group	Platform design		Treatment content	
	Identified suggestion	Modification	Identified suggestion	Modification
PRPs[a]	- Move the interview films - Include a chat function with the psychologist - Automatic message sent directly via email or SMS when something new is commented on in the portal	- Interview video content moved to introduction - Chat function considered but currently not implemented - Feature under construction, to be added	- Remove deadlines for homework tasks - Mandatory fear steps as opposed to only optional	- Replaced with recommended submission date instead - First cardiac-related fear step made mandatory, with 2 steps as optional
Expert panel	- Shorten interview video content - Add a follow-up video of patient interview - Visual representation of intervention progress - Captioning of videos	- Videos shortened - Video with patient with MINOCA[b] added to include an update posttreatment - Stepwise diagram created and added to each step - Text captions added to videos	- Include measurements every step/week of the treatment - Add sleep-related content - Give a clearer message about exercise and physical exertion - Exposure therapy content that focuses less on habituation	- GAD-2[c] and PHQ-2[d] questionnaire measurements added weekly - Addition of “things to think about” regarding sleep in step 4 - Short text added to clarify “mixed messages” around the benefits of exercise for those who can - The heart-related fear section slightly rewritten to be more in line with an inhibitory learning model of exposure [41]

[a]PRPs: patient research partners.

[b]MINOCA: myocardial infarction with nonobstructive coronary arteries.

[c]GAD-2: 2-item Generalized Anxiety Disorder scale.

[d]PHQ-2: 2-item Patient Health Questionnaire.

PRPs' Views Surrounding the Participatory Action Research Process

The final meeting with PRPs also included a short discussion about their participatory perspective of being involved in the development of an intervention. Many felt that being part of a panel contributing to the development of the intervention was a positive thing overall. They discussed that this process played a positive role in their lives and hopefully in the lives of other patients with MINOCA who will benefit from their input. Many expressed a motivation to contribute to better care for others, but also reported gaining some benefits to their own lives in being part of this process. Being surrounded by others with similar experiences was described by PRPs as something that led to a reduced feeling of loneliness, but they also reported that it exposed them to the fact that MINOCA affects others differently, something they saw as positive.

Discussion

Principal Results

The current paper describes the stepwise processes involved in the development of a web-based CBT program for patients with MINOCA experiencing psychological distress. Collaboration with PRPs and the expert panel resulted in a 6- to 9-week internet-delivered intervention with a CBT-focused approach aiming to reduce symptoms of high stress and anxiety. Continuous testing from PRPs, all of whom had received a diagnosis of MINOCA, helped to ensure that the intervention was relevant to the target patient group and hopefully more effective as a result.

Similar studies that have used PPI in intervention development have established beneficial impacts on the relevancy and acceptability of their interventions [42-44], and systemic reviews have also reported on the impact on study design and enhanced quality of research in general [45,46]. The outcome of our collaboration with 7 research partners enabled us to fulfill this aspect of PPI, hopefully producing an intervention better suited to the needs of the target user and that fulfilled the requirements shown in Textbox 1.

The process of using an iterative approach was the main advantage in the intervention creation. This meant that development was continuously modified so that each component was not finalized until participant feedback was provided. After feedback, it was fine-tuned even further, concluding with a final group meeting with the PRPs’ and the expert panel’s insight. This type of approach also ensured that the intervention was kept up to date during the development process, always tuned to the current needs or opinions of the patient group.

Participant feedback helped steer the direction of the development. While some treatment and design preferences might have appeared more in line with existing studies, the inclusion of a group of patients with a unique diagnosis brought some revelations to the surface that might not have otherwise been considered. A unique example of this is demonstrated in the use and consideration of certain terminology used to describe the diagnosis. PRPs expressed discontent with the term “broken heart syndrome” (often used, even among doctors), finding it to be downgrading or glamorizing the condition. A recent review of medical or nonmedical terms used to describe TTC supports this view and claimed that metaphorical language should be

discouraged while the pathophysiological and diagnostic knowledge is still being established [47].

A couple of limitations should, however, be noted. First, the team worked with an existing health care platform to provide the intervention, and while many features were added in response to the PRP's requests, it was not possible to cater to all requests, in part due to platform constraints. We could therefore only offer as much as the platform allowed. Some PRPs liked the idea of a video chat feature, for example. However, this feature was not possible to fully implement at the time of the feasibility testing phase, but it is on course to be available for later testing in a future randomized controlled trial.

Furthermore, while the work with PRPs ensured that patients with MINOCA were able to contribute their specific views and experiences towards the intervention, it should be noted that the transferability to current and future patients with MINOCA might be limited due to the sample being recruited from one main city region in Sweden. It should therefore be considered that across different hospitals in Sweden, the experience and views might differ with regard to treatment needs and rehabilitation. In addition, PRPs were included at different stages after the acute event, with one PRP having had MINOCA just a few weeks before inclusion, whereas another had a MINOCA 6 years prior. However, while the intervention will aim to be used just a few weeks after the MINOCA, putting the validity of the feedback from the PRPs with older MINOCA diagnoses into question, it is still felt to be valuable to get insight from PRPs who had MINOCA a long time ago.

Several challenges were inevitably experienced during the design process. As expected, the psychologists struggled at times to receive feedback that could be used to change the content more specifically. This mostly occurred when the PRPs did not feel like they related to that specific section of the intervention. However, since no group of patients is homogenous, it was not unexpected that the group would present with diverse problems and experiences. Scoring over a specific threshold on measures of stress, anxiety, or depression was not a requirement for the PRPs involved in the cocreation of the intervention, but it will be a prerequisite in the randomized control trial planned to follow. Therefore, some of the problems the intervention aimed to tackle might not have been personally problematic for the PRPs involved. Nevertheless, the PRPs agreed that the presented content and exercises under those circumstances were good to include and would be very helpful for the target user with MINOCA.

One future improvement feature to consider is individualization of the program for each participant. In the initial planning phases, we discussed the possibility of tailoring parts of the intervention, but this was deemed to result in a more complex intervention demanding too much of participants making informed choices and too difficult to construct. Since cardiac patients in previous internet-based interventions have described the treatment as burdensome and long [37,48], we aimed to produce a simple yet effective web-based treatment that would work for the target group and minimize dropout. However, individualized therapy in which participants can select the content they would like to work through themselves could be a future consideration for the mix of preferences patients with MINOCA might have with regard to psychological treatment. This was compensated for partly by offering the choice of the optional heart-related fear content, but future improvement could work towards making the intervention more customized. However, this intervention, as specified in the aforementioned a priori requirements (Textbox 1), aimed at offering treatment that can suit the majority of patients with MINOCA, something we believe was achieved. Therefore, customization of the intervention to a detailed degree could be something left to future studies.

Conclusion

The process of involving PRPs in the development of an internet-based psychological treatment provided substantial insights and ideas. These contributed substantially to what we believe is a treatment both relevant and helpful for patients with MINOCA who also report moderate levels of stress, depression, or anxiety following their diagnosis. The diversity of the group enabled a range of viewpoints to be represented, and PRPs expressed that being part of the panel was a positive experience. Moreover, including an expert panel to provide feedback and suggestions for improvement ensured that input was considered from researchers and clinicians outside of the core research group. Using a design process that structuralized components of the intervention into prototypes and built upon these in an iterative manner with constant PRP feedback resulted in an intervention that has been tailored to specific MINOCA patient groups in focus. The next step of evaluation will be through a subsequent feasibility analysis and randomized controlled trial to test the result of this design process and the intervention's effectiveness as a whole.

Acknowledgments

This study was funded by grants from the Swedish Research Council (Dnr 2018-02655) and the Swedish Heart and Lung Association. This study is also a part of U-CARE, a strategic research environment funded by the Swedish Research Council (Dnr 2009-1093).

The authors would like to thank the patient research partners for their invaluable participation and contribution throughout the entirety of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Translated interview guide from the final patient research partner meeting.
[DOCX File , 18 KB - jmir_v22i9e19066_app1.docx]

Multimedia Appendix 2
Screenshot of the intervention.
[PNG File , 93 KB - jmir_v22i9e19066_app2.png]

Multimedia Appendix 3
Screenshot of the intervention.
[PNG File , 69 KB - jmir_v22i9e19066_app3.PNG]

References

1. Larsen AI, Galbraith PD, Ghali WA, Norris CM, Graham MM, Knudtson ML, APPROACH Investigators. Characteristics and outcomes of patients with acute myocardial infarction and angiographically normal coronary arteries. Am J Cardiol 2005 Jan 15;95(2):261-263. [doi: 10.1016/j.amjcard.2004.09.014] [Medline: 15642564]
2. Widimsky P, Stellova B, Groch L, Aschermann M, Branny M, Zelizko M, et al. Prevalence of normal coronary angiography in the acute phase of suspected ST-elevation myocardial infarction: Experience from the PRAGUE studies. Canadian Journal of Cardiology 2006 Nov;22(13):1147-1152. [doi: 10.1016/s0828-282x(06)70952-7]
3. Patel MR, Chen AY, Peterson ED, Newby LK, Pollack CV, Brindis RG, et al. Prevalence, predictors, and outcomes of patients with non-ST-segment elevation myocardial infarction and insignificant coronary artery disease: results from the Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA Guidelines (CRUSADE) initiative. Am Heart J 2006 Oct;152(4):641-647. [doi: 10.1016/j.ahj.2006.02.035] [Medline: 16996828]
4. Agewall S, Beltrame JF, Reynolds HR, Niessner A, Rosano G, Caforio ALP, WG on Cardiovascular Pharmacotherapy. ESC working group position paper on myocardial infarction with non-obstructive coronary arteries. Eur Heart J 2017 Jan 14;38(3):143-153. [doi: 10.1093/eurheartj/ehw149] [Medline: 28158518]
5. Collste O, Sörensson P, Frick M, Agewall S, Daniel M, Henareh L, et al. Myocardial infarction with normal coronary arteries is common and associated with normal findings on cardiovascular magnetic resonance imaging: results from the Stockholm Myocardial Infarction with Normal Coronaries study. J Intern Med 2013 Feb;273(2):189-196 [FREE Full text] [doi: 10.1111/j.1365-2796.2012.02567.x] [Medline: 22742529]
6. Beltrame JF. Assessing patients with myocardial infarction and nonobstructed coronary arteries (MINOCA). J Intern Med 2013 Mar;273(2):182-185 [FREE Full text] [doi: 10.1111/j.1365-2796.2012.02591.x] [Medline: 22998397]
7. Safdar B, Spatz ES, Dreyer RP, Beltrame JF, Lichtman JH, Spertus JA, et al. Presentation, Clinical Profile, and Prognosis of Young Patients With Myocardial Infarction With Nonobstructive Coronary Arteries (MINOCA): Results From the VIRGO Study. J Am Heart Assoc 2018 Jun 28;7(13):e009174 [FREE Full text] [doi: 10.1161/JAHA.118.009174] [Medline: 29954744]
8. Agewall S, Beltrame JF, Reynolds HR, Niessner A, Rosano G, Caforio ALP, WG on Cardiovascular Pharmacotherapy. ESC working group position paper on myocardial infarction with non-obstructive coronary arteries. Eur Heart J 2017 Jan 14;38(3):143-153. [doi: 10.1093/eurheartj/ehw149] [Medline: 28158518]
9. Pais JL, Izquierdo Coronel B, Galán Gil D, Espinosa Pascual MJ, Martinez Peredo CG, Awamleh García P, et al. Psycho-emotional disorders as incoming risk factors for myocardial infarction with non-obstructive coronary arteries. Cardiol J 2018;25(1):24-31 [FREE Full text] [doi: 10.5603/CJ.a2017.0139] [Medline: 29240964]
10. Daniel M, Agewall S, Berglund F, Caidahl K, Collste O, Ekenbäck C, et al. Prevalence of Anxiety and Depression Symptoms in Patients with Myocardial Infarction with Non-Obstructive Coronary Arteries. The American Journal of Medicine 2018 Sep;131(9):1118-1124. [doi: 10.1016/j.amjmed.2018.04.040]
11. Feng H, Chien W, Cheng W, Chung C, Cheng S, Tzeng W. Risk of anxiety and depressive disorders in patients with myocardial infarction: A nationwide population-based cohort study. Medicine (Baltimore) 2016 Aug;95(34):e4464 [FREE Full text] [doi: 10.1097/MD.0000000000004464] [Medline: 27559951]
12. Daniel M, Agewall S, Caidahl K, Collste O, Ekenbäck C, Frick M, et al. Effect of Myocardial Infarction With Nonobstructive Coronary Arteries on Physical Capacity and Quality-of-Life. Am J Cardiol 2017 Aug 01;120(3):341-346. [doi: 10.1016/j.amjcard.2017.05.001] [Medline: 28610801]
13. Compare A, Grossi E, Bigi R, Proietti R, Shonin E, Orrego PS, et al. Stress-induced cardiomyopathy and psychological wellbeing 1 year after an acute event. J Clin Psychol Med Settings 2014 Mar;21(1):81-91. [doi: 10.1007/s10880-013-9380-1] [Medline: 24217954]
14. Wallström S, Ulin K, Määttä S, Omerovic E, Ekman I. Impact of long-term stress in Takotsubo syndrome: Experience of patients. Eur J Cardiovasc Nurs 2016 Dec;15(7):522-528 [FREE Full text] [doi: 10.1177/1474515115618568] [Medline: 26572162]

15. Dahlviken RM, Fridlund B, Mathisen L. Women's experiences of Takotsubo cardiomyopathy in a short-term perspective--a qualitative content analysis. Scand J Caring Sci 2015 Jul;29(2):258-267. [doi: 10.1111/scs.12158] [Medline: 24953349]
16. Sundelin R, Bergsten C, Tornvall P, Lyngå P. Self-rated stress and experience in patients with Takotsubo syndrome: a mixed methods study. Eur J Cardiovasc Nurs 2020 Jul 03:1474515120919387. [doi: 10.1177/1474515120919387] [Medline: 32491953]
17. Salmoirago-Blotcher E, Rosman L, Wittstein IS, Dunsiger S, Swales HH, Aurigemma GP, et al. Psychiatric history, post-discharge distress, and personality characteristics among incident female cases of takotsubo cardiomyopathy: A case-control study. Heart Lung 2016;45(6):503-509. [doi: 10.1016/j.hrtlng.2016.07.008] [Medline: 27553636]
18. National Collaborating Centre for Mental Health. Common mental health problems: identification and pathways to care. In: Common Mental Health Disorders: Identification and Pathways to Care. NICE Clinical Guidelines No. 123. Leicester, UK: British Psychological Society; 2011.
19. Gulliksson M, Burell G, Vessby B, Lundin L, Toss H, Svärdsudd K. Randomized controlled trial of cognitive behavioral therapy vs standard treatment to prevent recurrent cardiovascular events in patients with coronary heart disease: Secondary Prevention in Uppsala Primary Health Care project (SUPRIM). Arch Intern Med 2011 Jan 24;171(2):134-140. [doi: 10.1001/archinternmed.2010.510] [Medline: 21263103]
20. Cafarella PA, Effing TW, Usmani Z, Frith PA. Treatments for anxiety and depression in patients with chronic obstructive pulmonary disease: a literature review. Respirology 2012 May;17(4):627-638. [doi: 10.1111/j.1440-1843.2012.02148.x] [Medline: 22309179]
21. Linden W, Phillips MJ, Leclerc J. Psychological treatment of cardiac patients: a meta-analysis. Eur Heart J 2007 Dec;28(24):2972-2984. [doi: 10.1093/eurheartj/ehm504] [Medline: 17984133]
22. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
23. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
24. Zetterqvist K, Maanmies J, Ström L, Andersson G. Randomized controlled trial of internet-based stress management. Cogn Behav Ther 2003;32(3):151-160. [doi: 10.1080/16506070302316] [Medline: 16291546]
25. Lindsäter E, Axelsson E, Salomonsson S, Santoft F, Ejeby K, Ljótsson B, et al. Internet-Based Cognitive Behavioral Therapy for Chronic Stress: A Randomized Controlled Trial. Psychother Psychosom 2018;87(5):296-305. [doi: 10.1159/000490742] [Medline: 30041167]
26. Salomonsson S, Santoft F, Lindsäter E, Ejeby K, Ljótsson B, Öst LG, et al. Cognitive-behavioural therapy and return-to-work intervention for patients on sick leave due to common mental disorders: a randomised controlled trial. Occup Environ Med 2017 Dec;74(12):905-912. [doi: 10.1136/oemed-2017-104342] [Medline: 28756415]
27. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
28. Wallin EEK, Mattsson S, Olsson EMG. The Preference for Internet-Based Psychological Interventions by Individuals Without Past or Current Use of Mental Health Treatment Delivered Online: A Survey Study With Mixed-Methods Analysis. JMIR Ment Health 2016 Jul 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
29. Peynenburg VA, Mehta S, Hadjistavropoulos HD. Postsecondary student perceptions and preferences for the treatment of depression and anxiety: Comparison of internet-delivered cognitive behaviour therapy to face-to-face cognitive behaviour therapy and medication. Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement 2020 Jul 05;52(3):220-230. [doi: 10.1037/cbs0000165]
30. Higgins RO, Rogerson M, Murphy BM, Navaratnam H, Butler MV, Barker L, et al. Cardiac Rehabilitation Online Pilot: Extending Reach of Cardiac Rehabilitation. J Cardiovasc Nurs 2017;32(1):7-13. [doi: 10.1097/JCN.0000000000000297] [Medline: 26422638]
31. Webster R, Thompson AR, Norman P, Goodacre S. The acceptability and feasibility of an anxiety reduction intervention for emergency department patients with non-cardiac chest pain. Psychol Health Med 2017 Jan;22(1):1-11 [FREE Full text] [doi: 10.1080/13548506.2016.1144891] [Medline: 26924523]
32. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
33. Hewlett S, Wit MD, Richards P, Quest E, Hughes R, Heiberg T, et al. Patients and professionals as research partners: challenges, practicalities, and benefits. Arthritis Rheum 2006 Aug 15;55(4):676-680 [FREE Full text] [doi: 10.1002/art.22091] [Medline: 16874772]
34. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. A systematic review of the impact of patient and public involvement on service users, researchers and communities. Patient 2014;7(4):387-395. [doi: 10.1007/s40271-014-0065-0] [Medline: 25034612]

35. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015 Jan 23;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]
36. Popay J, Collins M. The Public Involvement Impact Assessment Framework Guidance. PiiAF. URL: http://piiaf.org.uk/documents/piiaf-guidance-jan14.pdf [accessed 2020-08-28]
37. Norlund F, Wallin E, Olsson EMG, Wallert J, Burell G, von Essen L, et al. Internet-Based Cognitive Behavioral Therapy for Symptoms of Depression and Anxiety Among Patients With a Recent Myocardial Infarction: The U-CARE Heart Randomized Controlled Trial. J Med Internet Res 2018 Mar 08;20(3):e88 [FREE Full text] [doi: 10.2196/jmir.9710] [Medline: 29519777]
38. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Fernández ME. Planning health promotion programs: An intervention mapping approach. Hoboken, NJ: John Wiley & Sons; 2011.
39. Lyon AR, Koerner K. User-Centered Design for Psychosocial Intervention Development and Implementation. Clin Psychol (New York) 2016 Jul;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
40. Tang W, Kreindler D. Supporting Homework Compliance in Cognitive Behavioural Therapy: Essential Features of Mobile Apps. JMIR Ment Health 2017 Jul 08;4(2):e20 [FREE Full text] [doi: 10.2196/mental.5283] [Medline: 28596145]
41. Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A. Optimizing inhibitory learning during exposure therapy. Behav Res Ther 2008 Jan;46(1):5-27. [doi: 10.1016/j.brat.2007.10.003] [Medline: 18005936]
42. Ali K, Roffe C, Crome P. What patients want: Consumer involvement in the design of a randomized controlled trial of routine oxygen supplementation after acute stroke. Stroke 2006;37(3):865-871. [doi: 10.1161/01.STR.0000204053.36966.80] [Medline: 16456122]
43. Birch R, Simons G, Wähämaa H, McGrath CM, Johansson EC, Skingle D, et al. Development and formative evaluation of patient research partner involvement in a multi-disciplinary European translational research project. Res Involv Engagem 2020;6:6 [FREE Full text] [doi: 10.1186/s40900-020-0178-7] [Medline: 32099665]
44. Warner G, Baghdasaryan Z, Osman F, Lampa E, Sarkadi A. 'I felt like a human being'-An exploratory, multi-method study of refugee involvement in the development of mental health intervention research. Health Expect 2019 Dec 09:10. [doi: 10.1111/hex.12990] [Medline: 31705620]
45. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650. [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
46. Blackburn S, McLachlan S, Jowett S, Kinghorn P, Gill P, Higginbottom A, et al. The extent, quality and impact of patient and public involvement in primary care research: a mixed methods study. Res Involv Engagem 2018;4:16 [FREE Full text] [doi: 10.1186/s40900-018-0100-8] [Medline: 29850029]
47. Efferth T, Banerjee M, Paul NW. Broken heart, tako-tsubo or stress cardiomyopathy? Metaphors, meanings and their medical impact. Int J Cardiol 2017 Mar 01;230:262-268. [doi: 10.1016/j.ijcard.2016.12.129] [Medline: 28041712]
48. Wallin E, Norlund F, Olsson EMG, Burell G, Held C, Carlsson T. Treatment Activity, User Satisfaction, and Experienced Usability of Internet-Based Cognitive Behavioral Therapy for Adults With Depression and Anxiety After a Myocardial Infarction: Mixed-Methods Study. J Med Internet Res 2018 Mar 16;20(3):e87 [FREE Full text] [doi: 10.2196/jmir.9690] [Medline: 29549067]

Abbreviations

CAD: coronary artery disease
CBT: cognitive behavioral therapy
FGD: focus group discussion
iCBT: internet-based cognitive behavioral therapy
MI: myocardial infarction
MINOCA: myocardial infarction with nonobstructive coronary arteries
PPI: patient and public involvement
PRP: patient research partner
TTC: takotsubo cardiomyopathy
U-CARE: Uppsala University Psychosocial Care Programme
UCD: user-centered design

Edited by G Eysenbach; submitted 02.04.20; peer-reviewed by K Schenck-Gustafsson, C Reidy; comments to author 20.06.20; revised version received 01.07.20; accepted 26.07.20; published 17.09.20.

Please cite as:
Humphries SM, Rondung E, Norlund F, Sundin Ö, Tornvall P, Held C, Spaak J, Lyngå P, Olsson EMG
Designing a Web-Based Psychological Intervention for Patients With Myocardial Infarction With Nonobstructive Coronary Arteries: User-Centered Design Approach
J Med Internet Res 2020;22(9):e19066
URL: http://www.jmir.org/2020/9/e19066/
doi:10.2196/19066
PMID:32940615

Original Paper

Factors Affecting the Implementation of Electronic Antiretroviral Therapy Adherence Monitoring and Associated Interventions for Routine HIV Care in Uganda: Qualitative Study

Jessica E Haberer[1,2], MS, MD; Lindsey Garrison[1], MPH; John Bosco Tumuhairwe[3], MPH; Robert Baijuka[3], MA; Edna Tindimwebwa[3], BS; James Tinkamanyire[3], BS; Bridget F Burns[4], MCP; Stephen Asiimwe[3], MBChB, PhD

[1]Center for Global Health, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Medicine, Harvard Medical School, Boston, MA, United States

[3]Kabwohe Clinical Research Centre, Kabwohe, Uganda

[4]Department of Urban Studies and Planning, Massachusetts Institute of Technology, Boston, MA, United States

Corresponding Author:
Jessica E Haberer, MS, MD
Center for Global Health
Massachusetts General Hospital
125 Nashua St, Suite 722
Boston, MA
United States
Phone: 1 617 724 0351
Email: jhaberer@partners.org

Abstract

Background: High, sustained adherence to HIV antiretroviral therapy (ART) is critical for achieving viral suppression, which in turn leads to important individual health benefits and reduced secondary viral transmission. Electronic adherence monitors record a date-and-time stamp with each opening as a proxy for pill-taking behavior. These monitors can be combined with interventions (eg, data-informed adherence counseling, SMS-based adherence support, and/or alarms) and have been shown to improve adherence in multiple settings. Their use, however, has largely been limited to the research context.

Objective: The goal of the research was to use the Consolidated Framework for Implementation Research (CFIR) to understand factors relevant for implementing a low-cost electronic adherence monitor and associated interventions for routine HIV clinical care in Uganda.

Methods: We conducted in-depth qualitative interviews with health care administrators, clinicians, and ART clients about likes and dislikes of the features and functions of electronic adherence monitors and associated interventions, their potential to influence HIV care, suggestions on how to measure their value, and recommendations for their use in routine care. We used an inductive, content analysis approach to understand participant perspectives, identifying aspects of CFIR most relevant to technology implementation in this setting.

Results: We interviewed 34 health care administrators/clinicians and 15 ART clients. Participants largely saw the monitors and associated interventions as favorable and beneficial for supporting adherence and improving clinical outcomes through efficient, differentiated care. Relevant outside factors included structural determinants of health, international norms around supporting adherence, and limited funding that necessitates careful assessment of costs and benefits. Within the clinic, the adherence data were felt likely to improve the quality of counseling and thereby morale, as well as increase the efficiency of care delivery. Existing infrastructure and care expenditures and the need for proper training were other noted considerations. At the individual level, the desire for good health and a welcomed pressure to adhere favored uptake of the monitors, although some participants were concerned with clients not using the monitors as planned and the influence of poverty, stigma, and need for privacy. Finally, participants felt that decisions around the implementation process would have to come from the Ministry of Health and other funders and would be influenced by sustainability of the technology and the target population for its use. Coordination across the health care system would be important for implementation.

Conclusions: Low-cost electronic adherence monitoring combined with data-informed counseling, SMS-based support, and/or alarms have potential for use in routine HIV care in Uganda. Key metrics of successful implementation will include their impact

on efficiency of care delivery and clinical outcomes with careful attention paid to factors such as stigma and cost. Further theory-driven implementation science efforts will be needed to move promising technology from research into clinical care.

Trial Registration: ClinicalTrials.gov NCT03825952; https://clinicaltrials.gov/ct2/show/NCT03825952

(J Med Internet Res 2020;22(9):e18038) doi:10.2196/18038

KEYWORDS

adherence; antiretroviral therapy; electronic monitoring; implementation science

Introduction

High and sustained adherence to antiretroviral therapy (ART) is critical for achieving HIV viral suppression, which in turn leads to important individual health benefits and reduced secondary viral transmission [1,2]. However, the attention paid to adherence in clinical settings varies. Health care providers may ask about missed doses or perform pill counts, but these measures tend to overestimate adherence [3-5]. Pharmacy refill data is less biased and can improve detection of nonadherence [6]; yet all of these adherence measures are obtained with clinic visits and potentially after viremia has led to drug resistance [7]. Electronic adherence monitoring involves smart pill containers that record a date-and-time stamp with each opening as a proxy for pill-taking behavior. Real-time versions of electronic adherence monitors contain modems that transmit these data via cellular networks for internet-based review. Although limited by the need to use the monitor for each dose, these monitors provide daily adherence records that can trigger timely adherence interventions, potentially before the loss of viral suppression.

Several studies have suggested the effectiveness of electronic monitoring for adherence support. One study conducted in China found an increase in ART adherence when short message service (SMS) reminders triggered by real-time detection of missed doses were combined with data-informed counseling (ie, adherence records were used at clinic visits to develop solutions to adherence challenges) [8]. A similar study of triggered SMS reminders in South Africa (that did not include supported counseling) observed a decrease in sustained treatment interruptions [9]. Further, a randomized controlled trial in Uganda tested real-time adherence monitoring plus SMS reminders to patients and SMS notifications to social supporters (ie, friends or family who could help support adherence) and found improved average adherence and a reduction in sustained interruptions [10]. Other studies have used non–real-time electronic adherence monitoring data to inform counseling and also found increased ART adherence [11,12]. Importantly, these devices have been shown to be feasible and acceptable in these and other settings, although sometimes with technical challenges [13].

Despite this evidence of improved adherence and promise for HIV outcomes, electronic adherence monitors and associated interventions have largely remained in the research context—a fate common among mobile health (mHealth) interventions [14]. One clear initial barrier to implementation has been cost. Electronic monitors have traditionally cost more than US $100 each and require data transmission and hosting fees; SMS may necessitate additional development and other fees. Recently, a low-cost electronic monitor with integrated SMS messaging was developed with total costs of less than US $30 per patient per year. To contextualize these costs, a modeling analysis found that adherence monitoring–based interventions could be considered cost-effective in sub-Saharan Africa at up to $50 per person-year [15]. Intervention adoption, however, is influenced by many factors other than cost and can be holistically considered through the Consolidated Framework for Implementation Research (CFIR) [16]. The CFIR includes 5 domains: (1) intervention characteristics (eg, design, cost), (2) outer setting (eg, organizational knowledge of patient needs, external policies), (3) inner setting (eg, culture, relative prioritization), (4) individual characteristics (eg, beliefs about the intervention, self-efficacy), and (5) process (eg, planning for implementation, engaging leaders).

Here we present an exploratory analysis guided by CFIR and involving ART clients, clinicians, and health care administrators in which we sought to understand factors relevant for implementing electronic adherence monitoring and associated interventions for routine HIV clinical care in Uganda.

Methods

Study Setting

This study was based at the Kabwohe Clinical Research Centre (KCRC) in rural southwestern Uganda. The KCRC ART Clinic provides PEPFAR-subsidized care for more than 6000 individuals living with HIV. It is a health center level IV facility, which also provides other comprehensive primary health care services and is governed by the Ministry of Health (MoH) and a district health officer. Specialized care is available through regional and national referral hospitals, and community-level care is provided through lower-level health facilities and community health workers. The research team met with a local community advisory board and KCRC leadership prior to initiating the study and incorporated their feedback in the study design.

Study Participants

We stratified ART clients by duration of ART use (less than vs more than 6 months) and residence type (rural vs periurban); within these categories, we identified clients randomly (ie, every 10th patient attending clinic) to understand the average experience in the clinic. Given the hierarchical nature of the Ugandan health care system, we identified up to 5 health care administrators/clinicians from each of the following cadres: MoH officials; regional referral hospital administrators; district health officers; and health center III/IV clinic administrators, physicians, nurses, and ART adherence counselors. Inclusion

criteria for all participants were aged over 18 years and engagement in HIV care through one of the above-defined roles. Additionally, ART clients had to have HIV infection per clinic records and own a cellular phone (familiarity with cellular technology was felt important to inform their input on the intervention). Exclusion criteria for all participants were unwillingness or inability to provide informed consent.

Electronic Adherence Monitor and Associated Interventions

We studied the evriMED electronic adherence monitor (Wisepill Technologies, Figure 1), which can function with or without real-time data transmission. It can be paired with any combination of the interventions presented in Table 1.

Figure 1. The evriMED real-time adherence monitor.

Table 1. Interventions to be combined with electronic adherence monitoring.

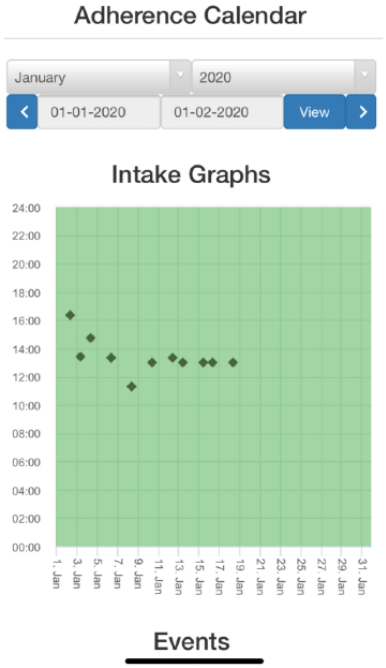
Intervention	Description
Data-informed counseling	Records of monitor openings are reviewed on a smartphone, tablet, or computer and discussed at each clinic visit to identify specific challenges and develop effective solutions to overcome future adherence barriers.
One-way scheduled SMS[a] to patients	SMS messages are sent to patients daily to encourage adherence (eg, through establishing the habit of daily pill taking and/or reminding patients that the clinic supports them). SMS are sent regardless of the recorded adherence.
One-way triggered SMS	When real-time monitors are used, SMS are sent to patients when one or more doses are taken late or missed. The SMS are sent to the patient and/or a social supporter (ie, a person who knows the patient's HIV status and is willing to provide support).
Two-way SMS	Both scheduled and triggered SMS allow for a callback from study staff to provide support directly at that time.
Alarms	Monitors are programmed to make audio-visual alerts when it is time to take medication.

[a]SMS: short message service.

Qualitative Interviews

Interviews were conducted by authors JBT and RB, who are both bilingual in the local language (Runyankole) and English and highly experienced, well-trained male qualitative research assistants. One-time interviews were digitally recorded for later transcription and took place in private settings; most occurred in the study office or near the clinic, although all MoH interviews were conducted in the participants' offices in Kampala, and some participants were interviewed at home, per their preference. Interviews with health care administrators/clinicians were conducted in English, which is commonly used in professional settings; interviews with ART clients were conducted in Runyankole or English per participant preference. Interviews began with an introduction to the research assistants, followed by statements of no conflicts of interest, a desire for honest perceptions (favorable or unfavorable), and the overall goals of the study. Participants were then asked for basic demographic data. A description of the electronic adherence monitors, associated interventions, evidence for their use, logistical requirements, and costs was subsequently read to participants (Multimedia Appendix 1). Participants were also shown an electronic adherence monitor and the software interface for displaying adherence data (Figure 2). Interview guides (Multimedia Appendix 2 and 3) were designed to obtain unbiased impressions of the technology and its potential for supporting ART in routine care, while also assessing each of the 5 domains in the CFIR. The guides were tailored for anticipated perspectives of health care administrators/clinicians versus ART clients. Initial questions in both guides asked about likes and dislikes of the features and functions of monitor and associated interventions and were followed by questions about their potential to influence HIV care, suggestions on how to measure their value, and recommendations for their use in routine care. Health care administrators/clinicians were also asked about the technology in relation to other health care priorities (ie, the outer setting). Questions were informally pretested with KCRC staff and clients and revised to ensure clarity and utility.

Figure 2. Adherence data display: each dot indicates the date and time of a monitor opening as a proxy of medication ingestion.



Research assistants wrote debriefs after each interview to capture body language, participant mood, and any other nonverbal aspects of the interviews. Transcripts were reviewed for quality among authors LG, BFB, JBT, and RB and corrected as needed. Participants were interviewed until thematic saturation was achieved.

Analysis

We used an inductive, content analysis approach [17] to explore factors that could influence the implementation of electronic adherence monitors plus associated interventions. We identified the aspects of CFIR [16] that participants indicated were most relevant to their context and potential implementation of the technology. In an iterative process, authors LG and JEH read the first 20% of transcripts, formulated codes, and assembled and pilot-tested a codebook. LG subsequently used the codebook to code the qualitative data, which was entered into qualitative analysis software (Dedoose, SocioCultural Research Consultants LLC). JEH and LG then developed categories by characterizing core concepts, developing labels, writing operational definitions, and selecting illustrative quotes from the interviews. Themes were reviewed with the qualitative research assistants but not participants to ensure accurate reflection of the participants' stated perspectives.

Ethics

All participants provided written informed consent. This study was reviewed and approved by the institutional review boards at the Mbarara University of Science and Technology, Ugandan National Council for Science and Technology, and Partners Healthcare. The study was registered with ClinicalTrials.gov [NCT03825952].

Results

Participant Characteristics

We interviewed 34 health care administrators/clinicians with a mean age of 37 (SD 10) years; 56% (19/34) were female. Four administrators worked in the MoH, while 5 worked in regional referral hospitals and 5 at the district level; 5 clinicians each were doctors, nurses, and adherence counselors. We also interviewed 15 ART clients with a mean age of 40 (SD 13) years; 60% (9/15) were female, and 67% (10/15) had taken ART for less than 6 months. A total of 60% (9/15) lived in periurban settings, and 40% (6/15) lived in rural settings. None of the individuals approached for participation declined. Interviews lasted an average of 51 minutes.

Overview

Table 2 presents the main factors participants felt would influence implementation of the adherence monitors and associated interventions. The factors are organized within the 5 domains of CFIR. Thematic saturation was achieved with the 49 study participants. Note that all perspectives reflect hypothetical use of the monitors and associated interventions.

Table 2. Factors influencing implementation of adherence monitors and associated interventions according to the Consolidated Framework for Implementation Research domains. The category of participant endorsing each factor is indicated per the footnote.

Intervention	Outer	Inner	Individual	Process

• Metrics of value
  • Improved monitoring and support for adherence[a]
  • Tool for management of viremia[b]
  • Improved clinical outcomes[a]
  • Facilitation of differentiated care[b]
• Device and SMS[c] features[a]
• Real-time monitoring versus stored records[a]
	• Structural determinants of health[b]
	• International norms[b]
	• Funding in the context of costs and benefits[b]
		• Data-informed counseling[a]
		• Quality of counseling[a]
		• Clinic morale[a]
		• Existing infrastructure and care expenditures[b]
		• Facilitation of efficient care delivery[b]
		• Need for in-service training, staffing, and support[b]
			• Desire for health[a]
			• Desire to please clinicians[d]
			• Not using monitors as planned[a]
			• Poverty, stigma, and privacy[a]
				• Decision making by the Ministry of Health and other funders[a]
				• Sustainability[b]
				• Target population[a]
				• Coordination across the health care system[b]

[a]Both health care administrators/clinicians and antiretroviral therapy client.

[b]Health care administrators/clinicians only.

[c]SMS: short message service.

[d]Antiretroviral therapy clients only.

Intervention

Both health care administrators/clinicians and ART clients uniformly stated that the value of the adherence monitor and associated interventions would be seen in their ability to improve knowledge of and support for adherence. They were seen as better than the current, limited approach to adherence monitoring.

> *The evriMED device is a good monitor in this era of technology and it will improve on monitoring of patient's adherence because we have been relying on backward methods like pill counting, which have their own disadvantages.* [District health officer, male, 54 years (health care administrator; HCA-1-023)]

> *If I used to miss taking my drugs for days or take beyond the time I am supposed to be swallowing the pills, and after using the monitor, I find that I no longer miss or take my drugs late, I would know that the monitor has really helped me to improve my adherence.* [ART client, female, 35 years (individual taking ART; ITA-1-007)]

Participants stated they would support using the monitor and associated interventions if they improved clinical outcomes (eg, viral suppression). Health care administrators/clinicians felt the data could also be used to differentiate care by targeting adherence counseling only to those in need, as well as helping to determine if viremia is due to nonadherence or drug resistance. This information would enable rapid and appropriately tailored care.

> *I think this would make them [ART clients] to remember taking their drugs the next day before it becomes a habit and worsens the situation.* [ART client, female, 45 years (ITA-1-002)]

> *I know if people are monitored well on adherence and if people at all do adhere because of this monitor, then you are improving their livelihood. Their chances of developing resistance will be very low because they are taking their ARVs or in case they are failing, then it is very easy for you to make a decision [on how to provide care].* [Physician, male, 30 years (HCA-1-030)]

> *If people are adhering to their treatment with support of such monitor and they are taking their medicine well, we expect they should suppress...and they can be moved to the differentiated delivery model which doesn't require them to come a lot to the clinic because now they move to the state of being stable and they decongest the clinic and they give us much time to look at the people who are very sick.* [Physician, male, 42 years (HCA-1-033)]

Both categories of participants identified several features of the monitors and associated interventions as advantageous for uptake, including size and secure storage capacity.

> *It is portable, it is clean, and it keeps all the ART client's drugs there.* [Physician, female, 39 years (HCA-1-018)]

> *The size should remain as it is so that I put in drugs for one month... The size should not be bigger than it is because it will become bulky and difficult to carry.* [ART client, female, 48 years (ITA-1-012)]

> *What I have liked about it is that it's like as if you go to the shop and buy a suitcase for keeping your clothes...so that they don't get damaged, it's the same thing I have liked about the evriMED because it will keep my drugs safely rather than keeping them in a piece of paper.* [ART client, female, 45 years (ITA-1-009)]

On the contrary, other participants felt the device was too big or too small, and several logistical requirements were seen as potentially limiting uptake, including the need to charge

batteries. Other features could lead to unintended and undesirable HIV status disclosure (eg, flashing lights or alarms).

> *My concerns are about electricity. For example, power may be off and the battery of the evriMED gets low. So, what will happen? Because it will not monitor when the battery is off.* [ART client, female, 55 years (HCA-1-014)]

> *If this device makes an alarm and you tell the patient to go with drugs to a place which is crowded, say it is a party, and it is going to make an alarm, I think that will be a challenge.* [ART client, male, 42 years (HCA-1-033)]

Views differed on the importance of real-time data transmission versus stored records that could be reviewed during clinic visits. Opinions were guided by prioritization of optimal adherence compared with cost and staffing resources. SMS reminders were seen as a way to extend the reach and support from the clinic. Both categories of participants felt that real-time intervention would reduce the necessary focus on adherence in clinic.

> *Real-time is much better as far as responses are concerned because there is a patina of poor adherence which people adopt, and it keeps on recurring and when we don't do real-time, we might lose time as well. Real-time is better, and I know the cost will come down.* [Physician, male, 42 years (HCA-1-033)]

> *We don't have the money and if this device is coming in to increase the cost of care per patient, I don't think it will be scalable but if it is not increasing cost per client then it will be good. Therefore, storing information and discussing it after is better.* [MoH administrator, male, 42 years (HCA-1-025)]

> *If a client receives a message informing him or her that he/she swallowed medicine very well last week; therefore, he or she should continue swallowing medicine very well. The client will feel happy that health workers are caring about him or her which will encourage client to continue swallowing his or her medicine very well.* [Adherence counselor, female, 31 years (HCA-1-007)]

> *It will influence my care in a great way because I may have forgotten to take my drugs or I am about to forget, but the SMS could remind me to take me drugs when it comes.* [ART client, female, 35 years (ITA-1-012)]

While participant views varied on specific features of the monitors and associated interventions, nearly all saw considerable potential for positive impact on HIV care.

Outer Setting

Beyond the intervention, health care administrators/clinicians noted that several structural factors would influence the success or failure of the interventions. For example, unstable personal life circumstances and community-level influences, like stigma and discrimination, may not be addressed by the proposed interventions.

> *It is very good for settings in the village whereby routine for day work is regular, but in town where people keep on moving from one place to another it might be challenging... The assumption may not hold that people are forgetting to take their drugs. There might be other reasons that are causing them not to take their drugs including the peer influence, pill burden fatigue, denial, stigma and discrimination.* [MoH administrator, male, 43 years (HCA-1-022)]

That said, this category of participants recognized the developing international norms for supporting adherence. They cited the Joint United Nations Programme on HIV and AIDS (UNAIDS) 90-90-90 policy that calls for 90% viral suppression; high adherence is critical for achieving this goal and would thus favor uptake of the technology.

> *Now adherence is one of the issues that we are looking up in the next strategic plan to increase efficiencies... We are almost enrolling 90% of our targets, but that is not good enough to end HIV. We need to increase efficiencies in the programs, so devices like the evriMED that increase efficiencies are very necessary to achieve the end to HIV.* [MoH administrator, male, 43 years (HCA-1-022)]

Another major factor was funding. Participants weighed benefits versus the costs in the context of other HIV expenditures.

> *I think we have to discuss it and look at the budget and we have a way out of it. But when you look at the benefits and you weigh against the cost, I think it is beneficial.* [Regional referral hospital administrator, female, 32 years (HCA-1-034)]

> *If, for example, our clinic was performing at 80% in good care, it can move our performance from 80% to 90%... But still we are performing without it. I hope you understand me. I am not saying that it doesn't have value and it can improve us... It will reduce the expenses on patients we move from first line to second line, so it has a lot of value. But I don't know the value when you mixed it with the cost.* [Physician, female, 45 years (HCA-1-029)]

Overall, the health care administrators/clinicians indicated that competing demands for limited funding would necessitate documented benefit from the intervention at a low total cost.

Inner Setting

Within the clinic, both categories of participants indicated that adherence monitor data would likely have considerable effects on counseling that would influence implementation. Most felt that the above-noted objectivity would change the dynamic between counselors and clients, allowing for more honesty and impact. The anticipated higher quality adherence counseling and more open relationship were seen as valuable for both clients and clinicians. This dynamic could improve job satisfaction for clinicians, thus motivating continued employment and continuity of care. Some participants, however, cautioned about the potentially negative impact of documented nonadherence.

> *When [the counselor] knows that I am adhering to my drugs well, it will also give him/her morale and encouragement to continue doing his/her job. We disturb them a lot... We don't tell them that we got some challenges that made us adhere poorly. You remain in the village and they bring you back to the clinic when your health has already deteriorated. You start blaming the clinic, saying they have not done enough for your health or cared about you, yet it's because of your ignorance.* [ART client, female, 35 years (ITA-1-007)]

> *Interviewer: What makes you think that it is not good for counselor seeing this information [adherence graph] and presenting it to you during your clinic visit? Participant: Because if you don't swallow your medicine very well, then counselor will think that he/she wasted time teaching you how to swallow your medicine.* [ART client, female, 35 years (ITA-1-006)]

> *When our clients are well or when they are healthy, I also feel okay because I know that I have done something great.* [Clinician, male, 58 years (HCA-1-008)]

Health care administrators/clinicians stated that the effects on counseling would have to be assessed within the context of clinic infrastructure and expenditures. Because the clinic already uses SMS with clients, SMS-based interventions would be affordable and easy to implement. The price of the monitor was seen as low when considered against other routine costs (eg, transportation). However, the overall cost of the monitors plus supporting infrastructure was a concern.

> *The cost of messages is not on the very high side. Currently as a clinic, we already have a platform that sends messages to patients who have raised viral load, and this doesn't cost us a lot.* [MoH administrator, male, 42 years (HCA-1-028)]

> *It is a very useful device... So, what can make me to invest is the cost vis-à-vis the number of monitors I am going to buy. And by the way the cost of the monitor together with the cost of the running infrastructure and the internet and all other things to ensure that everything is running well.* [Physician, female, 45 years (HCA-1-029)]

Health care administrators/clinicians emphasized that the above-noted ability to differentiate care would facilitate efficient care delivery in the clinic. For example, ART counselors would be better able to tailor their adherence counseling and allow the clients who are doing well to minimize their time in clinic. These efficiencies would help justify the cost of the monitors and SMS.

> *Those clients who will be found to be adhering well based on the monitor data and have no adherence issues may need no counseling while those that are doing badly in respect to adherence will be given adequate counseling or even switched to more experienced and skilled counselors for proper counseling or management.* [District health office, male, 54 years (HCA-1-023)]

> *The research will find out whether it [monitor] increases on the time each client spends with the doctor on clinic visit and it is good for the patient to spend more time with the doctor instead of rushing through.* [MoH administrator, male, 42 years (HCA-1-025)]

To enable uptake of the monitoring and associated interventions, participants also highlighted the need to provide in-service training. Opinions differed as to whether additional staff and/or infrastructure would be needed.

> *There is need to teach the clinic staff about the technology, how to use it, and how to do counseling based on that data from the monitor. If they teach staff on how to use it, there will be no challenges with clinical implementation.* [Adherence counselor, female, 49 years (HCA-1-011)]

> *The additional support the clinic might need is training of the personnel that will be involved in the use of this device/technology and providing logistics to the clinics necessary to use this technology which could be either from government or development partners. Things like tablets, computers if necessary and other accompanying equipment.* [District health officer, male, 42 years (HCA-1-028)]

Within the clinic setting, participants stated that the adherence monitors and associated interventions could have a significant impact on clinical care delivery, which they felt should be considered in any implementation plans.

Individual

Both categories of participants also identified several individual-level characteristics that could impact implementation. For instance, ART clients who prioritized improved health expressed enthusiasm to use the intervention. They described a welcomed pressure to adhere.

> *What I have liked is that it may report you to the clinic that you are not taking your drugs well, and as a patient, this will force you to take your drugs well and live longer.* [ART client, male, 37 years (ITA-1-005)]

Additionally, clients indicated that the monitors would enable them to demonstrate good adherence to the health care workers. This motivation reflected their appreciation of the care they receive in clinic.

> *I will know that the counselor is trying to help me on the basis of that data, to make sure that I live a healthy life and I cannot feel bad about it because I will know that she/he cares about my life... This counseling will help me to change my behavior because I will know that they will keep posted with my adherence.* [ART client, female, 48 years (ITA-1-012)]

That said, both categories of participants felt inaccuracies may arise in the data if clients do not use the monitors as planned, which would limit the value of the adherence data.

> *He/she might remove the pills, take them somewhere, and forgets or some time passes before he/she swallows the medicine. I would feel there would be a camera or a way of being sure that the client has taken the drugs after opening and removing them.* [Clinic administrator, female, 34 years (HCA-1-020)]

Poverty was seen as a principal influencing factor. Many participants felt clients could not pay for the monitors or associated SMS or pay for electricity to charge the monitors or their cell phones to receive or send SMS. Participants were also concerned that both traditional illiteracy and technical illiteracy (ie, ability to use technology)—two indirect effects of poverty—would limit some clients' ability to use the SMS.

> *[SMS] reminders in Uganda have not really yield the positive, because we have had reminders to mothers to attend antenatal... but the issue is that the phone must be charged, and the person must be with the phone and able to read the message.* [MoH administrator, male, 43 years (HCA-1-022)]

> *You know majority of patients are illiterate... so you have to continue emphasizing education on how to use the monitor. If you don't do that then they forget in a very short time.* [District health officer, male, 42 years (HCA-1-035)]

Nearly all participants commented on the potential for stigma in the event of lost privacy (eg, others seeing the monitors or SMS); however, views differed on the importance of these factors in using the intervention. Opinions seemed to stem from the extent to which clients accepted their HIV diagnosis and disclosed their status to others.

> *So you have to educate them on the advantages and functions of these monitors. If they feel out place carrying them, they may feel stigmatized. They may feel segregated and end up not using them.* [Regional referral hospital administrator, female, 32 years (HCA-1-034)]

> *I have no problem with the evriMED bringing stigma. I don't fear that because I am not the first person to have HIV or to be taking ART drugs. I have no problem with it.* [ART client, female, 23 years (ITA-1-002)]

> *Some people may not feel comfortable getting into their privacy in that someone knowing that there is a monitor that is recording him and information accessed by the third party, that alone may either increase stigma.* [Physician, female, 23 years (ITA-1-002)]

Participants felt that these individual-level factors were critical for uptake of the monitors and associated interventions. Challenges would be important to address through counseling and education.

Process

Consistent with the above-noted emphasis on cost-effectiveness, health care administrators/clinicians described funders as playing a key role in the implementation process. Support and resources were expected from the MoH and/or other organizations contracted for HIV services delivery in Uganda (eg, USAID-supported Elizabeth Glazer Pediatric AIDS Foundation). Most ART clients and clinicians reported having insufficient funding themselves to implement the monitors or interventions.

> *The clinic has no funding specifically to buy this [evriMED] because it receives medicines from the Ministry of Health through national medical stores and the staff belongs to the Ministry of Health and currently some of them are employed by RHITES Southwest, which is implemented by EGPAF... one needs to market it at the ministry level.* [MoH administrator, male, 42 years (HCA-1-028)]

> *I am not willing to pay for reminder messages or anything that evriMED does.* [ART client, male, 58 years (ITA-1-008)]

> *They will think that health workers have a hidden agenda in collecting that money which may spoil image of the health workers and the facility since our facility is a government facility and they think everything is for free.* [Adherence counselor, female, 31 years (HCA-1-010)]

In addition to upfront costs, health care administrators/clinicians highlighted the importance of program sustainability. They indicated the need to consider long-term costs, which depend partly on monitor durability.

> *That is one [question], affordability, sustainability and how do we service it supposing the funder left several of these and the funders pulled out, whoever the funder may be. Can the clinic program manage to service it, sustain it, so that it can be able to continue to work?* [Physician, female, 45 years (HCA-1-029)]

Health care administrators/clinicians also indicated that sustainability of the monitors and associated interventions would depend on the target population. Opinions on the target population, however, varied widely in both categories of participants. Some suggested including all ART clients, while others wanted to select those with high risk or documented adherence challenges. Participants felt that large numbers of clients would limit implementation, although excluding clients could also cause challenges.

> *For something to be routine, it should be able to be affordable for everybody and some of our clinics are very big like 12,000 [clients]...for example, is it something which is affordable for routine care when we are even failing to have Septrin to give to patients?* [Physician, female, 45 years (HCA-1-029)]

> *But this monitor can be a backup for a specific group of people, maybe the adolescents whose adherence is difficult, those patients who are nonsuppressed, maybe the mothers with high viral load...not everybody, because I think it maybe not sustainable, because it is very expensive.* [Physician, female, 45 years (HCA-1-029)]

> *All patients on ART should use the evriMED if possible.* [ART client, male, 37 years (ITA-1-005)]

> *It's a problem that you may not explain to every patient what the monitor is about. You may talk to those to whom you intend to give it, but patients talk to one another and they share a lot. So some may feel they are left out and feel that they are being segregated and denied a service.* [Regional referral hospital administrator, female, 32 years (HCA-1-034)]

If the decision to implement were made, health care administrators/clinicians advised coordination among multiple levels of health care organization through to the client to ensure successful care delivery.

> *There will be a need for like orientation for a capacity-building plan right from national level through the regions, districts, and the facilities. So, that has to be planned very well because people would want to know how to use the monitor and give the correct information to people.* [MoH administrator, female, 41 years (HCA-1-024)]

> *Attitude of the patients is a very important factor. You can do anything but if the patients' attitude towards the device is bad, you waste a lot of money.* [Physician, male, 30 years (HCA-1-030)]

Participants took a holistic view of the implementation process, including all of the above-noted factors related to the intervention, individual, inner setting, and outer setting.

Discussion

Principal Findings

This qualitative study of ART clients, health care administrators, and clinicians explored factors that may influence the implementation of low-cost electronic ART monitors and associated interventions for routine HIV care in Uganda. To our knowledge, this study is the first to use an implementation science framework to explore the means to move a technology-based adherence intervention from research into clinical care. The intervention was largely seen as favorable and beneficial for improved clinical outcomes with efficient, differentiated care. Concerns centered primarily around potential for stigma, device misuse, possible need for additional resources, and cost in the setting of competing demands for limited resources.

Participants identified improvements for the intervention that could address some of their concerns. For instance, stigma could be reduced by simple alterations, such as variable monitor sizes and making alarms optional. Counseling specifically around disclosure could also alleviate concerns for stigma, and education could support fidelity of use, even with low levels of literacy among clients. Within the clinic, leveraged use of existing infrastructure and staff would increase efficiency and reduce concerns about cost. And, perhaps most importantly, demonstrated value in clinical outcomes for the lowest possible cost could position implementation well against competing demands for limited resources in HIV care. These insights are particularly valuable, as most research on electronic adherence monitoring has focused only on measurement, even when considering the context of routine clinical care [18].

The potential value of an improved approach to adherence monitoring and support was endorsed by all participants. Steady progress is being made toward the UNAIDS 90-90-90 goals, and many are already achieving the high adherence necessary for viral suppression. Yet sustained adherence remains a challenge over time. Up to one-third of clients in sub-Saharan Africa are viremic at 2 years of therapy [19] and similar numbers have stopped ART and been lost from care at 5 years [20]. Increased adherence monitoring and support could play a critical role in reaching the 10-10-10 currently eluding the current care models.

Several tuberculosis treatment programs globally have begun to implement electronic adherence monitoring and support as part of routine care [21], although little evidence has been published on the implementation of these approaches. Studies driven by implementation science frameworks could facilitate uptake, assess for fidelity of implementation, and understand the impact on clinical outcomes. These data will be critical to determine how well potential benefits translate into real-world settings. Indeed, some preliminary reports with other digital adherence monitoring approaches suggest challenges with patient engagement and accuracy [22] that will need to be systematically addressed.

Limitations

This study has limitations. First, although we interviewed health care workers and administrators from all levels of the health care system, the ART clients came from a single site. That said, KCRC and the client population characteristics are largely reflective of ART delivery in rural East Africa. Second, study findings reflect perceived views that were not influenced by actual use of the monitors or associated interventions. Future work will involve deployment of the technology with subsequent reflections on the implementation process. Strengths of the paper include use of a comprehensive implementation science framework and in-depth exploration of factors relevant for the implementation process.

Conclusions

In conclusion, we found that low-cost electronic adherence monitoring combined with data-informed counseling, SMS-based support, and/or alarms has potential for use in routine HIV care in Uganda. Key metrics of successful implementation will include their impact on efficiency of care delivery and clinical outcomes with careful attention paid to factors such as stigma and cost. Given that most interventions fail to progress from research to practice, further theory-driven implementation science efforts will be needed to realize the benefits of this promising technology.

Acknowledgments

We would like to thank the study participants and clinical staff at KCRC. We also appreciate the support and input from Dr Josh Metlay. This study was funded by the US National Institute of Mental Health (K24MH114732). Neither the funder nor the manufacturer of the technology studied had any influence on the design, implementation, or data interpretation in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the electronic adherence monitor and associated interventions.
[DOCX File , 14 KB - jmir_v22i9e18038_app1.docx]

Multimedia Appendix 2
Qualitative interview guide for clients taking antiretroviral therapy.
[DOCX File , 15 KB - jmir_v22i9e18038_app2.docx]

Multimedia Appendix 3
Qualitative interview guide for healthcare administrators and clinicians.
[DOCX File , 15 KB - jmir_v22i9e18038_app3.docx]

References

1. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
2. Conway B. The role of adherence to antiretroviral therapy in the management of HIV infection. J Acquir Immune Defic Syndr 2007 Jun 01;45 Suppl 1:S14-S18. [doi: 10.1097/QAI.0b013e3180600766] [Medline: 17525686]
3. Haberer JE, Kiwanuka J, Nansera D, Ragland K, Mellins C, Bangsberg DR. Multiple measures reveal antiretroviral adherence successes and challenges in HIV-infected Ugandan children. PLoS One 2012;7(5):e36737 [FREE Full text] [doi: 10.1371/journal.pone.0036737] [Medline: 22590600]
4. Kagee A, Nel A. Assessing the association between self-report items for HIV pill adherence and biological measures. AIDS Care 2012;24(11):1448-1452. [doi: 10.1080/09540121.2012.687816] [Medline: 22670758]
5. Okatch H, Beiter K, Eby J, Chapman J, Marukutira T, Tshume O, et al. Brief report: apparent antiretroviral overadherence by pill count is associated with HIV treatment failure in adolescents. J Acquir Immune Defic Syndr 2016 Aug 15;72(5):542-545 [FREE Full text] [doi: 10.1097/QAI.0000000000000994] [Medline: 26990822]
6. Sangeda R, Mosha F, Prosperi M, Aboud S, Vercauteren J, Camacho R. Pharmacy refill adherence outperforms self-reported methods in predicting HIV therapy outcome in resource-limited settings. BMC Public Health 2014;14:1035. [doi: 10.1186/1471-2458-14-1035] [Medline: 25280535]
7. Haberer JE. Actionable adherence monitoring to optimise intervention. Lancet HIV 2017 Jan;4(1):e5-e6. [doi: 10.1016/S2352-3018(16)30191-6] [Medline: 27864001]
8. Sabin LL, Bachman DM, Gill CJ, Zhong L, Vian T, Xie W, et al. Improving adherence to antiretroviral therapy with triggered real-time text message reminders: the China adherence through technology study. J Acquir Immune Defic Syndr 2015 Aug 15;69(5):551-559 [FREE Full text] [doi: 10.1097/QAI.0000000000000651] [Medline: 25886927]
9. Orrell C, Cohen K, Mauff K, Bangsberg DR, Maartens G, Wood R. A randomized controlled trial of real-time electronic adherence monitoring with text message dosing reminders in people starting first-line antiretroviral therapy. J Acquir Immune Defic Syndr 2015 Dec 15;70(5):495-502. [doi: 10.1097/QAI.0000000000000770] [Medline: 26218411]
10. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
11. de Bruin M, Hospers HJ, van Breukelen GJP, Kok G, Koevoets WM, Prins JM. Electronic monitoring-based counseling to enhance adherence among HIV-infected patients: a randomized controlled trial. Health Psychol 2010 Jul;29(4):421-428. [doi: 10.1037/a0020335] [Medline: 20658830]
12. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 2012;52:275-301. [doi: 10.1146/annurev-pharmtox-011711-113247] [Medline: 21942628]
13. Musiimenta A, Atukunda EC, Tumuhimbise W, Pisarski EE, Tam M, Wyatt MA, et al. Acceptability and feasibility of real-time antiretroviral therapy adherence interventions in rural Uganda: mixed-method pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 May 17;6(5):e122 [FREE Full text] [doi: 10.2196/mhealth.9031] [Medline: 29773527]

14. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
15. Phillips AN, Cambiano V, Nakagawa F, Bansi-Matharu L, Sow PS, Ehrenkranz P, et al. Cost effectiveness of potential AST adherence monitoring interventions in sub-Saharan Africa. PLoS One 2016;11(12):e0167654 [FREE Full text] [doi: 10.1371/journal.pone.0167654] [Medline: 27977702]
16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
17. Patton M. Qualitative Research & Evaluation Methods. 3rd edition. Thousand Oaks: SAGE Publications; 2002.
18. Rotzinger A, Cavassini M, Bugnon O, Schneider MP. Development of an algorithm for analysing the electronic measurement of medication adherence in routine HIV care. Int J Clin Pharm 2016 Jul 29;38(5):1210-1218. [doi: 10.1007/s11096-016-0354-x]
19. Barth RE, van der Loeff MFS, Schuurman R, Hoepelman AIM, Wensing AMJ. Virological follow-up of adult patients in antiretroviral treatment programmes in sub-Saharan Africa: a systematic review. Lancet Infect Dis 2010 Mar;10(3):155-166. [doi: 10.1016/S1473-3099(09)70328-7] [Medline: 20185094]
20. Haas AD, Zaniewski E, Anderegg N, Ford N, Fox MP, Vinikoor M, et al. Retention and mortality on antiretroviral therapy in sub-Saharan Africa: collaborative analyses of HIV treatment programmes. J Int AIDS Soc 2018 Feb;21(2):1 [FREE Full text] [doi: 10.1002/jia2.25084] [Medline: 29479867]
21. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
22. Thomas B, Kumar VK, Chiranjeevi M, Khandelwale A, Shah D, Subbaraman R. Updated findings from an evaluation of the accuracy of 99DOTS, a digital technology for monitoring tuberculosis medication adherence in HIV co-infected and uninfected patients. 2019 Presented at: 50th Union World Conference on Lung Health; 2019 October 30; Hyderabad.

Abbreviations

ART: antiretroviral therapy
CFIR: Consolidated Framework for Implementation Research
KCRC: Kabwohe Clinical Research Centre
mHealth: mobile health
MoH: Ministry of Health
PEPFAR: President's Emergency Plan for AIDS Relief
SMS: short message service
UNAIDS: Joint United Nations Programme on HIV and AIDS

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by YH Kwan, M Paschen-Wolff; comments to author 01.04.20; revised version received 24.04.20; accepted 22.06.20; published 10.09.20.

Please cite as:
Haberer JE, Garrison L, Tumuhairwe JB, Baijuka R, Tindimwebwa E, Tinkamanyire J, Burns BF, Asiimwe S
Factors Affecting the Implementation of Electronic Antiretroviral Therapy Adherence Monitoring and Associated Interventions for Routine HIV Care in Uganda: Qualitative Study
J Med Internet Res 2020;22(9):e18038
URL: http://www.jmir.org/2020/9/e18038/
doi:10.2196/18038
PMID:32687473

Original Paper

Motivation Predicts Change in Nurses' Physical Activity Levels During a Web-Based Worksite Intervention: Results From a Randomized Trial

Jennifer Brunet[1,2,3], PhD; Heather E Tulloch[4], PhD; Emily Wolfe Phillips[1], BSc; Robert D Reid[4,5], MBA, PhD; Andrew L Pipe[4,5], CM, MD; Jennifer L Reed[4,5], RKin, PhD

[1]Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada

[2]Institut du savoir de l'Hôpital Montfort, Hôpital Montfort, Ottawa, ON, Canada

[3]Cancer Therapeutic Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

[4]Division of Prevention and Rehabilitation, University of Ottawa Heart Institute, Ottawa, ON, Canada

[5]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Jennifer Brunet, PhD
Faculty of Health Sciences
University of Ottawa
Montpetit Hall 339
125 University Private
Ottawa, ON, K1N 6N5
Canada
Phone: 1 613 562 5800 ext 3068
Fax: 1 613 562 5497
Email: jennifer.brunet@uottawa.ca

Abstract

Background: Low physical activity levels can negatively affect the health of nurses. Given the low physical activity levels reported by nurses, there is a clear need for brief and economical interventions designed to increase physical activity levels in this population. We developed a web-based intervention that used motivational strategies to increase nurses' physical activity levels. The intervention provided the nurses with feedback from an activity monitor coupled with a web-based individual, friend, or team physical activity challenge.

Objective: In this parallel-group randomized trial, we examine whether nurses' motivation at baseline predicted changes in objectively measured physical activity levels during the 6-week intervention.

Methods: The participants were 76 nurses (n=74, 97% female; mean age 46, SD 11 years) randomly assigned to 1 of 3 physical activity challenge conditions: (1) individual, (2) friend, or (3) team. The nurses completed a web-based questionnaire designed to assess motivational regulations for physical activity levels before the intervention and wore a Tractivity activity monitor before and during the 6-week intervention. We analyzed data using multilevel modeling for repeated measures.

Results: The nurses' physical activity levels increased (linear estimate=10.30, SE 3.15; *P*=.001), but the rate of change decreased over time (quadratic estimate=−2.06, SE 0.52; *P*<.001). External and identified regulations (ß=−2.08 to 11.55; *P*=.02 to .04), but not intrinsic and introjected regulations (ß=−.91 to 6.29; *P*=.06 to .36), predicted changes in the nurses' physical activity levels.

Conclusions: Our findings provide evidence that an intervention that incorporates self-monitoring and physical activity challenges can be generally effective in increasing nurses' physical activity levels in the short term. They also suggest that drawing solely on organismic integration theory to predict changes in physical activity levels among the nurses participating in web-based worksite interventions may have been insufficient. Future research should examine additional personal (eg, self-efficacy) and occupational factors (eg, shift length and shift type) that influence physical activity levels to identify potential targets for intervention among nurses.

Trial Registration: ClinicalTrials.gov NCT04524572; https://clinicaltrials.gov/ct2/show/NCT04524572

(J Med Internet Res 2020;22(9):e11543) doi:10.2196/11543

KEYWORDS

physical activity; motivation; wearable technology; nurses

Introduction

Background

Engaging in regular physical activity can improve cardiovascular function and musculoskeletal strength, reduce the risk of morbidity and mortality due to chronic disease, and decrease the risk of mental health problems such as anxiety and depression [1-3]. In addition, engaging in physical activity can reduce work-related stress and the incidence of burnout [4-7]—a major problem for health care workers [8]. Many investigators have shown that nurses, who represent 48% of the health care workforce [9], report high levels of work-related stress and burnout, low levels of job satisfaction, and poor health [10-14]. Despite the known benefits of physical activity, nurses' physical activity levels remain low [15-18]. Common barriers to physical activity reported by nurses include busy schedules, irregular shifts, long hours, and a lack of time, suggesting that the worksite may be an ideal place to intervene to increase nurses' physical activity levels [19]. Beyond personal health benefits, worksite interventions seeking to increase nurses' physical activity levels have the potential to improve employee performance, lower employee health care costs, and decrease absenteeism rates, which are higher in nurses than in other occupational groups [20].

The internet is a promising way to deliver worksite interventions, as it affords timely access and the ability to reach a larger population [21]. It may be particularly appropriate for nurses whose long working hours and irregular shifts preclude opportunities to participate in traditional face-to-face interventions that are often scheduled to accommodate those with relatively fixed schedules. There is mounting evidence that web-based interventions can help increase physical activity levels among working adults [22-25]. Nevertheless, some web-based worksite interventions have not led to significant increases in physical activity levels [26,27].

Although it has been recognized that changes in behaviors associated with a particular intervention may be influenced by the personal characteristics of the participants, few researchers evaluating the effects of web-based worksite interventions have sought to identify which characteristics, apart from sociodemographic factors, have an influence on behavior change [26,28-32]. Consequently, there is limited knowledge of other factors that may predict physical activity levels among web-based worksite intervention participants. An examination of additional factors that might predict physical activity levels in web-based worksite interventions is critical to acquire an enhanced understanding of the forces that impel change. There is robust evidence that motivation is a strong predictor of participation in physical activity [33]; researchers might, therefore, consider drawing on motivational theories such as the organismic integration theory [34,35]—1 of the 6 mini theories of self-determination theory—to ascertain whether motivation predicts physical activity levels among web-based worksite intervention participants.

Objectives

To address the aforementioned gaps in the literature, we developed a web-based worksite intervention for nurses working in a tertiary care cardiovascular institute. We created individual, friend, and team challenge groups in which the nurses would track their physical activity levels using a Tractivity activity monitor and upload their activity data at times and frequencies of their choosing because of the effectiveness of self-monitoring [36]. The nurses randomized to the friend and team challenge groups would also share their physical activity levels in deidentified format with one other nurse (friend challenge group) or a team of nurses (team challenge group) randomly chosen, which was presumed to motivate them to be more active to make a positive impression on members of their group according to self-presentation perspectives [37,38]. Furthermore, based on the principles of the social comparison theory [39], it was presumed that allowing the nurses randomized to the friend and team challenge groups to exchange physical activity level data would serve as a basis for social comparison, and such comparisons would impel further behavior change. For example, social comparisons could allow the nurses to develop an internal norm of what a *good* physical activity level is and encourage them to adjust their levels if there was a discrepancy. In this regard, observing better-performing nurses would prompt the nurses to increase their physical activity levels to reduce the discrepancy between themselves and others to make themselves feel good about their current levels.

Using data collected as part of a trial evaluating changes in physical activity levels and the impact on cardiovascular risk factors among nurses participating in a web-based worksite intervention [40], we examined whether the nurses' motivation predicted changes in their objectively measured daily physical activity levels. Using the organismic integration theory [34], we assessed 5 core motivational *regulations*: intrinsic motivation (ie, a person pursues an activity for the inherent pleasure and enjoyment of the activity), identified regulation (ie, a person pursues an activity that they deem personally valuable and important to attain a desired outcome), introjected regulation (ie, a person pursues an activity to avoid feelings of guilt and shame and/or protect feelings of worth and ego), external regulation (ie, a person pursues an activity because of external demands, eg, punishments, threats, and/or possible rewards), and amotivation (ie, a person has a relative absence of intrinsic or extrinsic motivation and lacks a reason to act). On the basis of the organismic integration theory [34] and past research [33,41], we hypothesized that self-determined motivational regulations (ie, intrinsic motivation and identified regulation) would positively predict initial levels of and changes in objectively measured daily physical activity levels among the study participants. As researchers have observed inconsistent associations between non–self-determined motivational regulations (ie, introjected regulation, external regulation, and amotivation) and physical activity levels [33,41], we further hypothesized that these regulations would be unrelated or

negatively associated with initial levels of and changes in objectively measured daily physical activity levels.

Methods

Setting and Procedures

Following ethics approval by the University of Ottawa Heart Institute Research Ethics Board (Protocol No. 20130429), nurses working at the University of Ottawa Heart Institute—a tertiary care cardiovascular institute—were recruited to participate in this parallel-group randomized trial. Further details about the study design and procedures have been reported previously [40]. Briefly, recruitment took place between September 2013 and November 2013 via posters distributed throughout the University of Ottawa Heart Institute, word of mouth, and announcements during nursing meetings and morning rounds. The nurses were eligible if they were (1) a registered nurse, (2) able to walk unassisted, (3) willing to wear a stretchable ankle band that contained a physical activity monitoring device (ie, accelerometer) and had access to the internet, and (4) able and willing to provide written informed consent. The nurses were not eligible if they were (1) pregnant or lactating, (2) unable to read and understand English, (3) having medical contraindications to exercise, and/or (4) already using an activity monitor to track their physical activity levels. Nurses who were interested and believed they were eligible were invited to contact the study staff who confirmed final eligibility.

Once eligibility was confirmed, the nurses attended a study enrollment session with study staff where they provided written informed consent and then received a Tractivity activity monitor along with instructions for using it and instructions for logging onto and uploading data to their Tractivity web account (Multimedia Appendix 1). The nurses were instructed to wear the activity monitor from waking to bedtime (except during water activities) throughout the baseline (1 week) and intervention (6 weeks) phases. In addition, they were asked to complete self-report measures (eg, sociodemographics, work-related characteristics, and motivational regulations for physical activity) at baseline and had their resting blood pressure, heart rate, and anthropometric measurements (ie, height, body mass, waist circumference, and body fat percentage) taken by research staff who were blinded to the assigned groups of the participants. Further details regarding these assessments can be found in a study by Reed et al [40].

Participants

In total, 76 nurses contacted the research staff, met eligibility criteria, and consented to participate in this study (Figure 1). Their mean age was 46.3 (SD 10.9) years, mean BMI was 27.5 (SD 5.6) kg/m^2, and their mean resting blood pressure was 115 (SD 12)/75 (SD 8) mm Hg. On the basis of these values, they were categorized as being mostly overweight and normotensive. Most were female (74/76, 97%), worked only day shifts (40/76, 53%), and performed clinical duties (53/76, 70%). Only 3 of the 76 (4%) participants met the current physical activity guidelines at baseline. Additional information describing the participants' demographics, anthropometrics, shift profiles, and nursing roles are presented by Reed et al [40] (Table 1).

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of nurses recruited and reasons for withdrawals.

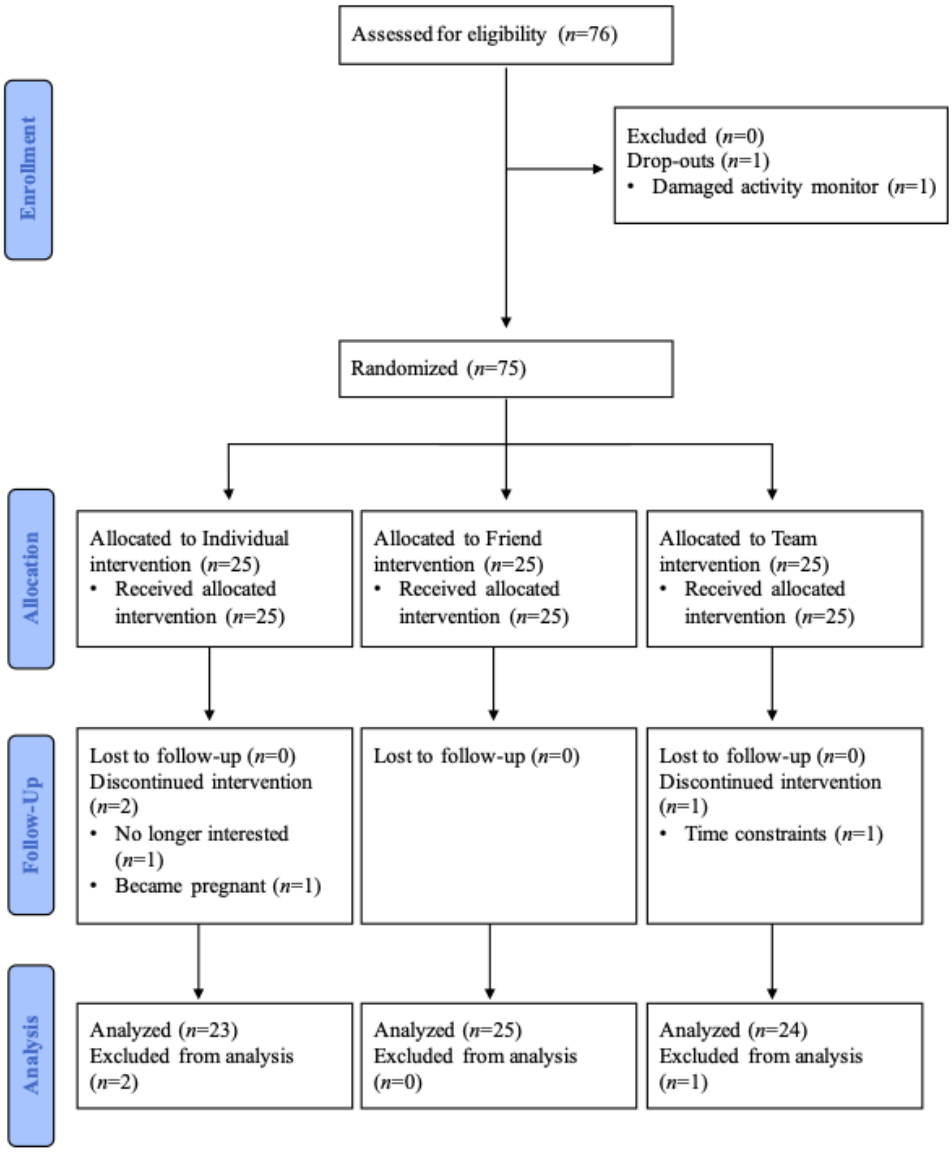


Table 1. Fixed effects and fit statistics for the multilevel growth models of moderate-to-vigorous physical activity.

Variables	Model 1		Model 2		Model 3		Model 4	
	Estimate (SE)	*P* value	Estimate (SE)	*P* value	Estimate (SE)	*P* value	Estimate (SE)	*P* value
Intercept	44.43 (5.13)	<.001	34.71 (4.63)	<.001	30.49 (12.24)	.02	34.85 (4.45)	<.001
Time	−1.96 (0.79)	.02	10.30 (3.15)	.002	15.04 (8.37)	.08	10.26 (3.03)	.001
Time squared	—[a]	—	−2.06 (0.52)	<.001	−2.83 (1.38)	.04	−2.05 (0.50)	<.001
Group	—	—	—	—	2.14 (5.74)	.71	—	—
Group×time	—	—	—	—	−2.40 (3.92)	.54	—	—
Group×time squared	—	—	—	—	0.39 (0.65)	.55	—	—
External	—	—	—	—	—	—	−1.60 (7.11)	.82
Introjected	—	—	—	—	—	—	5.65 (4.82)	.24
Identified	—	—	—	—	—	—	11.43 (7.75)	.14
Intrinsic	—	—	—	—	—	—	1.51 (6.33)	.81
External×time	—	—	—	—	—	—	11.55 (4.83)	.02
Introjected×time	—	—	—	—	—	—	−6.29 (3.27)	.06
Identified×time	—	—	—	—	—	—	11.38 (5.35)	.04
Intrinsic×time	—	—	—	—	—	—	−4.05 (4.37)	.36
External×time squared	—	—	—	—	—	—	−1.87 (0.80)	.02
Introjected×time squared	—	—	—	—	—	—	0.91 (0.54)	.10
Identified×time squared	—	—	—	—	—	—	−2.08 (0.89)	.02
Intrinsic×time squared	—	—	—	—	—	—	0.75 (0.73)	.31
2 restricted log likelihood	4803.96	N/A[b]	4747.55	N/A	4739.56	N/A	4696.33	N/A
Akaike information criterion	4811.96	N/A	4761.55	N/A	4753.56	N/A	4710.33	N/A
Schwarz Bayesian information criterion	4828.61	N/A	4790.68	N/A	4782.65	N/A	4739.28	N/A

[a]There are no results to report.

[b]N/A: not applicable.

Randomization and Intervention Groups

Of the 76 nurses who provided consent, 75 (99%) were randomized to the individual, friend, or team physical activity challenge groups, and 1 (1%) dropped out following the baseline assessment because of a damaged Tractivity activity monitor (Figure 1). Randomization to the 3 groups was conducted by research staff using the *RAND* function of a software spreadsheet program (Microsoft Excel) in a 1:1:1 ratio. The participants were notified of their assigned group via email. In the individual challenge group, the participants were able to log onto their Tractivity web account at any time during the intervention phase to track their physical activity levels (ie, distance [km], steps [counts], active time [min], and calories [kcal]) displayed in a graphical format in the web-based Tractivity program. In the friend and team challenge groups, the participants were also able to log onto their Tractivity web account at any time during the intervention phase to track their own physical activity levels, but they could also monitor the physical activity levels of either one other participant (friend challenge group; Multimedia Appendix 2) or 4 other participants (team challenge group; Multimedia Appendix 3). The participants in the friend and team challenge groups were blinded in keeping with ethical considerations; none knew the identity of the other participant or team members in their group.

Study Assessments

Motivational regulations for physical activity were assessed at baseline using the 19-item Behavioral Regulation in Exercise Questionnaire-2 (BREQ-2) [42]. The participants were presented with the stem, “Using the scale below, please indicate to what extent each of the following items is true for you,” followed by items representing amotivation (4 items; eg, “I can’t see why I should bother exercising”), external (4 items; eg, “I feel under pressure from my friends/family to exercise”), introjected (3 items; eg, “I feel ashamed when I miss an exercise session”), identified (4 items; eg, “I value the benefits of exercise”), and intrinsic (4 items; eg, “I exercise because it’s fun”) regulations. Items were rated on a 5-point Likert scale, ranging from 0 (*not true for me*) to 4 (*very true for me*). Integrated regulation is not assessed on this scale because it is difficult to differentiate between integrated and identified regulation [43]. We calculated subscale scores by averaging responses of items belonging to the same subscale; however, only the external, introjected, identified, and intrinsic regulation subscales were analyzed in this study because of the extremely low variance and the high

number of zeros for the amotivation subscale. The reliability and validity of BREQ-2 scores have been previously demonstrated [44,45].

Physical activity was measured regularly during the baseline (1 week) and intervention (6 weeks) phases using the Tractivity activity monitor), which is a lightweight, compact accelerometer that uses a proprietary signal processing algorithm to determine step counts in 1-min intervals. The activity monitor provides no visible feedback and stores up to 30 days of data (ie, distance, steps, active time, and calories). Research staff uploaded the participants' activity data into the web-based Tractivity program at the end of the baseline and intervention phases. The participants uploaded their activity data at times and frequencies of their choice throughout the intervention phase. The Tractivity activity monitor has been shown to be a valid measure of step counts in comparison with direct observation [46]. Activity monitors were calibrated for stride length before the baseline week by having the participants walk 10 steps (at their usual walking speed) in a straight line on a large indoor track. These measures were performed in triplicate, and the average was entered into the web-based Tractivity program to assist the proprietary signal processing algorithm in calculating step counts.

The monitors provided us with consecutively ordered min-by-min activity data (ie, steps [counts], distance [km], active time [min], and calories [kcal]) during each day of the baseline and intervention phases for all the participants. We used a Hypertext Preprocessor (PHP, version 7.0) script to process the data. All activity monitor data were screened to identify valid and nonvalid days. Data were considered valid and included in the analysis if the wear time was at least 10 hours [47]. Step counts were used to calculate min of moderate-to-vigorous physical activity (MVPA) levels in bouts of at least 10 min [48,49]. Previously established cut-points (ie, >100 steps per min) [50] were used to calculate daily min of MVPA.

Sample Size

A post hoc power analysis revealed that the sample size of 76 participants provided adequate power ($1-\beta=.92$) to detect significant differences in physical activity levels within and between groups of small magnitude (ie, eta-squared value of 0.022 with an α of .05).

Statistical Analysis

All data analyses were performed using SPSS (version 24; IBM Corp), and $P<.05$ was considered statistically significant. Descriptive characteristics of the study sample were summarized using mean (SD) or frequencies (%). Data were analyzed using multilevel growth modeling as repeated observations were nested within the participants who were nested within the groups [51]. When analyzing longitudinal data, multilevel growth modeling also offers the following advantages: (1) equally spaced periods are not required, (2) the number of time points may vary across the participants, allowing for the use of data from all the participants to provide unbiased estimates of the outcomes, assuming data are missing at random, and (3) missing data are not problematic as long as they are missing at random [52]. Before these analyses, a 2-step approach for transforming continuous, nonnormalized variables to normal variables was applied to the MVPA data [53], as preliminary analysis revealed that the MVPA data were not normally distributed.

We then estimated an unconditional multilevel linear growth model for MVPA (model 1) and compared it with an unconditional quadratic growth model (model 2) to formally test the optimal functional form of growth. In doing so, we created a new variable *time*, for which baseline was coded as *0* to serve as the reference point, and subsequent time points were assigned the following values: 1, 2, 3, 4, 5, and 6. This coding accounts for any differences in time intervals between points and allowed for the interpretation of the intercept as predicted MVPA levels at baseline. In addition, fixed and random effects for time were included because it was assumed that not all the participants had the same baseline MVPA levels or the same exact rate of change over time. The fixed effects provide estimates of the average levels at baseline and average rate of change for the sample, whereas the random effects serve to ascertain whether there is variability in baseline levels and in the rate of change. These competing models were compared using a likelihood ratio test and 2 commonly used information criteria, namely, Akaike information criterion and Schwarz Bayesian information criterion. The model that minimized Akaike information criterion and Bayesian information criterion values was retained.

Next, we expanded the retained unconditional growth model by adding group, group by time, and group by time squared as predictors of MVPA to test the effect of group (model 3). There was no statistically significant main effect for group ($P=.71$) or interaction between time (and time square) and group (Table 1, model 3; $P=.54$ to .55). This demonstrated that there were no group differences in MVPA levels at baseline or in change over time. We calculated the intraclass correlation coefficient to further assess dependence in the grouped data. As the coefficient was less 0.05 (indicating that dependence related to group membership could be ignored) [54], we proceeded to test subsequent models without group, group by time, and group by time squared.

Finally, we added the motivational regulations as predictors to the unconditional quadratic growth model to test the effect of each regulation on MVPA (model 4). To fit this conditional growth model, we added the main effects of each regulation along with their interaction with time (and time squared). Of note, each regulation was grand-mean centered by subtracting the sample mean from each observed value to make the interpretation of the model parameters easier.

Results

Participants

Of the 75 participants randomized, 72 (96%) completed all study assessments, including 92% (23/25) assigned to the individual challenge, 100% (25/25) assigned to the friend challenge, and 96% (24/25) assigned to the team challenge. A one-way analysis of variance was performed to compare baseline MVPA levels between the participants who dropped out and

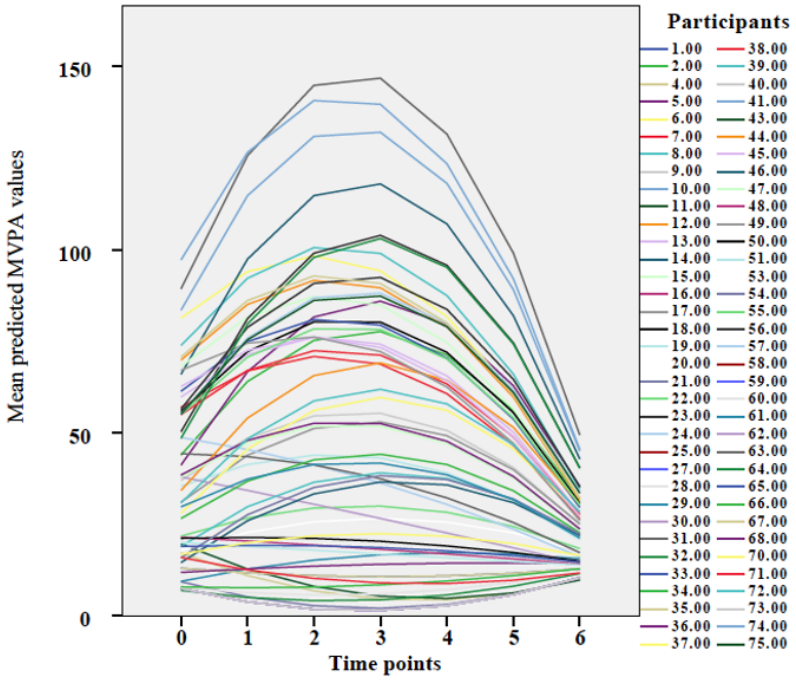
those who completed the 6-week intervention; results revealed no significant differences in baseline MVPA levels.

Main Results

Visual inspection of the plotted trajectories using predicted normalized MVPA values (Figure 2) suggested that it may not be adequate to summarize the pattern of change over time with a linear trajectory, but rather a quadratic trajectory over time. In addition, there were several indications that a quadratic growth model was the most appropriate for representing the individual growth trajectories of MVPA levels (Table 1, models 1 and 2). First, the Akaike information criterion and Bayesian information criterion values were smaller for the quadratic growth model. Second, the fixed quadratic effect and variance components of the quadratic model were significant and of nontrivial magnitude. Third, after refitting the 2 models with full information maximum likelihood, a likelihood ratio test comparing the linear model with the quadratic model indicated that the former should be rejected in favor of the latter. The fixed effects were significant in the quadratic unconditional growth model (Table 1, model 2), demonstrating that the mean MVPA baseline level was 34.71 (SE 4.63; P<.001) min per week and that levels changed significantly over time in a curvilinear (ie, inverted U shape) fashion (linear estimate=10.30, SE 3.15; P=.002; and quadratic estimate=−2.06, SE 0.52; P<.001). In addition, the random effects for (1) the intercept (estimate=815.85, SE 258.40; P=.002), (2) the slopes for MVPA (linear estimate=269.39, SE 122.85; P=.03; and quadratic estimate=8.07, SE 3.47; P=.02), and (3) the covariance between the intercepts and quadratic slopes (estimate=−50.86, SE 21.22; P=.02) were significant. These findings demonstrate that there was meaningful variability in (1) MVPA levels at baseline between the participants, and (2) changes in MVPA levels over time. Furthermore, the participants who engaged in more MVPA at baseline tended to have greater increases in MVPA levels initially, followed by steeper decreases.

The results of the conditional quadratic growth model, in which we added the main effects of each grand-mean centered motivational regulation along with their interaction with time (and time squared), are presented in Table 1 (model 4). No significant main or interaction effects were observed for introjected regulation or intrinsic motivation (P=.06 to .81). In contrast, there were significant effects for external and identified regulations. Specifically, there was a significant interaction between time (and time squared) and external regulation (P=.02) as well as between time (and time squared) and identified regulation (P=.02 to .04). These findings indicate that there are differences in the rate of change in MVPA levels as a function of the external and identified regulations levels of the participants at baseline. To better understand the nature of these relationships, we probed both interactions by a test of simple slopes at specific values of external and identified regulations, namely, at high (1 SD above the mean), medium (at the mean), and low (1 SD below the mean) levels of each regulation [55,56]. Probing showed that initial increases in MVPA levels were significant for the participants reporting medium (estimate=−3.93, SE 1.58; P=.02) and high (estimate=−5.70, SE 2.49; P=.03) levels of external regulation at baseline, but not for those with low levels of external regulation (estimate=−2.16, SE 1.12; P=.06). Similarly, probing showed that initial increases in MVPA levels were significant for the participants reporting medium (estimate=−3.96, SE 1.57; P=.01) and high (estimate=−6.76, SE 2.48; P=.009) levels of identified regulation at baseline, but not for those with low levels of identified regulation (estimate=−1.16, SE 1.10; P=.30).

Figure 2. A plot illustrating the individual trajectories for MVPA levels at baseline (week 0) and throughout the intervention phase (weeks 1-6). More negative slopes correspond to greater decreases in MVPA levels. MVPA: moderate-to-vigorous physical activity.



Discussion

Principal Findings

Given the importance of identifying factors that may help to promote changes in physical activity levels within the workplace setting, the aim of this study was to examine whether motivation predicted changes in objectively measured daily physical activity levels among participating nurses. Our principal finding is that some, namely, external and identified regulations, but not all types of motivation, predicted changes in the nurses' physical activity levels throughout the intervention. According to the World Health Organization, the workplace is an ideal setting to implement health promotion initiatives to reduce noncommunicable disease risk factors [57]. Although worksite interventions seeking to increase physical activity levels among health care workers (eg, allied health care providers and administrative staff) have been developed and implemented, few have targeted nurses specifically. Only half of these interventions significantly increased physical activity levels (eg, steps, daily min, and energy expenditure) [58]. Nurses differ from other hospital workers because they may work long shifts (ie, 8- to 12-hour shifts), irregular hours (ie, rotating day, evening, or night shifts), and undertake physically demanding tasks (eg, transfer patients between beds, chairs, and wheelchairs, reposition patients, push or pull beds, chairs, and wheelchairs, and carry equipment) [59], all of which can adversely affect their health. The traditional mode of delivering worksite interventions is face-to-face, but worksite and job characteristics may hinder nurses' ability to participate in such programs. Recognizing that nurses report low physical activity levels [60]—a known risk factor for the onset of noncommunicable diseases [61] and that the internet may offer a way to reach nurses—we developed and implemented a web-based worksite intervention for nurses working in a tertiary care cardiovascular institute.

Compared to nurses with low levels of external regulation at baseline, the nurses with medium and high levels of external regulation at baseline had greater increases in MVPA levels at the start of the intervention phase in this study. Contrary to the belief that external incentives can decrease people's motivation to participate in physical activity [35,62,63], these findings suggest that external incentives (eg, financial incentives, competition prizes, and recognition from others), pressures, and sanctions may play a role in initially increasing physical activity levels. However, the effects of external incentives appear to be beneficial only in the short term, as the nurses' physical activity levels were not maintained at the end of the intervention phase (Figure 2). As previously observed by other researchers [62,63], external incentives, pressures, and sanctions may undermine people's self-determined motivation to participate in physical activity, which raises questions concerning the use of such strategies to help nurses *maintain* physical activity levels over time. More research is needed to understand whether and when external incentives, pressures, and sanctions could be used to increase nurses' physical activity levels. It is possible that they help nurses who are not regularly active commence activities until they recognize and enjoy the intrinsic rewards that accompany physical activity (eg, healthy weight, better sleep, stress management, and improved psychological health).

Motivation has been shown to positively influence physical activity behavior when pursuing an activity that is deemed personally valuable and in which it is important to attain a desired outcome [33]. In support of these findings, the nurses in this study who possessed medium and high levels of identified regulation at baseline tended to have greater initial increases in MVPA. This suggests that it is necessary to help nurses recognize and enjoy the physical, psychological, and social benefits that accompany physical activity to help them increase their physical activity levels [33]. However, this approach may not be sufficient long-term; the findings of this study also showed that the nurses with medium and high levels of identified regulation had greater decreases in physical activity levels at the end of the intervention phase. There are several factors that may have interfered with the nurses' ability to maintain physical activity levels over time. On a personal level, nurses have often identified barriers to physical activity such as high workloads, conflicting schedules, and physical and emotional stress in their workplace. These barriers may have interfered with the nurses' ability to maintain physical activity levels in this study. It is also possible that, despite being motivated, the nurses lacked confidence and skills to sustain high physical activity levels. To test this hypothesis, interventions seeking to enhance nurses' physical activity confidence and skills by providing teaching, training, and/or counseling on goal setting, self-monitoring, and action planning should be developed and evaluated to determine if this leads to sustained changes over a longer period [64-66]. Furthermore, drawing on evidence-based behavior change techniques [36], providing: (1) coaching, (2) social support from family, friends, and staff, (3) feedback on progress and barrier identification or problem solving, (4) follow-up prompts, and (5) health checks may help to reinforce long-term changes in physical activity.

In addition to the personal factors that may have hindered sustained physical activity change in this study, targeting the entire worksite environment as opposed to the nurses within it might have increased the effectiveness of our intervention, as occupational constraints may have further inhibited the nurses' ability to maintain physical activity levels. Speculatively, worksite characteristics such as management structure, leadership, culture, and support for physical activity within the workplace may have been insufficient for the nurses in this study to translate their intention into long-term physical activity changes. Accordingly, comprehensive interventions that target both personal (eg, motivation and self-management) and macro-level factors (eg, worksite environment) may be warranted. With regard to the latter, policy interventions (eg, arranging physical activity breaks during work) or environmental changes (eg, using physical activity level prompts in common areas [break rooms, bathrooms, and elevators or stairwells], forming lunchtime physical activity groups, promoting stairway signs, and having indoor and outdoor walking routes) may help to promote sustained physical activity.

Fostering social support in the workplace has been shown to be an effective way to increase physical activity levels [66,67]. In this study, the nurses were randomized to individual, friend, or

group challenges; however, no significant differences in initial levels of or changes in physical activity levels were observed between the groups. As the identities of the friend or group members were not disclosed to the nurses, it is possible that an opportunity to foster social support was missed. Future interventionists should consider permitting the participants to know the identity of their fellow participants and facilitate social support and relatedness (rather than simply social comparison) among them. Indeed, within basic psychological needs theory [34]—another mini theory of self-determination theory—perceived relatedness (ie, experience of belongingness and connectedness to others), as well as perceived competence (ie, feelings of effectiveness and ability to achieve desired outcomes) and autonomy (ie, experience of self-determination and volition when carrying out an activity), has been identified as necessary for promoting adaptive behavioral outcomes [34,35]. Several studies indicate that individuals benefit from feeling connected to others [33].

Limitations

Although promising, the results of this study should be interpreted with caution. First, this study was conducted at a single worksite, a tertiary care cardiovascular institute, and the sample size was relatively small. Although there are similarities in some nursing roles, the local context may impact nurses' physical activity levels as a result of differing systems of nursing care, facilities, patient load, and resources. The generalizability of the findings of this study to nurses working in other health care settings and systems merits further exploration with larger sample sizes. Motivational regulations were only assessed at baseline. Thus, it is not clear to what extent the intervention impacted the nurses' motivation over time and to what extent this was related to changes in their physical activity levels. For example, the nurses' motivation may have increased or waned when comparing their activity level with others or as they gained more experience and confidence exercising over the 6-week intervention. There is a risk that the present findings reflect a selection bias as the participants were self-selecting. It is possible that the nurses who participated in this study may only be those who felt that they were healthy and fit enough to engage in a physical activity intervention and valued such activity. The nonsignificant associations between certain motivational regulations and changes in physical activity levels may be explained by the fact that the nurses in this study had relatively low (or high) scores at baseline, which may have precluded the ability to detect significant associations because of the limited variability in scores. There was some attrition, although dropout was distributed evenly across the groups. Finally, the intervention was only 6 weeks long, which may not have permitted time to facilitate long-term physical activity change.

Implications for Future Research and Practice

Understanding how best to promote physical activity among nurses remains an important endeavor, as their low physical activity levels suggest that they are at increased risk of chronic diseases and, consequently, are at higher risk of being absent from work. To allow for a better consideration of the potential impact of interventions on nurses' physical activity levels, more nurse-only intervention studies drawing on theories from the fields of psychology, sociology, behavioral economics, and/or management are needed to identify personal, situational, environmental, structural, and lifestyle factors that influence participation and effectiveness. For example, as the nurses randomized to the friend or team challenge groups might have formed expectations of how much physical activity they ought to accumulate based, in part, on how others in their group perform and made changes to their behavior accordingly, researchers could draw on social comparison theory [39] and test the role of social comparisons. In addition, given that the nurses' working environment and job characteristics can have detrimental effects [59,68-70], the extent to which nurses' workload, responsibilities, and working hours (eg, shift length and type of shift worked) influence their ability to engage in physical activity and the effectiveness of physical activity interventions should be studied. Finally, as increases in physical activity levels were not maintained over the course of the intervention, further research is clearly warranted to determine how web-based worksite interventions for nurses can be improved to support long-term changes in physical activity. On the basis of previous research [30,71-75], providing (1) individually tailored lifestyle advice, (2) physical activity plans and targets, (3) information on the benefits of physical activity, (4) physical activity self-monitoring devices, (5) interactive lectures, (6) weekly aerobic exercise classes, and/or (7) short exercise breaks at work should be considered. Furthermore, integrating behavior change techniques, implementing cognitive behavioral training, and manipulating the worksite may increase the effectiveness of the intervention [64,76]. Thus, future research to assess the effectiveness of physical activity interventions should be (1) tailored to the nurses' individual needs, (2) address macro-level changes (ie, policy changes and environmental modifications), and (3) designed, implemented, and evaluated based on theory.

Conclusions

Although the International Council of Nurses has called for nurses to make "a personal commitment to eat healthily, exercise appropriately, drink sensibly and avoid the use of tobacco" [77] and the growing expectation that nurses should embody those behaviors they wish to promote [78], most nurses report low physical activity levels [14,79,80]. Despite this, worksite interventions aimed at increasing physical activity levels among nurses are scarce. Moreover, little consideration has been given to factors that may predict changes in physical activity levels within an intervention. The principal conclusion of this study is that external and identified regulations for physical activity predicted changes in objectively measured physical activity levels. Accordingly, strategies to promote motivation for physical activity, external and integrated regulations in particular, should be part of larger strategies to promote physical activity in future interventions. Nevertheless, as initial increases were not maintained over time, the findings also highlight that changing nurses' long-term physical activity behavior is difficult and requires continued effort. Work-related circumstances (eg, job strain, nurse shortages, workload, long hours, and night or irregular shifts) may introduce barriers (eg, fatigue and lack of time) for physical activity. It is necessary to continue to investigate both personal and occupational factors that could

help the nurses sustain physical activity levels in the long term. Using qualitative methods (eg, in-depth interviews, focus group discussions, and observations) may aid in the understanding of such factors and provide insight into what the nurses themselves thought of the intervention. Finally, key stakeholders should be involved in the development, implementation, and evaluation of future worksite physical activity interventions for nurses to ensure they are feasible, sustainable, and adaptable to specific workplace demands.

Acknowledgments

The authors wish to thank all the nurses who participated in this study and the staff at the University of Ottawa Heart Institute for their support in conducting this study. This was an investigator-initiated research. Funding was provided by the University of Ottawa Heart Institute to purchase the equipment required for this study. The last author (JR) is currently supported by a New Investigator Award in Clinical Rehabilitation by the Canadian Institutes of Health Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online Tractivity® program which displayed participants distance, steps, active time and calories expended on an hourly, daily, weekly, and monthly basis.
[PDF File (Adobe PDF File), 69 KB - jmir_v22i9e11543_app1.pdf]

Multimedia Appendix 2
Friend challenge in online Tractivity® program which displayed the total distance and steps of another participant randomized to the friend challenge.
[PDF File (Adobe PDF File), 44 KB - jmir_v22i9e11543_app2.pdf]

Multimedia Appendix 3
Team challenge in online Tractivity® program which displayed the total distance and steps of others teams randomized to the team challenge.
[PDF File (Adobe PDF File), 47 KB - jmir_v22i9e11543_app3.pdf]

References

1. Ströhle A. Physical activity, exercise, depression and anxiety disorders. J Neural Transm 2009;116(6):777-784. [doi: 10.1007/s00702-008-0092-x] [Medline: 18726137]
2. Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. Am J Prev Med 2013;45(5):649-657. [doi: 10.1016/j.amepre.2013.08.001] [Medline: 24139780]
3. Warburton DE, Bredin SS. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
4. Maslach C, Schaufeli WB, Leiter MP. Job burnout. Annu Rev Psychol 2001;52:397-422. [doi: 10.1146/annurev.psych.52.1.397] [Medline: 11148311]
5. Toker S, Biron M. Job burnout and depression: unraveling their temporal relationship and considering the role of physical activity. J Appl Psychol 2012;97(3):699-710. [doi: 10.1037/a0026914] [Medline: 22229693]
6. Lindwall M, Gerber M, Jonsdottir IH, Börjesson M, Ahlborg G. The relationships of change in physical activity with change in depression, anxiety, and burnout: a longitudinal study of Swedish healthcare workers. Health Psychol 2014;33(11):1309-1318. [doi: 10.1037/a0034402] [Medline: 24245832]
7. Jonsdottir IH, Rödjer L, Hadzibajramovic E, Börjesson M, Ahlborg G. A prospective study of leisure-time physical activity and mental health in Swedish health care workers and social insurance officers. Prev Med 2010;51(5):373-377. [doi: 10.1016/j.ypmed.2010.07.019] [Medline: 20691721]
8. Stansfeld S, Candy B. Psychosocial work environment and mental health--a meta-analytic review. Scand J Work Environ Health 2006;32(6):443-462 [FREE Full text] [doi: 10.5271/sjweh.1050] [Medline: 17173201]
9. World Health Organization. World Health Statistics 2016: Monitoring Health for the Sustainable Development Goals (SDGs). Geneva, Switzerland: World Health Organization; 2020.
10. McVicar A. Workplace stress in nursing: a literature review. J Adv Nurs 2003;44(6):633-642. [doi: 10.1046/j.0309-2402.2003.02853.x] [Medline: 14651686]
11. Jamal M, Baba VV. Job stress and burnout among Canadian managers and nurses: an empirical examination. Can J Public Health 2000;91(6):454-458 [FREE Full text] [Medline: 11200738]

12. Aiken LH, Clarke SP, Sloane DM, Sochalski JA, Busse R, Clarke H, et al. Nurses' reports on hospital care in five countries. Health Aff 2001;20(3):43-53. [doi: 10.1377/hlthaff.20.3.43] [Medline: 11585181]
13. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. J Am Med Assoc 2002;288(16):1987-1993. [doi: 10.1001/jama.288.16.1987] [Medline: 12387650]
14. Reed JL, Prince SA, Pipe AL, Attallah S, Adamo KB, Tulloch HE, et al. Influence of the workplace on physical activity and cardiometabolic health: results of the multi-centre cross-sectional champlain nurses' study. Int J Nurs Stud 2018;81:49-60 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.02.001] [Medline: 29499395]
15. Blake H, Malik S, Mo PK, Pisano C. 'Do as say, but not as I do': are next generation nurses role models for health? Perspect Public Health 2011;131(5):231-239. [doi: 10.1177/1757913911402547] [Medline: 21999028]
16. McElligott D, Siemers S, Thomas L, Kohn N. Health promotion in nurses: is there a healthy nurse in the house? Appl Nurs Res 2009;22(3):211-215. [doi: 10.1016/j.apnr.2007.07.005] [Medline: 19616170]
17. Phiri LP, Draper CE, Lambert EV, Kolbe-Alexander TL. Nurses' lifestyle behaviours, health priorities and barriers to living a healthy lifestyle: a qualitative descriptive study. BMC Nurs 2014;13(1):38 [FREE Full text] [doi: 10.1186/s12912-014-0038-6] [Medline: 25506262]
18. Blake H, Harrison C. Health behaviours and attitudes towards being role models. Br J Nurs 2013;22(2):86-94. [doi: 10.12968/bjon.2013.22.2.86] [Medline: 23587891]
19. Caruso CC. Negative impacts of shiftwork and long work hours. Rehabil Nurs 2014;39(1):16-25 [FREE Full text] [doi: 10.1002/rnj.107] [Medline: 23780784]
20. Trends in Own Illness- or Disability-Related Absenteeism and Overtime among Publicly-Employed Registered Nurses: Quick Facts 2017. Canada's Nurses. URL: https://nursesunions.ca/wp-content/uploads/2017/05/Quick_Facts_Absenteeism-and-Overtime-2017-Final.pdf [accessed 2018-07-10] [WebCite Cache ID 70o2wq5yi]
21. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
22. Cotie LM, Prince SA, Elliott CG, Ziss MC, McDonnell LA, Mullen KA, et al. The effectiveness of eHealth interventions on physical activity and measures of obesity among working-age women: a systematic review and meta-analysis. Obes Rev 2018;19(10):1340-1358. [doi: 10.1111/obr.12700] [Medline: 30156044]
23. van den Berg MH, Schoones JW, Vlieland TP. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]
24. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
25. Neville LM, O'Hara B, Milat A. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. Int J Behav Nutr Phys Act 2009;6:30 [FREE Full text] [doi: 10.1186/1479-5868-6-30] [Medline: 19490649]
26. van Wier MF, Ariëns GA, Dekkers JC, Hendriksen IJ, Smid T, van Mechelen W. Phone and e-mail counselling are effective for weight management in an overweight working population: a randomized controlled trial. BMC Public Health 2009;9:6 [FREE Full text] [doi: 10.1186/1471-2458-9-6] [Medline: 19134171]
27. Hughes SL, Seymour RB, Campbell RT, Shaw JW, Fabiyi C, Sokas R. Comparison of two health-promotion programs for older workers. Am J Public Health 2011;101(5):883-890 [FREE Full text] [doi: 10.2105/AJPH.2010.300082] [Medline: 21421955]
28. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007;9(2):e17 [FREE Full text] [doi: 10.2196/jmir.9.2.e17] [Medline: 17581811]
29. Sternfeld B, Block C, Quesenberry CP, Block TJ, Husson G, Norris JC, et al. Improving diet and physical activity with ALIVE: a worksite randomized trial. Am J Prev Med 2009;36(6):475-483. [doi: 10.1016/j.amepre.2009.01.036] [Medline: 19460655]
30. Hess I, Borg J, Rissel C. Workplace nutrition and physical activity promotion at Liverpool hospital. Health Promot J Austr 2011;22(1):44-50. [doi: 10.1071/he11044] [Medline: 21717837]
31. Irvine L, Barton GR, Gasper AV, Murray N, Clark A, Scarpello T, et al. Cost-effectiveness of a lifestyle intervention in preventing type 2 diabetes. Int J Technol Assess Health Care 2011;27(4):275-282. [doi: 10.1017/S0266462311000365] [Medline: 22004767]
32. Slootmaker SM, Chinapaw MJ, Schuit AJ, Seidell JC, van Mechelen W. Feasibility and effectiveness of online physical activity advice based on a personal activity monitor: randomized controlled trial. J Med Internet Res 2009;11(3):e27 [FREE Full text] [doi: 10.2196/jmir.1139] [Medline: 19674956]
33. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
34. Deci E, Ryan R. Intrinsic Motivation and Self-determination in Human Behavior. New York, USA: Springer; 1985.

35. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
36. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
37. Leary M. Self-presentational processes in exercise and sport. J Sport Exerc Psychol 1992;14(4):339-351 [FREE Full text] [doi: 10.1123/jsep.14.4.339]
38. Hausenblas HA, Brewer BW, van Raalte JL. Self-presentation and exercise. J Appl Sport Psychol 2004;16(1):3-18 [FREE Full text] [doi: 10.1080/10413200490260026]
39. Festinger L. A theory of social comparison processes. Human Relat 2016;7(2):117-140. [doi: 10.1177/001872675400700202]
40. Reed JL, Cole CA, Ziss MC, Tulloch HE, Brunet J, Sherrard H, et al. The impact of web-based feedback on physical activity and cardiovascular health of nurses working in a cardiovascular setting: a randomized trial. Front Physiol 2018;9:142. [doi: 10.3389/fphys.2018.00142] [Medline: 29559917]
41. Wilson PM, Sabiston CM, Mack DE, Blanchard CM. On the nature and function of scoring protocols used in exercise motivation research: an empirical study of the behavioral regulation in exercise questionnaire. Psychol Sport Exerc 2012;13(5):614-622 [FREE Full text] [doi: 10.1016/j.psychsport.2012.03.009]
42. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exerc Psychol 2004;26(2):191-196 [FREE Full text] [doi: 10.1123/jsep.26.2.191]
43. Wilson PM, Rodgers WM, Fraser SN, Murray TC. Relationships between exercise regulations and motivational consequences in university students. Res Q Exerc Sport 2004;75(1):81-91. [doi: 10.1080/02701367.2004.10609136] [Medline: 15532364]
44. Rose EA, Markland D, Parfitt G. The development and initial validation of the exercise causality orientations scale. J Sports Sci 2001;19(6):445-462. [doi: 10.1080/0264041013001493393] [Medline: 11411780]
45. Wilson PM, Rodgers WM, Blanchard CM, Gessell J. The relationship between psychological needs, self‐determined motivation, exercise attitudes, and physical fitness. J Appl Soc Psychol 2006;33(11):2373-2392 [FREE Full text] [doi: 10.1111/j.1559-1816.2003.tb01890.x]
46. Warburton DE, Jeklin A, Shannon S, Bredin D. The validity of the tractivity motion sensor during walking. Health Fit J Can 2013;6(3):114-122. [doi: 10.14288/hfjc.v6i3.163]
47. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
48. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2011.
49. Canadian Society for Exercise Physiology. Canadian physical activity guidelines for adults. Can J Appl Physiol 1993;18(4):395-442. [doi: 10.1139/h93-034]
50. Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps per day in US adults. Med Sci Sports Exerc 2009;41(7):1384-1391. [doi: 10.1249/MSS.0b013e318199885c] [Medline: 19516163]
51. Nezlek JB. An introduction to multilevel modeling for social and personality psychology. Social Pers Psych Compass 2008;2(2):842-860 [FREE Full text] [doi: 10.1111/j.1751-9004.2007.00059.x]
52. Snijders T, Bosker R. Multilevel Analysis: an Introduction to Basic and Applied Multilevel Analysis. London, UK: Sage Publications; 2012.
53. Templeton GF. A two-step approach for transforming continuous variables to normal: implications and recommendations for IS research. Commun Assoc Inf Syst 2011;28:41-58. [doi: 10.17705/1cais.02804]
54. Kenny DA, Kashy D, Bolger N. Data analysis in social psychology. In: Gilbert S, Fiske S, Lindzey G, editors. The Handbook of Social Psychology. New York, USA: McGraw-Hill; 1998:233-265.
55. Preacher KJ, Curran PJ, Bauer DJ. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. J Educ Behav Stat 2016;31(4):437-448 [FREE Full text] [doi: 10.3102/10769986031004437]
56. Curran P, Bauer D, Willoughby M. Testing and probing interactions in hierarchical linear growth models. In: Bergeman CS, Boker SM, editors. Methodological Issues in Aging Research (Notre Dame Series on Quantitative Methodology). Mahwah, NJ: Lawrence Erlbaum Associate; 2006:99-129.
57. Quintiliani L, Sattelmair J, Sorensen G. The Workplace as a Setting for Interventions to Improve Diet and Promote Physical Activity. Geneva, Switzerland: World Health Organization; 2007.
58. Torquati L, Pavey T, Kolbe-Alexander T, Leveritt M. Promoting diet and physical activity in nurses. Am J Health Promot 2017;31(1):19-27. [doi: 10.4278/ajhp.141107-LIT-562] [Medline: 26389980]
59. Geiger-Brown J, Trinkoff AM. Is it time to pull the plug on 12-hour shifts?: part 1. The evidence. J Nurs Adm 2010;40(3):100-102. [doi: 10.1097/NNA.0b013e3181d0414e] [Medline: 20485206]
60. Malik S, Blake H, Batt M. How healthy are our nurses? New and registered nurses compared. Br J Nurs 2011;20(8):489-496. [doi: 10.12968/bjon.2011.20.8.489] [Medline: 21537281]

61. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
62. Standage M, Duda JL, Ntoumanis N. A model of contextual motivation in physical education: using constructs from self-determination and achievement goal theories to predict physical activity intentions. J Educ Psychol 2003;95(1):97-110. [doi: 10.1037/0022-0663.95.1.97]
63. Gunnell KE, Crocker PR, Mack DE, Wilson PM, Zumbo BD. Goal contents, motivation, psychological need satisfaction, well-being and physical activity: a test of self-determination theory over 6 months. Psychol Sport Exerc 2014;15(1):19-29 [FREE Full text] [doi: 10.1016/j.psychsport.2013.08.005]
64. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
65. Pearson ES. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. Patient Educ Couns 2012;87(1):32-42. [doi: 10.1016/j.pec.2011.07.018] [Medline: 21852063]
66. Jirathananuwat A, Pongpirul K. Promoting physical activity in the workplace: a systematic meta-review. J Occup Health 2017;59(5):385-393 [FREE Full text] [doi: 10.1539/joh.16-0245-RA] [Medline: 28740029]
67. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
68. Torquati L, Kolbe-Alexander T, Pavey T, Persson C, Leveritt M. Diet and physical activity behaviour in nurses: a qualitative study. Int J Health Promot Educ 2016;54(6):268-282 [FREE Full text] [doi: 10.1080/14635240.2016.1169943]
69. Huntington A, Gilmour J, Tuckett A, Neville S, Wilson D, Turner C. Is anybody listening? A qualitative study of nurses' reflections on practice. J Clin Nurs 2011;20(9-10):1413-1422. [doi: 10.1111/j.1365-2702.2010.03602.x] [Medline: 21414055]
70. Nahm E, Warren J, Zhu S, An M, Brown J. Nurses' self-care behaviors related to weight and stress. Nurs Outlook 2012;60(5):e23-e31. [doi: 10.1016/j.outlook.2012.04.005] [Medline: 22770679]
71. Furukawa F, Kazuma K, Kawa M, Miyashita M, Niiro K, Kusukawa R, et al. Effects of an off-site walking program on energy expenditure, serum lipids, and glucose metabolism in middle-aged women. Biol Res Nurs 2003;4(3):181-192. [doi: 10.1177/1099800402239623] [Medline: 12585782]
72. Pfister PB, Niedermann K, Sidelnikov E, Bischoff-Ferrari HA. Active over 45: a step-up jogging programme for inactive female hospital staff members aged 45+. Eur J Public Health 2013;23(5):817-822. [doi: 10.1093/eurpub/ckt027] [Medline: 23504626]
73. Strijk JE, Proper KI, van der Beek AJ, van Mechelen W. A worksite vitality intervention to improve older workers' lifestyle and vitality-related outcomes: results of a randomised controlled trial. J Epidemiol Community Health 2012;66(11):1071-1078 [FREE Full text] [doi: 10.1136/jech-2011-200626] [Medline: 22268128]
74. McElligott D, Capitulo KL, Morris DL, Click ER. The effect of a holistic program on health-promoting behaviors in hospital registered nurses. J Holist Nurs 2010;28(3):175-183. [doi: 10.1177/0898010110368860] [Medline: 20595638]
75. Brox JI, Frøystein O. Health-related quality of life and sickness absence in community nursing home employees: randomized controlled trial of physical exercise. Occup Med 2005;55(7):558-563. [doi: 10.1093/occmed/kqi153] [Medline: 16251373]
76. Lemon SC, Zapka J, Li W, Estabrook B, Rosal M, Magner R, et al. Step ahead a worksite obesity prevention trial among hospital employees. Am J Prev Med 2010;38(1):27-38 [FREE Full text] [doi: 10.1016/j.amepre.2009.08.028] [Medline: 20117554]
77. Delivering Quality, Serving Communities: Nurses Leading Chronic Care. HRH Global Resource Center. 2008. URL: https://www.hrhresourcecenter.org/node/5094.html [accessed 2020-08-17]
78. While AE. Are nurses fit for their public health role? Int J Nurs Stud 2014;51(9):1191-1194. [doi: 10.1016/j.ijnurstu.2014.01.008] [Medline: 24513293]
79. Ratner PA, Sawatzky R. Health status, preventive behaviour and risk factors among female nurses. Health Rep 2009;20(3):53-61 [FREE Full text] [Medline: 19813439]
80. Tucker SJ, Harris MR, Pipe TB, Stevens SR. Nurses' ratings of their health and professional work environments. AAOHN J 2010;58(6):253-267. [doi: 10.3928/08910162-20100526-03] [Medline: 20677722]

Abbreviations

BREQ-2: Behavioral Regulation in Exercise Questionnaire-2
MVPA: moderate-to-vigorous physical activity

Edited by G Eysenbach; submitted 11.07.18; peer-reviewed by J Connelly, D Pekmezi, J Gay; comments to author 27.10.18; revised version received 27.12.18; accepted 03.09.19; published 11.09.20.

Please cite as:
Brunet J, Tulloch HE, Wolfe Phillips E, Reid RD, Pipe AL, Reed JL
Motivation Predicts Change in Nurses' Physical Activity Levels During a Web-Based Worksite Intervention: Results From a Randomized Trial
J Med Internet Res 2020;22(9):e11543
URL: http://www.jmir.org/2020/9/e11543/
doi:10.2196/11543
PMID:32915158

Original Paper

Digital Health Tools and Patients With Drug Use Disorders: Qualitative Patient Experience Study of the Electronic Case-Finding and Help Assessment Tool (eCHAT)

Melinda Ada Choy[1], BMed, MD, DCH, MMed, FRACGP, AFHEA; Elizabeth Sturgiss[2,3], BMed, MPH, MForensMed, PhD, FRACGP, FHEA; Felicity Goodyear-Smith[4], MBChB, MD, FRNZCGP, FFFLM; Gavin JD Smith[5], MA, MRes, PhD

[1]Academic Unit of General Practice, College of Health and Medicine, The Australian National University, Garran, Australia

[2]Research School of Population Health, College of Health and Medicine, The Australian National University, Garran, Australia

[3]Department of General Practice, Monash University, Melbourne, Australia

[4]Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand

[5]School of Sociology, The Australian National University, Canberra, Australia

Corresponding Author:
Melinda Ada Choy, BMed, MD, DCH, MMed, FRACGP, AFHEA
Academic Unit of General Practice
College of Health and Medicine
The Australian National University
Building 4, Level 2
The Canberra Hospital
Garran, 2605
Australia
Phone: 61 5124 5592
Email: melinda.choy@anu.edu.au

Abstract

Background: One of the promises of digital health is to better engage patients and improve care for vulnerable populations. Patients with drug use disorders are a vulnerable population who often do not receive the care they need, both for their drug use disorders as well as their other health care needs. Appropriate primary care for patients with drug use disorders needs to be patient-centered, holistic, highly accessible, and engaging. The electronic Case-finding and Help Assessment Tool (eCHAT) was designed as a patient-centered tool for the identification and measurement of problematic health behaviors and mood states.

Objective: The aim of this study was to explore the patient experience of eCHAT at an Australian family medicine clinic for patients with drug use disorders.

Methods: A total of 12 semistructured interviews were conducted with patients, two interviews were conducted with doctors, and one focus group was conducted with patient advocates who were former patients of the clinic where the study took place. The transcripts were analyzed using inductive thematic analysis.

Results: The key themes identified from the interviews and the focus group were as follows: (1) eCHAT helped reduce stigma related to drug use in the doctor-patient consultation, (2) restricted answer options impacted the ability of patients to tell their stories, (3) patient-related response factors, (4) increased efficiency in the consultation process, and (5) divergence in level of concern around security and privacy.

Conclusions: eCHAT has the potential to help vulnerable patients in primary care to engage more with their doctors and reduce experiences of stigma. eCHAT may be a useful digital health intervention in a family medicine clinic for patients with drug use disorders. It has the potential to improve patient engagement and access to health care, which are crucial areas of need in this vulnerable population. However, it is important to clearly communicate the privacy risk of digital health tools and to implement eCHAT such that it will add value to, rather than displace, in-person consultations with the family doctor.

(J Med Internet Res 2020;22(9):e19256) doi:10.2196/19256

KEYWORDS

eCHAT; eHealth; mHealth; digital health; drug use disorders; patient experience; stigma; patient experience; family medicine; general practice

Introduction

The Promise of eHealth

The expanding presence of digital health and eHealth is driven by its potential to improve health care outcomes [1]. eHealth refers to the use of internet-based technology for health care and can be used by systems, providers, and/or patients [2]. eHealth is one component of the wider concept of digital health, which is the use of information and communication technology to improve patient well-being and health [3]. Some examples of eHealth used primarily by patients in Australia are electronic medical records, searching for online health information, and booking appointments online.

eHealth has the potential to increase health care access and better engage diverse groups of patients [4]. If eHealth can particularly improve access for vulnerable patients, one key anticipated outcome of increasing eHealth use should be an improvement in health care equity [4].

The Health Care Needs of Patients With Drug Use Disorders

One example of a vulnerable population affected by health care inequity is people with drug use disorders. They often do not receive adequate treatment support and are more likely to have disability and reduced social and emotional functioning [5,6]. Patients with drug use disorders require customized health care relevant to their drug use situation [5].

Patients with drug use disorders are also more likely to have mental health diagnoses and other chronic diseases [7]. Consequently, patients with drug use disorders often require effective, customized, and sustained general health care, which can be provided through primary care [7].

However, only a small proportion of patients with drug use disorders receive the care they require. One in six people suffering from drug use disorders received treatment for those disorders in 2016 [5]. They also have greater difficulty in access and engagement with general health care services for their family medicine needs [8].

Therefore, the challenge is to design more agile health care services that meet the specific and general needs of patients with drug use disorders [9]. It is crucial that health care services for patients with drug use disorders improve patient engagement and access, as this is a population group with higher health care needs but who also receive proportionally fewer health care services [10].

Can eHealth Improve Health Care for Patients With Drug Use Disorders?

The critical question is whether eHealth can truly facilitate better engagement and health care access for patients with drug use disorders. The majority of studies involving eHealth and patients with drug use disorders focus on the benefits of telemedicine for treating drug use disorders [11]. Telehealth is a branch of eHealth that creates more *mobile* connections between health care providers and patients [12]. A number of systematic reviews suggest that telemedicine, as well as other eHealth tools, can be effective for improving drug use disorders and creating patient satisfaction [11,13,14].

However, there is scarce literature that explores the role of eHealth in the general health care needs of patients with drug use disorders, particularly in the context of family medicine. Further, there is minimal research into the patient experience of, and engagement with, eHealth in those with drug use disorders. For a patient population with drug use disorders, it is particularly important to understand the patient experience of new eHealth interventions, as they engage health care services less frequently and sometimes with difficulty [8].

One potential example of eHealth in the family medicine setting is the electronic Case-finding and Help Assessment Tool (eCHAT), created in New Zealand [15]. eCHAT is designed as a patient-centered tool for the identification and measurement of problematic health behaviors and mood states. It is a screening survey completed by patients on a tablet computer in the waiting room of a family medicine clinic, with the results provided to the doctor at point of care prior to the consultation [15]. Pilot studies of eCHAT in New Zealand have had positive results in feasibility and acceptability studies [15]. However, prior to this study, eCHAT had not been trialed in Australia, nor had it been trialed at a family medicine clinic with a focus on vulnerable patients with drug use disorders.

Objective

Our objective was to explore the patient experience of eCHAT when it was trialed in an Australian family medicine clinic for patients with drug use disorders. We were particularly interested in understanding the impact of eCHAT on the patient's consultation experience and relationship with their family doctor, as these are key factors for health care access for this patient population.

Methods

Procedure

From May to November 2019, receptionists offered patients the option to complete eCHAT before their consultation. For each consenting patient, receptionists entered in their unique clinic identifier number into eCHAT on the tablet computer and then gave the patient the tablet. The patient completed eCHAT in the waiting room for 10 minutes and the information went onto a secure web server.

Patients who completed eCHAT answered questions about their recent drug use, mood, and physical activity. They indicated whether they wished to discuss anything further. The doctor reviewed their patient's answers by searching the patient's

unique clinic identifier number on eCHAT using their desktop computer, just before initiating the patient-doctor consultation.

Participants

We undertook semistructured interviews with patients and doctors who had used eCHAT. We also ran a focus group with former patients of the clinic who now worked as patient advocates but had not used eCHAT themselves. On days that the research team visited, receptionists invited patients to participate in the study who were (1) over 18 years old, (2) had the time to start eCHAT before their consultation with their doctor and, (3) had the time to participate in a 30-minute interview after their consultation with their doctor.

We used purposive sampling by actively selecting the most productive sample to answer the research question. We recruited participants seeing a diverse range of doctors, a mixture of established and new patients, and a diverse demographic. We continually reflected on our sample and whom we sought to recruit next in coding meetings.

Data Collection

The Australian National University Human Research Ethics Committee approved this study. Prior to the interview, participants provided written informed consent and completed a short demographic questionnaire. The interview guide was informed by a systematic literature review. After a pilot interview conducted by the first author (MC), the interview guide was further reviewed by an experienced coauthor (GS). The full authorial team collaboratively contributed to the refinement of the interview guide at coding meetings. Interviews and the focus group were audio recorded and transcribed anonymously by a third-party transcription service. The demographic survey results were recorded in Microsoft Excel. The transcription data were managed in both NVivo 11 (QSR International) and Microsoft Word.

Data Analysis

The interviews and focus group data were coded and processed by inductive thematic analysis. Two researchers (MC, ES) independently reviewed and coded the first five interview transcripts to ensure intercoder validity and reliability. Three researchers (MC, ES, GS) then met to reflect on and discuss the different codes, early themes, sampling frame, and the interview guide.

Following that, one researcher (MC) coded the rest of the interviews and selected key transcripts for the second researcher (ES) to independently code. Both researchers independently coded the doctor interviews and focus group data. Four coding meetings were held with different arrangements of researchers in order to continually reflect on and resolve disagreement over codes and themes. At the final coding meeting, a final list of themes was developed and agreed upon by all four authors.

Results

Participant Characteristics

The demographic characteristics of the patients we interviewed are shown in Table 1. The doctors who were interviewed included one male and one female participant. The focus group with patient advocates included three female participants and one male participant.

Table 1. Demographic characteristics of patients who were interviewed.

Characteristic	Value (N=12), n (%)
Gender	
Female	2 (17)
Male	10 (83)
Highest educational qualification	
Lower than year 10	2 (17)
Year 10	3 (25)
Year 12	1 (8)
Trade certification	3 (25)
Diploma	1 (8)
Bachelor’s degree	2 (17)
Occupation	
Not employed	4 (33)
Pension	4 (33)
Other	4 (33)

Themes

Overview

The key themes from the interviews and focus group were as follows: (1) eCHAT helped reduce stigma related to drug use in the doctor-patient consultation, (2) restricted answer options impacted the ability of patients to tell their stories, (3) patient-related response factors, (4) increased efficiency in the consultation process, and (5) divergence in level of concern around security and privacy.

Theme 1: eCHAT Helped Reduce Stigma Related to Drug Use in the Doctor-Patient Consultation

Some patients expressed that it was easier to be comfortable with a screen, compared to a person, when answering questions regarding their drug use and mental health.

> *Because maybe the person is not comfortable coming up and saying it to their doctor, straight to his face. But if they feel comfortable with that [eCHAT]...* [Patient #2, interview]

Further, some patients explained that they would be more honest, because interacting with a screen, rather than directly with a person, removed some of the shame associated with their drug use.

> *I know when I first came to terms with my struggles, I was so ashamed and so guilty I would lie and deny...I still would be so ashamed of why I was there that I would make it seem not as bad as what it actually was. Whereas if I was writing it down—if it was in question form like that—I would just be honest.* [Patient #5, interview]

One patient explained that the reason that it was easier to be honest with a screen was because a screen is less judgmental than a person.

> *Whereas with an electronic device, I'm just going to be completely honest with it. I've got no reason to—you know, it's not going to judge me or expect anything from me.* [Patient #3, interview]

Patients also felt that their doctors got more information from eCHAT, because eCHAT helped patients themselves tell their stories better. Some patients felt this was because using eCHAT helped them focus on what they wanted to speak to the doctor about.

> *...it made me separate my thoughts and that just to, it just felt, it made me sort of tell myself something in the doctor's appointment.* [Patient #7, interview]

Other patients commented that eCHAT gave them more confidence to have a starting point with their story.

> *...at least it's a starting point; you fill out something on there, then it'll give you more reason, it'll make you feel more confident talking to the doctor... [Patient #6, interview]*

Theme 2: Restricted Answer Options Impacted the Ability of Patients to Tell Their Stories

A large number of patients commented on how the restricted multiple-choice answer options on the eCHAT questionnaire limited their ability to fully explain their context and situation. Both patients who felt positively and those who felt negatively about eCHAT mentioned that eCHAT's key weakness was its limited answer options. Particularly in relation to drug use, patients wanted to be able to explain, beyond quantitative questions, about timing frequency.

> *Everybody will tell you that they smoke too much. Everybody will tell you that. But it's about "Can you quit, do you want to quit, how come you haven't?"* [Patient #11, interview]

Some patients felt that the restricted answers meant that their answers were less likely to be correct.

> *But I did find that some of the questions, I had to kind of pick one, even though I didn't feel like there was—I didn't feel like that was correct, either. It didn't seem like there was an option that I needed.* [Patient #4, interview]

One patient commented that the narrow scope of some questions presented an opportunity to be dishonest.

> *You've dabbled in drugs nearly your whole life, but if you're only answering it the last two weeks, so you just tick that and you go, "Oh, yeah, no worries," and then, you know, you're not saying anything.* [Patient #9, interview]

The idea that eCHAT was not providing a full story frustrated some patients. Some patients felt that this restriction meant that it was far more efficient and satisfying to speak to a doctor directly instead. There was a desire for a qualitative component that would grant the patient more autonomy to co-construct meaning.

> *It frustrated me that I couldn't have my own say, I couldn't, I had to choose one of the options, I couldn't actually have what actually happened...no one's going to fit into a little box. Especially someone like me. So I guess when you have little boxes and you don't have one you fit in, it gets frustrating after a while.* [Patient #1, interview]

Other patients felt that the restricted answers amplified stigma by placing everyone in the same box based on drug use, rather than understanding the context, story, and factors that mediate the practices of each person.

> *Well, the government puts everyone like us in the same box. We're addicts or whatever we are—we've been to jail...I think we should all be judged by our own merits—our own problems and that—and we shouldn't be boxed together.* [Patient #2, interview]

Theme 3: Patient-Related Response Factors

Some patients had physical disabilities, mostly related to eyesight, that affected their ability to fill out the eCHAT questionnaire. A few patients felt emotionally confronted by

some of the questions. One patient started crying when questions were asked about depression, and another patient felt that some of the questions about drug use raised traumatic memories.

> *They all sort of hit a sensitive spot inside. It brings up emotions to answer some of those questions...* [Patient #10, interview]

Patients who did not have an established relationship with their doctor sometimes felt that eCHAT improved their relationship with their doctor, but no one felt that eCHAT negatively affected their relationship with their doctor. Patients with an established patient-doctor relationship felt that eCHAT made no difference to that relationship.

The doctors reported that eCHAT was useful for more stable patients who had less urgent agendas, because eCHAT was then able to pick up important, but not pressing, information that had not been previously covered.

> *...from the kind of chaotic life and you're just dealing with crises, to the...you've gradually got someone engaged and you're starting to work on those things, that's the group that I think it's [eCHAT is] probably most useful with.* [Doctor #1, interview]

The patient advocates who participated in the focus group strongly agreed that a patient's stage in the recovery process was the most important factor in their willingness to be honest and to benefit from eCHAT.

> *If they're very invested in recovery, then they'll answer it truthfully either way [electronically or in person].* [Patient advocate #1, focus group]

Theme 4: Increased Efficiency in the Consultation Process

Most patients generally felt positive about the potential for eCHAT to harness waiting room time and provide their doctor with information that would direct or streamline the consultation.

> *Yeah, I think it's really good. I think it would save heaps of time...Nobody wants to sit in the doctor's waiting room all day. I think the faster the appointments are, it's better for everyone. I think that's what eCHAT would do, is speed up the whole process.* [Patient # 5, interview]

However, some patient advocates in the focus group felt that eCHAT would actually result in an overall loss of time, as patients would need to explain the meaning of their restricted answers.

> *I think the time that you save, not having to ask these questions personally in the room, I think there's a big opportunity to lose that time that you've saved by having to qualify all those questions.* [Patient advocate #1, focus group]

Theme 5: Divergence in Level of Concern Around Security and Privacy

Most of the patients interviewed did not have strong security concerns or register any awareness of how their personal data could be misused.

> *Why would somebody bother hacking into something like that? Yeah, it doesn't bother me.* [Patient #4, interview]

In contrast, the patient advocates in the focus group had stronger security and privacy concerns.

> *I personally don't do anything electronic if I can help it...because I don't trust the systems in place. There are too many holes, too many things can go wrong.* [Patient advocate #4, focus group]

A few patients who did have concerns in general about security were less worried about eCHAT because they trusted their clinic and knew their name did not go into the system. In general, however, the high level of trust in the clinic and lack of security concerns for some patients meant that they were willing to complete the eCHAT questionnaire if it was made mandatory at the clinic, although they were generally uncomfortable with the concept.

> *Every time? If I had to?...They've been very good to me here, and they've offered really good services. So I'd be okay with that.* [Patient advocate #3, focus group]

Discussion

Principal Findings

This paper reports on the findings of a qualitative study exploring the patient experience of an eHealth tool in a family medicine clinic for patients with drug use disorders. Some patients felt that the platform helped them to communicate more honestly, completely, and efficiently with their doctor. A few patients also said that being able to communicate more honestly in the past would have allowed them to get help for their drug use disorders earlier and faster.

Many patients commented on the restricted nature of the multiple-choice answer options. Some felt that these restrictions meant that they were not able to tell their story properly and that the interface compressed their capacity to explain and contextualize factors. A few patients found some of the questions asked in eCHAT confronting, both positively and negatively. Some patients had trouble filling in the eCHAT questionnaire due to a range of physical disabilities.

Very few patients voiced concerns about data security and privacy regarding the information they documented. Those that were aware of possible security issues felt limited concern because no identifying details went into eCHAT.

Patients reported that the use of eCHAT before the patient-doctor consultation sometimes had a positive effect on the patient-doctor relationship, sometimes had no effect, but never generated a negative effect. The optional use of eCHAT before the patient-doctor consultation did not appear to negatively impact patient engagement or access to their family medicine–based health care services.

Strengths

One strength of this study is the focus on the experience of eHealth for patients with drug use disorders. One key way to

improve patient engagement with eHealth is to make the patient experience the guiding principle in eHealth tool design [16]. The way that eHealth increases levels of patient engagement is a common theme in eHealth research [17]. Patient engagement is about creating a more active role for the patient and a more collaborative partnership between the patient and doctor, in order to create a greater protagonist role for the patient in their health care [17,18]. A systematic review exploring how eHealth improves patient engagement found that eHealth interventions generally increase the prominence of the patient's role in their health care [17]. This matches how eCHAT helped patients communicate their situation more effectively, thereby creating a more active role for the patient and improving the collaboration with the health care provider. As patient engagement and disengagement are highly relevant factors among patients with drug use disorders, the ability of eCHAT to moderately increase patient participation and representation in health care is a promising feature of eCHAT for this vulnerable population.

Another strength of this study is that it uniquely explores the role of eHealth in the general health care needs of patients with drug use disorders, particularly in the context of family medicine. While it is already known that eHealth tools can help with the treatment of drug use disorders and may improve patient satisfaction, this shows that eCHAT could improve holistic patient-centered family medicine care for a vulnerable population of patients with drug use disorders [11,13,14]. This study showed that eCHAT helped some patient-doctor relationships to improve and did not negatively affect any patient-doctor relationships.

Our finding that eHealth tools have the potential to reduce what might be interpreted as *transactional stigma*—that is, shame and stigma arising as a consequence of patients confessing their failings to an authority figure—is significant for a patient population with drug use disorders. Patients with drug use disorders are a highly stigmatized and marginalized group, which can reduce their access to health care and social services [19]. If eCHAT can help reduce stigma and improve patient access, it means that eHealth can have a role in improving access to care for this vulnerable group.

This study also illuminated some other stigma-related concerns of patients with drug use disorders when it comes to engaging with eHealth. A phenomenological study of people with drug use disorders found that when patients felt they were listened to properly by nurses, they responded better to treatment programs and experienced reduced feelings of stigma [20]. The importance of being properly listened to may explain why the restricted answers were problematic for this patient group. Therefore, the concern over restricted answers in combination with some of the more emotionally confronting questions in eCHAT make clear that the eHealth tool was best used when paired with the face-to-face patient-doctor consultation. Pairing eCHAT with a face-to-face consultation means that the patient can have extra emotional support but also be presented with the opportunity to explain their situation in more detail so that their drug-taking practices can be better contextually understood.

Limitations

While the sample size of this study is small, it uniquely explores the patient experience of an eHealth tool in a population of patients with drug use disorders in the context of their family medicine clinic. Further, the additional data from the focus group and doctor interviews serve as a useful compliment and contrast to the patient perspective.

For example, the prominent privacy and security concerns of the focus group members were in stark contrast to the perspectives of the patients. This contrast begs the question of whether the patients, as a consequence of their requiring and seeking health care, were less able to identify and raise critical concerns about potential privacy and data security issues than their more detached and independent patient advocate counterparts. It is possible that the interview setting (ie, the clinic where the patients receive care) and the interviewer (ie, a family medicine doctor from another practice) affected the response of the patient participants during the interviews. Even if this had some impact on the interview data, the reticence of patients to accentuate concerns pertaining to the use and processing of their personal data highlights the need for administrators of eHealth to clearly communicate the potential surveillance, security, and privacy risks related to the use of digital health tools, especially for those who are most vulnerable. Adequate risk communication is particularly appropriate when seeking informed consent for the use of an eHealth tool from patients who require access to health care.

Another example of how the different data worked together was that the concerns raised by patients about the restricted answer options in eCHAT were partly offset by the doctors, who explained that they principally used eCHAT to initiate conversations with their patients.

Implementation issues meant that not many patients completed the eCHAT questionnaire separate from the study taking place, which left doctors feeling that they were limited in their ability to comment broadly on the effect of eCHAT.

Future Considerations

The challenges of eCHAT that were identified by participants have been discussed with both the family medicine clinic where the study took place and the creators of eCHAT. Regarding implementation at the family medicine clinic, the risks of upsetting patients and miscommunicating the patient story with restricted answer options make it clear that eCHAT works best as an optional process paired with, and on the same day as, the doctor-patient consultation.

Regarding the design of eCHAT, the options of voice-prompted questions for those with disabilities and open-ended questions to counterbalance concerns around restricted answers have been discussed. There are further developed versions of eCHAT that now have an introductory video that patients can watch before completing the eCHAT questionnaire, which greatly assists with clearly and accessibly communicating security and privacy risks to patients.

A future option for long-term evaluation of eCHAT could build on this study with qualitative analysis of the patient experience of eCHAT in vulnerable populations from multiple clinics.

Our findings suggest that further quantitative, qualitative, and mixed methods research into the experience of vulnerable patients when using other digital health tools may assist these tools in contributing to improving health care equity within communities.

Conclusions

eCHAT has the potential to help vulnerable patients in primary care to engage more with their doctors and reduce their sense of stigma, but attention must be paid to how consent is obtained from patients and how eCHAT is paired with the consultation.

To move this forward, future research will need to explore how other eHealth tools, and how the concept of eHealth in general, affect the experience of health care for patients with drug use disorders. Understanding the situated experiences of eHealth for specific groups of vulnerable populations can help clarify aspects of design and implementation for eHealth tools that are more likely to improve patient access and equity.

Based on this study, eCHAT could be a useful eHealth intervention in a family medicine clinic for patients with drug use disorders. It has the potential to improve patient engagement and access to health care, which are crucial areas of need. However, care must be taken when implementing eCHAT to make the surveillance, privacy, and data security risks clear to patients, and emphasis needs to be placed on how eCHAT should complement, rather than replace, the face-to-face consultation with the family doctor.

Acknowledgments

We would like to thank the patients, doctors, and patient advocate focus group members who were interviewed for this project, as well as the family medicine clinic who hosted us and its community advisory group. We are thankful also to Dr Margot Darragh from the University of Auckland, who assisted in the setup of eCHAT at the clinic. This research project was supported by the Royal Australian College of General Practitioners with funding from the Australian Government under the Australian General Practice Training Program.

Conflicts of Interest

None declared.

References

1. Srivastava S, Pant M, Abraham A, Agrawal N. The technological growth in eHealth services. Comput Math Methods Med 2015;2015:894171 [FREE Full text] [doi: 10.1155/2015/894171] [Medline: 26146515]
2. Global Diffusion of eHealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2016. URL: http://apps.who.int/iris/bitstream/10665/252529/1/9789241511780-eng.pdf?ua=1 [accessed 2020-08-28]
3. Iyawa GE, Herselman M, Botha A. Digital health innovation ecosystems: From systematic literature review to conceptual framework. Procedia Comput Sci 2016;100:244-252 [FREE Full text] [doi: 10.1016/j.procs.2016.09.149]
4. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
5. United Nations Office on Drugs and Crime. World Drug Report 2018. New York, NY: United Nations Publications; Jun 2018.
6. Grant BF, Saha TD, Ruan WJ, Goldstein RB, Chou SP, Jung J, et al. Epidemiology of DSM-5 drug use disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. JAMA Psychiatry 2016 Jan;73(1):39-47 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.2132] [Medline: 26580136]
7. Berends L, Lubman DI. Obstacles to alcohol and drug care: Are Medicare Locals the answer? Aust Fam Physician 2013 May;42(5):339-342 [FREE Full text] [Medline: 23781538]
8. Neale J, Tompkins C, Sheard L. Barriers to accessing generic health and social care services: A qualitative study of injecting drug users. Health Soc Care Community 2008 Mar;16(2):147-154. [doi: 10.1111/j.1365-2524.2007.00739.x] [Medline: 18290980]
9. Samet JH, Friedmann P, Saitz R. Benefits of linking primary medical care and substance abuse services: Patient, provider, and societal perspectives. Arch Intern Med 2001 Jan 08;161(1):85-91 [FREE Full text] [doi: 10.1001/archinte.161.1.85] [Medline: 11146702]
10. Bauman AE, Fardy HJ, Harris PG. Getting it right: Why bother with patient-centred care? Med J Aust 2003 Sep 01;179(5):253-256. [Medline: 12924973]
11. Lin L, Casteel D, Shigekawa E, Weyrich MS, Roby DH, McMenamin SB. Telemedicine-delivered treatment interventions for substance use disorders: A systematic review. J Subst Abuse Treat 2019 Jun;101:38-49. [doi: 10.1016/j.jsat.2019.03.007] [Medline: 31006553]

12. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
13. Ferreri F, Bourla A, Mouchabac S, Karila L. e-Addictology: An overview of new technologies for assessing and intervening in addictive behaviors. Front Psychiatry 2018;9:51 [FREE Full text] [doi: 10.3389/fpsyt.2018.00051] [Medline: 29545756]
14. Young LB. Telemedicine interventions for substance-use disorder: A literature review. J Telemed Telecare 2012 Jan;18(1):47-53. [doi: 10.1258/jtt.2011.110608] [Medline: 22101610]
15. Goodyear-Smith F, Warren J, Elley CR. The eCHAT program to facilitate healthy changes in New Zealand primary care. J Am Board Fam Med 2013;26(2):177-182 [FREE Full text] [doi: 10.3122/jabfm.2013.02.120221] [Medline: 23471931]
16. Graffigna G, Barello S, Triberti S, Wiederhold BK, Bosio AC, Riva G. Enabling eHealth as a pathway for patient engagement: A toolkit for medical practice. Stud Health Technol Inform 2014;199:13-21. [Medline: 24875682]
17. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: A systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
18. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: A framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb;32(2):223-231. [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
19. Lloyd C. The stigmatization of problem drug users: A narrative literature review. Drugs 2012 Nov 29;20(2):85-95. [doi: 10.3109/09687637.2012.743506]
20. Sleeper JA, Bochain SS. Stigmatization by nurses as perceived by substance abuse patients: A phenomenological study. J Nurs Educ Pract 2013 Jan 14;3(7):92-98 [FREE Full text] [doi: 10.5430/jnep.v3n7p92]

Abbreviations

eCHAT: electronic Case-finding and Help Assessment Tool

Edited by G Eysenbach; submitted 10.04.20; peer-reviewed by GE Iyawa, D Pförringer; comments to author 12.06.20; revised version received 26.06.20; accepted 19.07.20; published 14.09.20.

Please cite as:
Choy MA, Sturgiss E, Goodyear-Smith F, Smith GJD
Digital Health Tools and Patients With Drug Use Disorders: Qualitative Patient Experience Study of the Electronic Case-Finding and Help Assessment Tool (eCHAT)
J Med Internet Res 2020;22(9):e19256
URL: http://www.jmir.org/2020/9/e19256/
doi:10.2196/19256
PMID:32924959

Original Paper

Sun Safe Partners Online: Pilot Randomized Controlled Clinical Trial

Sharon Manne[1], PhD; David Buller[2], PhD; Katie Devine[1], PhD; Carolyn Heckman[1], PhD; Sherry Pagoto[3], PhD; Sara Frederick[1], MS; Anna Mitarotondo[1], BA

[1]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

[2]Klein Buendel, Inc, Golden, CO, United States

[3]University of Connecticut, Storrs, CT, United States

Corresponding Author:
Sharon Manne, PhD
Rutgers Cancer Institute of New Jersey
CINJ
195 Little Albany Street
New Brunswick, NJ, 08903
United States
Phone: 1 732 235 9780
Email: mannesl@cinj.rutgers.edu

Abstract

Background: Harnessing supportive influences in close relationships is an innovative and potentially effective strategy to improve sun protection behaviors.

Objective: This pilot randomized controlled clinical trial evaluates the feasibility and impact of Sun Safe Partners Online, a web-based, couples-focused intervention to improve sun protection behavior.

Methods: A total of 75 couples reporting suboptimal levels of sun protection recruited from Facebook advertisements were randomized to receive a web-based intervention called Sun Safe Partners Online or a Generic Online Sun Safety Information intervention. Sun Safe Partners Online had 4 individual-focused modules and 4 couples-focused modules. Feasibility was assessed by study enrollment, engagement, follow-up survey completion, and intervention evaluation. Participants completed baseline and a 1-month postintervention survey assessing sun protection and exposure, along with individual and relationship attitudes about the importance of sun protection.

Results: Using Facebook as a recruitment strategy resulted in rapid enrollment and higher acceptance than for the prior telephone and print trial. The follow-up survey completion was higher in the Generic Online condition (100%) than in the Sun Safe Partners Online condition (87.2%). Engagement in Sun Safe Partners Online was high, with more than two-thirds of participants completing all modules. Evaluations of Sun Safe Partners Online content and features as well as ease of navigation were excellent. Sun Safe Partners Online showed small effects on sun protection behaviors and sun exposure on weekends compared with the Generic Online intervention and moderate effect size increases in the Sun Safe Partners Online condition.

Conclusions: This study uses a novel approach to facilitate engagement in sun protection by harnessing the influence of relationships among spouses and cohabiting partners. A couples-focused intervention may hold promise as a means to improve sun protection behaviors beyond interventions focused solely on individuals by leveraging the concern, collaboration, and support among intimate partners and addressing relationship-based barriers to sun protection.

Trial Registration: ClinicalTrials.gov NCT04549675; https://clinicaltrials.gov/ct2/show/NCT04549675

(J Med Internet Res 2020;22(9):e18037) doi:10.2196/18037

KEYWORDS

sun protection; behavior intervention; online interventions; couples; skin cancer prevention; mobile phone

Introduction

Skin cancer is the most common cancer in the United States. An estimated 96,480 cases of invasive melanoma and 5.4 million cases of nonmelanoma skin cancer were diagnosed in 2019 [1] Melanoma is the fifth most common malignancy in both men and women [2]. The rate of new melanoma cases has been rising by 1.5% on average each year over the last 10 years [2]. The incidence and mortality rates of non-melanoma squamous cell skin cancer are also increasing. The number of deaths caused by squamous cell skin cancer may soon be comparable to melanoma-related deaths. The rising number and per person costs of treatment for skin cancer has increased the average national annual treatment costs of skin cancer, estimated at US $8.2 billion per year [3]. On the basis of these facts, the United States Surgeon General's Call to Action to Prevent Skin Cancer [4] emphasized that skin cancer is a serious public health concern and suggested heightened skin cancer prevention efforts, including research, surveillance, and evaluation.

The primary risk factor for skin cancer is excess exposure to UV light, and the majority of skin cancers could be prevented if people consistently engaged in sun protection [5-7], The American Cancer Society [1] and the Skin Cancer Foundation [8] recommend minimizing exposure between daily peak hours for UV exposure, using sunscreen with a sun protection factor (SPF) of 30 or higher regularly and wearing protective clothing. Engagement in these recommendations is low. Up to 72% of US population do not use sunscreen regularly, wear protective clothing, or avoid the sun while outdoors [9,10]. Many studies have evaluated individual factors that contribute to sun protection behaviors, including demographic variables, objective risk factors, and attitudes and beliefs. For example, fewer perceived benefits of sun protection, more barriers to sun protection, and lower self-efficacy for using sun protection predict less sun protection [11-14].

The majority of sun protection interventions also focus on individuals. The potential role of the marital relationship as a motivator for sun protection is a less-studied, yet important, factor. Couples live together and typically engage in activities together. Thus, they share situations where UV exposure occurs (eg, sports events, vacations to sunny places), share sun protection equipment (eg, sunscreen bottles), and share environmental support for sun protection habits (eg, a car where sunglasses are stored). Overall, the high correlation between partners' sun protection practices (r=0.5-0.6) indicates significant couple similarity with regard to sun protection [15]. In terms of marital relationship influences, couples who discuss sun protection and endorse its benefits for the other partner and their relationship are more likely to engage in sun protection [15]. The marital relationship is an important influence on sun protection, and harnessing constructive marital influences offers a promising method to improve sun protection. Although no studies have evaluated the mechanisms of marital influence on sun protection, some have examined general family influence in persons with a family history of melanoma. These studies suggest that greater family support for sun protection is associated with higher levels of sun protection, and that communication about skin cancer occurs within families, particularly between parents and their minor children [16,17]. Additionally, family-focused behavioral interventions have shown efficacy in promoting health-related behaviors, including physical activity and diet as well as sun protection habits [18-20].

When considering how marital relationships may influence health behavior, Lewis et al [21] proposed an integrative framework based on an interdependence theory and communal coping framework to explain how couples' interactions may influence engagement in risk-reducing health behavior. This framework proposes that a strong interdependence in long-term, successful close relationships (ie, partners' influence on one another's behaviors and outcomes) transforms their motivations from doing what is in their self-interest (self-centered) to doing what is in the best interest of their relationship (relationship-centered). The transformation from self- to relationship-centered motivation occurs when partners ascribe health threats and subsequent health changes as having meaning for the relationship and/or their spouse. The model by Lewis et al [21-24] proposed 4 contributors to behavioral change: (1) predisposing factors of the couple (eg, individual perceptions of the health threat), (2) how much partners agree that health changes should be made together, (3) partners' commitment to the relationship, and (4) demographic factors. When relationship-centered motivation develops, communal coping begins. Communal coping efforts consist of joint decision making (eg, discussing the change) and planning how to make the change. Communal coping efforts lead to engagement in health behavior change for both partners [21-24]. In our prior work [15], we found high couple concordance with sun protection practices (r=0.5) and support for the interdependence and communal coping framework. Couples who reported that they discussed sun protection and endorsed its benefits for their partner and their relationship were more likely to engage in sun protection. Taken together, this suggests that harnessing constructive relationship influences via behavioral interventions may be a promising method for improving sun protection.

In a prior study, we developed and tested a couple-focused print and telephone counseling intervention called Sun Safe Partners [25]. Content was guided by the interdependence and communal coping framework. It included the provision of mailed small media materials, a couple-focused telephone counseling call, and a mailed summary letter. Results from a small, nonrandomized trial showed that couples' sun protection behaviors significantly increased after the intervention. We also observed increases in attitudes about the importance of one's own engagement in sun protection for the partner, relationship, and partner-centered motivations to engage in better sun protection [25]. However, intervention uptake was low, and implementation was challenging for our enrollees; it was difficult to schedule couples for the 1-hour phone call, deliver the content, and create implementation plans to improve sun protection for both partners.

To address these challenges, we created Sun Safe Partners Online and utilized a social media recruitment strategy rather than a web-based panel strategy. The web-based intervention allowed couples to work through the content at convenient times and at their own pace without the need for an interventionist.

In addition to standard individual-focused behavior change strategies such as goal setting and planning better sun protection, Sun Safe Partners Online content targeted couple-level influences by (1) raising awareness of the partner's skin cancer risk, (2) identifying benefits of improving sun protection for the partner and relationship, (3) helping partners learn and practice constructive communication to foster one another's sun protection, (4) identifying ways the partner can assist in working on sun protection behavioral goals, and (5) providing home assignments to help the couple discuss sun protection and ways to support each other's goals. Furthermore, content was added to address the risks of sun exposure to children, assess a child's risk factors, and set sun protection goals for the child for couples who have children in the home. We chose a social media advertisement recruitment strategy to examine whether this strategy resulted in better uptake than our prior work [25].

In this study, we report on the development, feasibility, and pilot testing of Sun Safe Partners Online. In a pilot and randomized feasibility trial, we compared Sun Safe Partners Online with a Generic Sun Safety Information-Only Online condition. The study had 2 aims. The first aim was to evaluate the feasibility and acceptability of Sun Safe Partners Online as compared with the Generic Online intervention. Feasibility was measured as enrollment, retention, and intervention use. Acceptability was assessed by a self-report evaluation of both interventions. We also compared our social media recruitment approach to the web-based panel approach utilized in our previous study [25]. The second aim was to assess the impact of Sun Safe Partners Online on the primary outcomes of sun protection and sun exposure and our intervention processes, which were individual and relationship attitudes and practices about sun protection. A 1-month, postbaseline follow-up survey was administered to examine the short-term impact of the intervention.

Methods

Development of Sun Safe Partners Online

Over a 10-month period, we worked with ITX Corporation to develop an interactive, online-mobile responsive (ie, can be accessed on a smartphone as well as a desk or laptop computer) web-based intervention. The web-based intervention focused on both the individual and the relationship with content divided into *My Stuff* (individual content) and *Our Stuff* (couple content). As shown in Table 1, we addressed key individual knowledge, attitudes, and behavior change constructs for sun protection as well as relationship content. Content from the original print and telephone intervention was used. We focused on individual factors drawn from Jackson and Aiken's psychosocial model of sun protection [13]. Sun Safe Partners Online also targeted proven behavior change techniques, including goal setting, action planning, and reviewing behavioral goals.

Table 1. Individual- and couple-focused objectives, targeted constructs, and tasks for Sun Safe Partners Online.

Objectives, Targeted constructs	Key tasks in Sun Safe Partners
Improve attitudes and skills for better sun protection	
• Personal risk for skin cancer • Sun protection benefits • Sun protection barriers • Improve confidence in sun protection practices • Action planning and goal setting	• Increase awareness of personal risk factors for skin cancer • Provide information about recommended sun protection • Assess current sun protection behaviors • Improve awareness of benefits of sun protection • Assess and address personal barriers to sun protection practices • Provide education about sunscreen application, sunscreen, sunglasses, minimizing exposure, unintentional sun exposure, and the dangers of tanning • Set sun protection behavioral goals, develop plans to address barriers to change
Build relationship focus and support for sun protection	
• Promote awareness of the benefits of sun protection for partner and relationship • Promote acceptance of partner support and influence	• Increase awareness of how the marital relationship can foster better health practices • Identify the benefits to the partner and the relationship for engaging in better sun protection • Increase awareness of partner's skin cancer risk (phenotype and current sun protection) • Increase willingness to accept influence from one's partner
• Promote supportive relationship behaviors regarding sun protection and including partner in goals	• Identify desired support for sun protection from the partner • Understand how to provide constructive support to one's partner • Build the ability to give and accept partner's influence for sun protection • Increase understanding of constructive communication to foster better sun protection habits • Include partner support for sun protection goals in goal-setting exercise

The program allowed couples to access the material while at home, but we did not convey the expectation that they would log in together and view the material at the same time. Rather, each partner had their own link and password to open the

intervention. In the Our Stuff modules, we designed activities that provided information about the other partner (eg, feedback about the other partner's skin cancer risk factors) as well as home assignments that ask couples to discuss topics covered in the modules and engage in setting joint goals and support each other's behavior change goals. Some of the activities (eg, completing quizzes) would be difficult to complete at the same time.

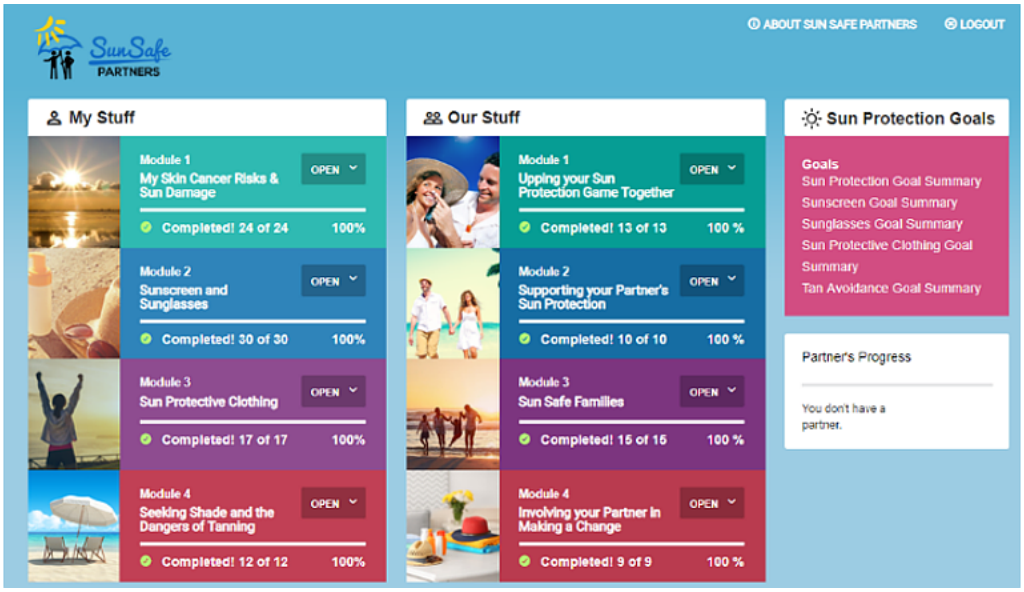
We developed 4 modules. Each *My Stuff* module included information displayed in colorful and engaging ways, tailored self-assessments (eg, Fitzpatrick skin type risk calculation, current sun protection, and sunscreen barriers), individual feedback (eg, participant's Fitzpatrick skin type), goal-setting exercises (eg, select a goal, identify barriers, and develop strategies to address barriers), and downloadable PDF files (eg, daily sun exposure diary). In *Our Stuff* modules, we created novel approaches to build relationship support. Each module contained basic information (eg, benefits of improving your health for your partner and relationship, the importance of staying healthy for your spouse and relationship, and shared environments for sun protection), personalized assessments (eg, my partner's risk factors for skin cancer, what my partner can do to help my sun protection, and how my partner can help with my sun protection goal), and couples home assignments (eg, share your sun protection goal with your partner). *Our Stuff* module 2 contained an animated video illustrating couples' communication about sun protection and partner assistance in completing a sun protection goal. Sun Safe Partners Online included a separate sun protection goal summary module, where participants could review and update their goals for sun protection, sunscreen, sunglasses, sun-protective clothing, and tanning avoidance. Sun Safe Partner's navigation page included a partner progress area, where the partner's progress was displayed. Participants could *nudge* their partner to log into the website or complete content on the landing page. Table 2 contains a summary of the content and assignments for each module, and the landing page for Sun Safe Partners Online is shown in Figure 1. After the initial development, we sent Sun Safe Partners Online to 6 couples to review and provide input and comments on navigation and content. Their feedback was incorporated into the intervention. A key change was made to unlock modules. Initially, the team planned to unlock the 4 modules weekly. However, owing to participant feedback, all modules were unlocked so that participants could complete the modules at their own pace.

Table 2. Information on content of the Sun Safe Partners Online modules.

Modules	My Stuff	Our Stuff
1	My skin cancer risks and sun damage: • Rationale for making changes as a couple • Basic information about skin cancer and sun damage • Assessment of skin cancer risk • Additional risks (sunburn history, tanning) • Ways to protect yourself from the sun • Assessment of current sun protection practices • Set and plan a sun protection goal	Upping your sun protection game together: • Importance of spouse support for health behavior changes • Health behavior changes made in the past that benefitted the partner and/or relationship • Choose relationship benefits for improving sun protection • Select way that partner can help with protection goal • Homework: share sun protection goal with partner or discuss relationship benefits of better sun protection
2	Sunscreen and sunglasses: • Homework review • Sunscreen recommendations, UV-A and UV-B, what is sun protection factor, chemical versus physical sunscreens • Set a sunscreen goal • Sunglasses: Ask the expert, barriers to wearing sunglasses • Set a sunglasses goal	Supporting your partner's sun protection: • Understanding your partner's skin cancer risk and current sun protection • How to support your partner improving his/her sun protection • Homework: Discuss skin cancer risk factors and sun protection behaviors that partners have in common/do not have in common or share your sun protection goal and make a plan about how you can help one another
3	Sun protective clothing: • Homework review • Recommended types of clothing • Unintentional sun exposure • Barriers to wearing clothing and hats • Set a sun protective clothing goal	Sun safe families (for couples who have children in the home): • Phenotypic risk assessment of child who has worst sun protection • Why children are at increased risk/guidelines • Sun safety in the home and outside the home • Assess child sun protection behaviors • Assess parent barriers for child sun protection • Setting a goal for child sun protection
4	Seeking shade and the dangers of tanning: • Homework review • Dangers of tanning • How to protect your skin by seeking shade • Rating current sun protection behaviors • Set a shade or tanning goal	Involving your partner in making a change: • Review of risk factors, sunscreen, sun glasses, protective clothing, and sun avoidance recommendations and how to involve your partner in the changes • List benefits to partner, benefits to relationship, and what partner can do to help you make the change
Goal setting	Goals from each section are imported into this section. Participants can view, add, and/or modify goals	N/A[a]

[a]N/A: not applicable.

Figure 1. Sun Safe Partners' landing page.



For the trial, consented participants were emailed a link to Sun Safe Partners Online. On the landing page, participants were instructed to complete home assignments before proceeding to the next module. As noted above, participants were instructed to move to the next module once the home assignments were completed.

Generic Online Intervention

The team reviewed publicly online available skin cancer and sun protection information and selected the following 4 links to send to participants as the Generic Online comparison condition: (1) The Skin Cancer Foundation's Skin Cancer Prevention guidelines [8], (2) the Centers for Disease Control and Prevention's information on risk factors for skin cancer [7], (3) the American Academy of Dermatology's information on how to select an appropriate sunscreen [26], and (4) The Skin Cancer Foundation's information on sun protective clothing [27]. Participants were emailed a link to one of the resources each week for 4 weeks.

Participants and Procedures

Facebook advertisements were created with Oxford Communications, an advertising company, to recruit couples for this study. The eligibility criteria were (1) both partners aged 18 to 75 years, (2) married or cohabiting with a significant other for at least one year, (3) partner #1 must be willing to provide contact information for partner #2, the (4) panel member (partner #1) and spouse responded with never, rarely, sometimes, or often to the question, "When outdoors in warm weather, how often do you protect your skin (by staying in the shade or covering your body with protective clothes or 30+ SPF sunscreen)?" (those who responded with *always* were excluded), (5) both partners had not been diagnosed with any type of skin cancer, and (6) both partners had an email account, internet access, and phone service (cell or landline).

Enrollment procedures were as follows: first, partner #1, the person who viewed the Facebook advertisement and indicated an interest in participating, clicked on the advertisement that took partner #1 to the eligibility survey. This survey included a consent to answer screening questions. If partner #1 was eligible, this person provided contact information for their cohabiting significant other (partner #2) and himself or herself. Next, Partner #2 was emailed a link to the eligibility survey. If both partners were eligible, then a member of the study team called the couple and spoke with both partners to confirm study eligibility. The team member confirmed study eligibility and sent an electronic link to the web-based consent and survey to eligible couples. Participants followed the link to acknowledge reading the web-based consent document before proceeding to the survey.

After both partners consented and completed the survey, couples were randomly assigned to either the Sun Safe Partners Online or the Generic Online condition. The cancer center's biostatistician created the randomization scheme Individual assignments were stored in a locked Excel file that could only be accessed by the study's project coordinator and accessed sequentially according to completion of the baseline assessment. Couples randomized to Sun Safe Partners Online were registered on the website and were provided a unique username and password. Participants were instructed to work at their own pace, but asked to do home assignments before logging into the next module. Home assignments were exercises completed with the other partner. Couples randomized to the Generic Online intervention were emailed the initial link, with 1 of 3 additional email links sent every 5 days. Participants were enrolled from May to August 2019, with participants recruited using 4 waves of advertisements. Follow-up surveys were completed between May and November 2019.

At baseline and the 1-month follow-up, participants completed surveys assessing sun protection, sun exposure, sun protection intentions, sun protection benefits, sunscreen and clothing barriers, and self-efficacy for sun protection as well as relationship benefits, motivation, and support. At the 1-month follow-up, a treatment acceptability measure was completed. Time spent in modules was downloaded from the Sun Safe Partners Online website. Participants were paid US $25 for the baseline and US $25 for the follow-up survey.

Measures: Primary Outcomes

Sun Protection

The Sun Habits survey [28] asked participants to rate their frequency of 5 sun protection behaviors (sunscreen, hat, shirt with sleeves, long pants, and sunglasses) on warm sunny days (1=never to 5=always). Studies evaluating the validity of self-reported sun protection with weekly electronic diaries [29] and observational assessments of sun protection have reported good correspondence [30,31]. Alpha reliabilities ranged from .52 to .64.

Sun Exposure

The Sun Habits survey [28] asked participants to rate the duration of outdoor time during peak hours on weekends and weekdays over the past summer months, 1=30 min or less, and 8=more than 6 hours. Self-report measures of time outdoors have shown satisfactory agreement with observational and dosimeter methods [29].

Measures: Intervention Processes

Individual Attitudes

Three items assessed the perceived risk of skin cancer (sample item: "If I don't protect myself from the sun, I would feel vulnerable to getting skin cancer in my lifetime") [15,32]. Items were rated on Likert-type response scales, ranging from 1 (strongly disagree) to 5 (strongly agree); alphas ranged from .91 to .92. Nine items assessed sun protection benefits (sample item: "Regularly wearing sunscreen when in the sun would reduce my chances of getting skin cancer"); alphas ranged from .85 to .87 [13,33]. Another 9 items measured sunscreen barriers (sample item: "For me, using sunscreen when I am outside on a warm sunny day is not part of my daily routine"); alphas ranged from .78 to .87 [13,34], both on 1 (strongly disagree) to 5 (strongly agree) Likert-type scales. Barriers to wearing sun-protective clothing [11] were measured with 7 Likert-type items. Sample item: "For me, wearing sun protective clothing when I am outside on a warm sunny day interferes with my work or leisure activities," 1 (strongly disagree) to 5 (strongly agree); alphas ranged from .82 to .95. Finally, self-efficacy for sun protection [13,33] was assessed with 9 items on confidence in performing sun protection behaviors. Sample item: "Are you confident that you can use sunscreen on every part of your body that is not covered by clothing?" rated from 1 (not at all confident) to 5 (very confident); alpha was .83 at both time points.

Relationship Attitudes

Twelve items measured the relationship benefits of sun protection for one's partner and relationship. Sample item: "I can think of reasons it would be beneficial for my relationship if I engage in sun protection," 1 (strongly disagree) to 5 (strongly agree); alphas ranged from .93 to .94 [15]. Five items assessed relationship motivations, that is, the degree to which partners perceive it is important to engage in sun protection because it is important to the other partner. Sample item: "I wear sunglasses when I go outside because it is important to my spouse that I do so," 1 (not at all true) to 5 (very true); alphas ranged from .80 to .87 [15]. Participants rated whether they engaged in 10 support behaviors for sun protection in the past month. Sample item: "Encouraged my spouse to apply sunscreen"; alphas ranged from .73 to .81 [25]. One item assessed the degree to which participants received support for sun protection from their spouse who supported their sun protection. Sample item: "How supportive is your partner of your sun protection practices?" (1=not at all supportive to 5=very supportive) [25].

Demographics

Age, sex, education, season of year enrolled, state residing in during childhood, relationship length, and phenotypic risk were measured at baseline. Phenotypic risk was measured using the Brief Cancer Risk Assessment scale [35]. Eight items assessed risk factors for skin cancer (sample item: "What is the color of your non-sun exposed skin?")

Intervention Acceptability, Satisfaction, and Use

At the 1-month follow-up, participants in both conditions completed a 9-item scale about the intervention they received. Sample items: "How helpful were the materials?" 1 (not at all helpful) to 7 (extremely helpful); "I learned something new from the materials/Sun Safe Partners website," "The information was easy to understand," and "I feel the materials were prepared with me and my partner in mind," 1 (strongly disagree) to 7 (strongly agree). Participants also rated how much of the materials they viewed. Sun Safe Partners: "How much of the Sun Safe Partners website did you review?"; Generic Online: "How much of the materials did you review?" 1 (just the overview) to 7 (viewed it many times) [25].

Sun Safe Partner Online participants completed a measure of ease of navigation, evaluation of content and features, and overall satisfaction. Twelve items assessed ease of navigation (sample items: "The Sun Safe Partners website was easy to use" and "The Sun Safe Partners website was user friendly"; rated on a 7-point Likert scale, 1=strongly disagree to 7=strongly agree). Eight items assessed features of the Sun Safe Partners program. Sample items: "What did you think about the goal-setting features throughout the program?" and "What did you think of the homework discussions with your partner?" (1=not at all helpful to 7=extremely helpful). Four items evaluated satisfaction. Sample item: "I am satisfied with the Sun Safe Partners website," 1 (strongly disagree) to 7 (strongly agree). The Sun Safe Partners Online website tracked logins and time in modules (eg, both partners individually in each couple).

Analytic Plan

For aim 1, in addition to basic descriptive information (eg, acceptance and survey completion rates), we compared the 2 study aims with regard to treatment evaluation. Both members of the couple participated in the study, hence the data were not independent. To handle this nonindependence, we used multilevel modeling treating dyad as the upper-level unit to compute tests of the intervention effect (Sun Safe Partners vs Generic Online). For aim 2, we adopted the same approach but in addition to the treatment condition, the fixed effect model also controlled for the person's baseline score on the outcome. Note that because this is a pilot study, we report Cohen *d* effect

sizes that were computed based on the *t* values and degrees of freedom for the condition effect from the multilevel models. Given the small data set and the goals of the pilot study, missing data were not imputed.

Results

Participants

The sociodemographic characteristics of the sample are shown in Table 3. The sample was 47.3% male and 52.3% female (there were several same-sex couples). Most (79.6%) participants were non-Hispanic White (9.5% Asian, 5.4% Black, and 3.4% Hispanic White), 88% had at least a high school certificate, the average age was 39.5 years (range 24 to 69 years), and the median relationship duration was 12 years (range 3 to 43 years). Nearly all participants (93.2%) had major medical insurance. Regarding sun exposure, 54.7% had experienced 3 or more blistering sunburns in their lifetime, and 42% had engaged in indoor tanning at least once in the past.

Table 3. Descriptive information for the study sample.

Variables	Sun Safe Partners Online		Generic Online intervention	
	Males	Females	Males	Females
Age (years), mean (SD)	40.9 (9.0)	38.2 (7.6)	41.0 (9.2)	38.1 (8.7)
Race and ethnicity, n (%)				
Non-Hispanic White	27 (75.0)	31 (77.5)	27 (77.0)	31 (83.8)
Non-Hispanic Black	2 (5.6)	2 (5.0)	1 (2.9)	1 (2.7)
Hispanic White	4 (11.1)	1 (2.5)	1 (2.9)	1 (2.7)
Asian	3 (8.3)	4 (10.0)	5 (14.3)	4 (10.8)
Indigenous people	0 (0)	1 (2.5)	1 (2.9)	0 (0.0)
Other	0 (0.0)	1 (2.5)	0 (0.0)	0 (0.0)
Education, n (%)				
Less than high school	1 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)
High school	6 (16.7)	4 (10.0)	4 (11.4)	3 (8.1)
Some college	9 (25.0)	11 (27.5)	7 (20.0)	6 (16.2)
Bachelor’s degree	11 (30.5)	15 (37.5)	11 (31.4)	16 (43.2)
Graduate degree	9 (25.0)	10 (25.0)	13 (37.2)	12 (32.5)
Relationship length (years), mean (SD)	13.8 (7.6)	12.9 (7.8)	13.0 (8.0)	12.5 (7.7)
Insurance status (yes), n (%)	33 (91.7)	36 (90.0)	33 (97.1)	35 (94.6)
Childhood residence (sun exposure), n (%)				
Northern latitude	29 (80.6)	29 (72.5)	23 (65.7)	26 (70.3)
Southern United States	1 (2.8)	0 (0.0)	0 (0.0)	1 (2.7)
Hawaii or Tropics	6 (16.6)	9 (22.5)	11 (31.4)	8 (21.6)
Unknown	0 (0.0)	2 (5.0)	1 (2.9)	2 (5.4)
Phenotypic risks, n (%)				
Fair to very fair skin	23 (63.9)	31 (77.5)	20 (58.8)	24 (68.6)
Blonde or red hair	8 (22.2)	10 (25.0)	4 (11.8)	9 (25.8)
History of 6 or more sunburns	6 (16.7)	12 (30.0)	4 (11.7)	10 (28.6)
More than 10 moles	2 (5.6)	2 (5.0)	0 (0.0)	1 (2.9)
Many freckles	7 (19.4)	12 (30.0)	3 (8.8)	12 (34.3)
Burn easily	17 (47.2)	28 (71.8)	19 (55.9)	22 (62.9)
Ability to tan none or light	11 (30.6)	20 (50.0)	12 (35.3)	18 (51.4)

Aim 1: Feasibility and Acceptability

Recruitment and Retention

The recruitment and retention of participants is shown in the Consolidated Standards of Reporting Trials (CONSORT) diagram in Figure 2. From Facebook advertisements, 572 eligible partner #1s were identified, and links were sent by the project coordinator to partner #2 to determine eligibility. Of these 572 partners #1s, 398 partner #2s (69.5%) did not complete the eligibility screener or were ineligible, and 174 partner #2s (30.5%) were eligible. These 174 couples were contacted by phone, and 77 couples (44%) were reached and confirmed to be an actual couple. Of these 77 couples, 74 couples completed baseline surveys and were randomized to the Sun Safe Partners Online or the Generic Online intervention (n=36 couples/74 participants assigned to the Sun Safe Partners condition and 38 couples/74 participants assigned to the Generic Online condition). This yielded an acceptance rate of 42.5% (74/174 couples) of eligible couples.

Figure 2. CONSORT diagram. BL: baseline; CONSORT: Consolidated Standards of Reporting Trials; P1: Patient #1; P2: Patient #2.

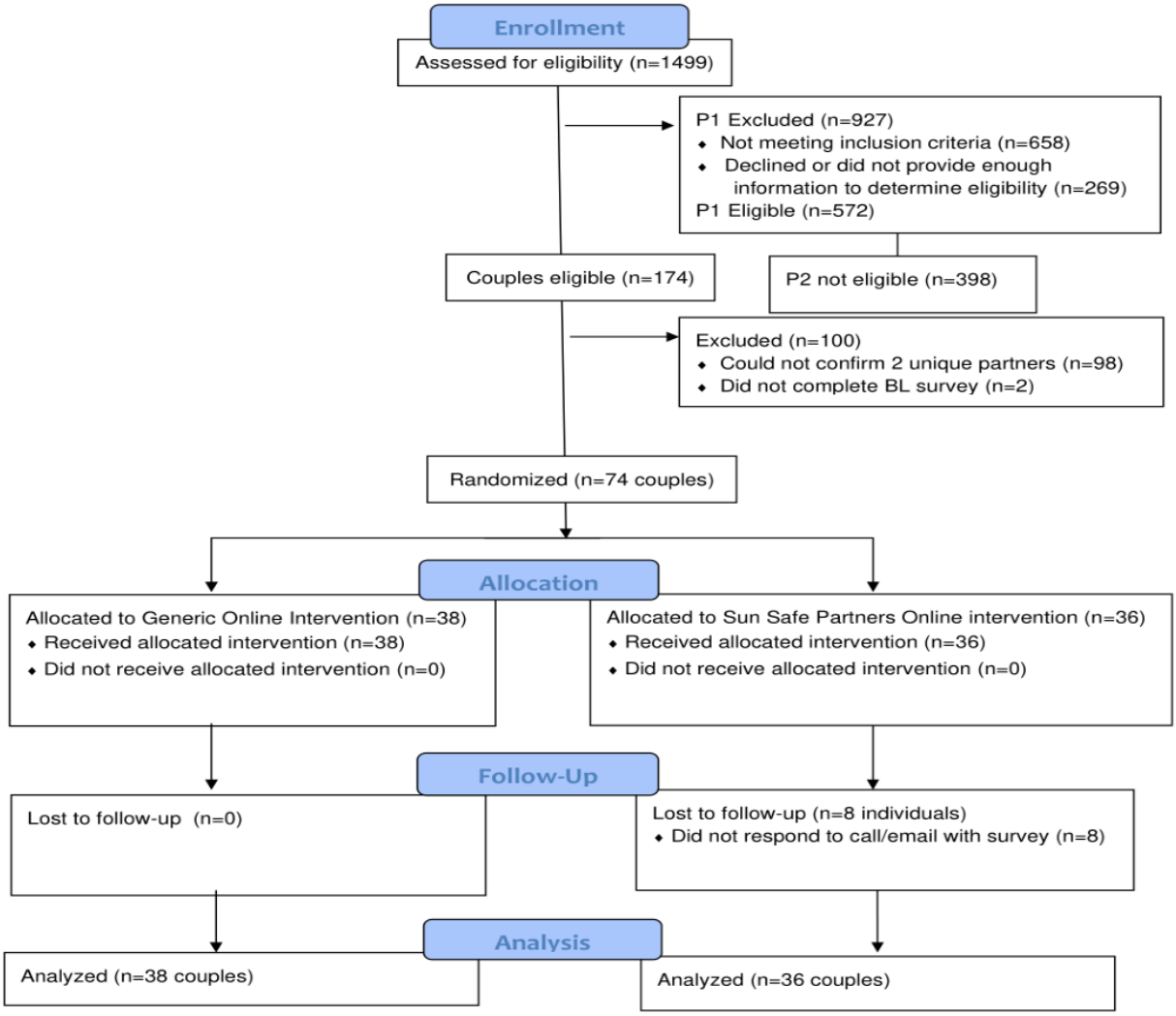


Figure 2 shows the retention rate. The follow-up survey completion rate was 93.3%. However, the return rate was higher in the Generic Online condition (100%) than in the Sun Safe Partners Online condition (87.2%). Comparisons between participants who completed the follow-up with participants who did not complete the follow-up with regard to demographic characteristics, baseline individual and relationship attitudes, sun protection, and sun exposure, did not show significant differences in any of the variables.

Sun Safe Partners Online Engagement and Evaluation

Of the 78 participants randomized to Sun Safe Partners Online, a review of data collected from the website indicated that 72 (92.3%) logged into the first module. The percentage of participants who opened the first page of each module are as follows: *My Stuff*: module 1 (91.5%), module 2 (91.5%), module 3 (83.5%), and module 4 (78.2%); *Our Stuff*: module 1 (84.6%), module 2 (76.9%), module 3 (76.9%), and module 4 (75.6%). The total time spent in Sun Safe Partners Online ranged from 1.5 min to 189.2 min (median 67.5 min, mean 69.3 min, SD 47.7 min). Of the 72 participants randomized to the Generic Online intervention, 60 participants (83.3%) reported reviewing all materials at least once, and 6 reported that they glanced over it or viewed a few sections (7.3%).

Of the participants randomized to the Sun Safe Partners Online, who completed the follow-up survey, 95.5% reported completing homework discussions with their partner. Among participants

reporting having a discussion with their partner, the most common home assignment topics that were discussed were: "Sharing my sun protection goal with my partner and how he or she can support me" (70.5%), "Making a plan about how you can help each other work on your sun protection goal"(62.8%), "Working together on building a sun safe home"(61.5%), "Discussing the benefits to your relationship, partner, and family of adopting better sun protection" (60.3%), and "Discussing skin cancer risk factors and sun protection habits that my partner and I have in common and do not have in common" (60.3%). Intervention acceptability is shown in Table 4.

Both interventions were evaluated positively. However, Sun Safe Partners Online had significantly higher ratings than the Generic Online intervention on helpfulness, learning something new, being valuable, being interesting, and being prepared with the couple in mind, making it easier to talk to the partner about better sun protection, and fostering an understanding of why it is helpful for the relationship and spouse to engage in better sun protection. Sun Safe Partners Online was also rated highly on ability to navigate it and its content. Videos, interactive quizzes, and homework assignments were rated positively. Positive aspects noted in open-ended questions were as follows: "interesting videos," "like connection with my partner," and "examples of how to talk to my husband about sun protection."

Table 4. Feasibility and acceptability of Sun Safe Partners Online and Generic Online intervention.

Intervention acceptability	Sun Safe Partners (n=67), mean (SD)	Generic Online (n=72), mean (SD)	*t* test (*df*)	*P* value
General characteristics				
Was helpful	6.10 (1.08)	5.60 (1.27)	2.01 (70)	.048
Contained valid information	6.55 (0.70)	6.39 (0.99)	1.06 (70)	.29
Learned something new	6.34 (1.04)	5.69 (1.33)	2.64 (70)	.01
Information was valuable to me	6.31 (0.93)	5.57 (1.37)	2.94 (68)	.005
Information was interesting	6.24 (1.00)	5.51 (1.34)	2.87 (68)	.005
Length of time to review it was sufficient	5.57 (1.56)	5.40 (1.50)	0.44 (70)	.66
Prepared with me and my partner in mind	6.30 (1.10)	5.32 (1.51)	3.80 (69)	<.001
Made it easier to talk to my partner about better sun protection	6.27 (1.25)	5.43 (1.42)	3.16 (70)	.002
Helped me understand why it was helpful for our relationship and why my spouse has to protect our skin from the sun	6.43 (1.06)	5.52 (1.33)	3.77 (70)	<.001
Sun Safe Partners Online navigation				
Easy to use	6.58 (0.68)	N/A[a]	N/A	N/A
Simple to use	6.56 (0.77)	N/A	N/A	N/A
User friendly	6.53 (0.75)	N/A	N/A	N/A
Required fewest steps possible to accomplish what I wanted to do	5.92 (1.33)	N/A	N/A	N/A
Flexible	6.29 (0.98)	N/A	N/A	N/A
Using it was effortless	6.09 (1.12)	N/A	N/A	N/A
Learned to use it quickly	6.52 (0.83)	N/A	N/A	N/A
Easy to remember how to use it	6.57 (0.72)	N/A	N/A	N/A
Easy to learn to use it	6.54 (0.84)	N/A	N/A	N/A
Quickly became skillful	6.43 (0.93)	N/A	N/A	N/A
Sun Safe Partners Online satisfaction				
Satisfied with it	6.45 (0.82)	N/A	N/A	N/A
Would recommend to a friend	6.22 (1.18)	N/A	N/A	N/A
Works the way I want it to	6.28 (0.93)	N/A	N/A	N/A
Feel the need to have it	5.70 (1.43)	N/A	N/A	N/A
Features of Sun Safe Partners				
Sun protection content	6.31 (1.00)	N/A	N/A	N/A
Videos	5.73 (1.58)	N/A	N/A	N/A
Home assignments for couple	5.93 (1.25)	N/A	N/A	N/A
Quizzes	6.30 (1.13)	N/A	N/A	N/A
Links between partner answers	6.17 (1.02)	N/A	N/A	N/A
Reminder to login	6.01 (1.28)	N/A	N/A	N/A
Goal-setting feature	5.76 (1.46)	N/A	N/A	N/A

[a]N/A: not applicable.

Aim 2: Impact of Sun Safe Partners Online Versus Generic Online Intervention on Outcomes

The descriptive statistics for baseline and follow-up as a function of condition as well as *t* tests testing the condition effect and Cohen *d* estimating the condition effect are presented in Table 5. As the levels of sun exposure on weekdays were quite low (33% reported 30 min or less on weekdays vs only 3% reported 30 min or less on weekends), we focused on weekend sun exposure as the sun exposure outcome.

Table 5. Comparisons of the Sun Safe Partners Online with the Generic Online intervention on outcomes and relationship and individual factors.

Outcomes[a]	Sun Safe Partner Online		Generic Online		Cohen *d*
	Baseline, mean (SD)	Follow-up, mean (SD)	Baseline, mean (SD)	Follow-up, mean (SD)	
Primary outcomes					
Sun Protection Behaviors	2.84 (0.67)	3.19 (0.73)	2.79 (0.60)	2.96 (0.59)	0.36
Weekend sun exposure	3.58 (1.59)	2.84 (1.33)	3.54 (1.58)	2.93 (1.21)	0.08
Intervention processes					
Relationship factors					
Relationship benefits	4.39 (0.66)	4.66 (0.52)	4.01 (0.71)	4.26 (0.65)	0.68
Relationship motivation	1.93 (0.72)	2.54 (0.96)	1.73 (0.72)	2.17 (0.87)	0.29
Support provided	4.33 (0.94)	4.59 (0.67)	4.06 (1.14)	4.58 (0.78)	0.18
Support received	4.45 (0.84)	4.47 (0.87)	3.94 (1.17)	4.31 (1.03)	0.18
Individual factors					
Sun protection intentions	4.35 (1.13)	5.35 (1.09)	4.00 (1.03)	4.77 (1.04)	0.48
Perceived risk	4.27 (0.94)	4.47 (0.94)	4.01 (0.96)	4.30 (0.87)	0.01
Sun protection benefits	4.37 (0.74)	4.76 (0.53)	4.17 (0.69)	4.49 (0.55)	0.45
Sunscreen barriers	2.74 (0.73)	2.39 (0.87)	2.90 (0.69)	2.58 (0.83)	0.14
Clothing barriers	3.41 (1.12)	3.07 (1.24)	3.74 (0.97)	3.28 (1.15)	0.01
Sun protection efficacy	3.29 (0.88)	3.64 (0.83)	3.14 (0.87)	3.44 (0.83)	0.16

[a]At baseline, Sun Safe Partners Online (n=76) and Generic Online (n=72). At follow-up, Sun Safe Partners Online (n=72). The *t* tests for follow-up differences as a function of condition were computed using multilevel modeling controlling for baseline score. Degrees of freedom for these tests ranged between 64 and 72 across the variables.

Sun Safe Partners Online showed small-to-moderate size effects on sun protection behaviors and sun exposure on weekends as compared with the Generic Online intervention. Sun Safe Partners Online also showed small-to-moderate size effects on relationship benefits and support provided to the partner. The small-to-moderate effect sizes for individual factors suggested that participants in Sun Safe Partners Online increased sun protection intentions and benefits.

Discussion

Engagement in Sun Safe Partners Online was high, with the vast majority of participants using it, and more than two-third of participants completing all modules. Online delivery may have allowed couples to engage in the program at convenient times and places, at their own pace, and when they were together.

Using Facebook as a recruitment strategy resulted in rapid recruitment. Higher acceptance rates were observed (43%) relative to our prior telephone and print trial (22.4%) [25], which used a Qualtrics online panel recruitment. Compared with a Qualtrics panel, Facebook users may also have a preference for online interventions, which could explain the higher acceptance rate. Facebook is also a networking platform that may have yielded participants who placed a higher premium on a relationship-based intervention. It would be informative to see if samples recruited through general internet advertising or offline means would be as engaged with the Sun Safe Partners Online intervention as the current sample derived from Facebook. Follow-up survey completion was also high, although the return rate was lower in the Sun Safe Partners Online condition than in the Generic Online condition. Perhaps the participants from Facebook expected to be engaged for a shorter period of time, and thus, some of them felt they had devoted enough time to the study when completing the Sun Safe Partner Online intervention. Still, follow-up in the Sun Safe Partners Online condition was still high (87%), reducing concerns that loss to follow-up would create substantial selection biases.

Compared with the Generic Online intervention, Sun Safe Partners Online showed small magnitude increases in sun protection behaviors and sun exposure on weekends. However, it should be noted that there were increases in sun protection and reductions in sun exposure in the Generic Online intervention. This was encouraging and suggests that the provision of basic sun protection and skin cancer risk information to participants may motivate some increase in sun protection behaviors. However, these changes must be interpreted with caution. Owing to a very small sample, we avoided significance testing, and the effect size may be imprecise, as noted in the published advice on interpreting pilot studies [36,37]. The results hold promise due to Sun Safe Partners Online's differential impact on relationship factors. Although both interventions showed increases in participants' ability to view sun protection from a relational perspective, there were moderate-sized increases in the Sun Safe Partners Online condition and small magnitude effects for Sun Safe Partners Online on relationship motivations and support for the spouse's sun protection. The results support the framework by

Lewis et al [38]. These findings are corroborated by the fact that participants reported supportive behaviors to help their partner adopt sun protection behaviors. In terms of individual factors, there were medium effect sizes in favor of Sun Safe Partners on sun protection benefits and intentions, but there were no differential effects on sunscreen barriers, barriers to wearing sun-protective clothing, and perceived risk. Overall, the pattern of findings implies that Sun Safe Partners Online had an impact on the relationship constructs as intended. However, a fully powered randomized trial is needed to provide inferential tests on the effectiveness of Sun Safe Partners Online [36,37].

A comparison with our previous noncontrolled clinical trial in which Sun Safe Partners involved a tailored counseling call as well as print materials [25] illustrates some important points about Sun Safe Partners Online. In our prior study, we reported a larger effect size for sun protection behaviors (Cohen *d*=1.29) at the 6-month follow-up than the effect size for sun protection behaviors in this trial (Cohen *d*=.36) at the 1-month follow-up. However, these 2 effect sizes are not directly comparable in the sense that the effect size from this study compares baseline to follow-up differences between the intervention and control groups, and the effect size from the previous study compared only differences between baseline and follow-up for the intervention group (ie, there was no control). We recomputed both effect sizes using data from only those individuals in the intervention group who completed both waves of data collection. On the basis of these data, we again found that the current web-based study produced a much smaller overtime effect size (Cohen *d*=.43 vs Cohen *d*=1.41). There are several possible explanations for this difference. First, the mean baseline score on sun protection behaviors for the study was considerably higher (mean 2.88, SD 0.70) than in the previous study (mean 2.47, SD 0.47). Thus, these differences may reflect sampling. A second potential explanation for this finding is that the prior intervention was more intensive, and because all individuals participated in the counseling call, the intervention *dose* was more consistent.

Although both interventions were positively evaluated, Sun Safe Partners Online was rated as more helpful, valuable, and interesting than the Generic Online intervention. Participants felt that it was prepared for the couple, it was viewed as promoting the ability to talk to one's spouse about better sun protection, and it was seen as fostering an understanding of why it is helpful for the relationship and spouse to engage in better sun protection. Participants felt Sun Safe Partners were easy to navigate, and their unique features were positively evaluated. These features may have increased partners' engagement with the program and collaboration on homework, and made it more likely to impact relationship factors than the Generic Online intervention.

These conclusions should be considered in light of the study limitations. Data collection spanned the late summer through the winter months. For participants who resided in southern climates, the follow-up occurred in a warm, sunny time frame; however, for participants who resided in nonsouthern climates, follow-up occurred in the early fall when UV levels decreased. Thus, sun protection behaviors may not have been as useful among those residing in nonsouthern climates. Second, nearly 80% of the participants were non-Hispanic Whites. Thus, the sample size was not as diverse as the general population. However, skin cancer is far more prevalent among non-Hispanic Whites, especially those with highly sun-sensitive skin, so the sample undoubtedly contained a large number of high-risk participants, the key target population. Finally, our participants may have been more motivated to improve sun protection behaviors than the general population because they volunteered for an intervention on this topic.

This study leveraged a novel approach to facilitating engagement in sun protection by harnessing the relationship between spouses and cohabiting partners. A couple-focused intervention may hold promise as a way to improve sun protection behaviors by leveraging the concern, collaboration, and support among intimate partners and addressing relationship-based barriers to sun protection. Participants felt Sun Safe Partners Online was valuable, and most participants completed all of the modules. On the basis of the outcome of our pilot study, a fully powered trial with a larger, more diverse sample and a longer follow-up time frame is warranted to evaluate the efficacy of Sun Safe Partners Online, which has the potential in its web-based format to be scaled up to a larger population of adults at risk for skin cancer.

Acknowledgments

This research was supported by Rutgers Cancer Institute of New Jersey Institutional funds. The sponsor was not involved in the project design, collection, analysis and interpretation of data, writing of the manuscript, or the decision to submit this paper for publication. This research was facilitated by ITX Incorporated and Oxford Incorporated. The authors would like to thank Mary Riley (MPhil) for input on recruitment methods and the participants in the study.

Conflicts of Interest

None declared.

This randomized study was only retrospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

References

1. Melanoma of the Skin. Cancer Statistics Center - American Cancer Society. 2018. URL: http://cancerstatisticscenter.cancer.org [accessed 2020-07-28]
2. Cancer Stat Facts: Melanoma of the Skin. Surveillance, Epidemiology, and End Results Program. 2019. URL: https://seer.cancer.gov/statfacts/html/melan.html [accessed 2020-07-28]
3. Guy GP, Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the US, 2002-2006 and 2007-2011. Am J Prev Med 2015 Feb;48(2):183-187 [FREE Full text] [doi: 10.1016/j.amepre.2014.08.036] [Medline: 25442229]
4. US Department of Health and Human Services, Office of the Surgeon General. The Surgeon General's Call to Action to Prevent Skin Cancer. Washington, DC: US Department of Health and Human Services; 2019.
5. van der Pols JC, Williams GM, Pandeya N, Logan V, Green AC. Prolonged prevention of squamous cell carcinoma of the skin by regular sunscreen use. Cancer Epidemiol Biomarkers Prev 2006 Dec;15(12):2546-2548 [FREE Full text] [doi: 10.1158/1055-9965.EPI-06-0352] [Medline: 17132769]
6. Green AC, Williams GM, Logan V, Strutton GM. Reduced melanoma after regular sunscreen use: randomized trial follow-up. J Clin Oncol 2011 Jan 20;29(3):257-263. [doi: 10.1200/JCO.2010.28.7078] [Medline: 21135266]
7. What Are the Risk Factors for Skin Cancer? Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/cancer/skin/basic_info/risk_factors.htm [accessed 2020-01-28]
8. Skin Cancer Prevention. Skin Cancer Foundation. 2020. URL: https://www.skincancer.org/skin-cancer-prevention/ [accessed 2020-08-12]
9. Fischer AH, Wang TS, Yenokyan G, Kang S, Chien AL. Sunburn and sun-protective behaviors among adults with and without previous nonmelanoma skin cancer (NMSC): a population-based study. J Am Acad Dermatol 2016 Aug;75(2):371-9.e5 [FREE Full text] [doi: 10.1016/j.jaad.2016.02.1236] [Medline: 27198078]
10. Buller DB, Cokkinides V, Hall HI, Hartman AM, Saraiya M, Miller E, et al. Prevalence of sunburn, sun protection, and indoor tanning behaviors among Americans: review from national surveys and case studies of 3 states. J Am Acad Dermatol 2011 Nov;65(5 Suppl 1):S114-S123. [doi: 10.1016/j.jaad.2011.05.033] [Medline: 22018060]
11. Bränström R, Kasparian NA, Chang Y, Affleck P, Tibben A, Aspinwall LG, et al. Predictors of sun protection behaviors and severe sunburn in an international online study. Cancer Epidemiol Biomarkers Prev 2010 Sep;19(9):2199-2210 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-0196] [Medline: 20643826]
12. Calderón TA, Bleakley A, Jordan AB, Lazovich D, Glanz K. Correlates of sun protection behaviors in racially and ethnically diverse US adults. Prev Med Rep 2019 Mar;13:346-353 [FREE Full text] [doi: 10.1016/j.pmedr.2018.12.006] [Medline: 30792951]
13. Jackson KM, Aiken LS. A psychosocial model of sun protection and sunbathing in young women: the impact of health beliefs, attitudes, norms, and self-efficacy for sun protection. Health Psychol 2000 Sep;19(5):469-478. [doi: 10.1037//0278-6133.19.5.469] [Medline: 11007155]
14. Weig EA, Tull R, Chung J, Brown-Joel ZO, Majee R, Ferguson NN. Assessing factors affecting sunscreen use and barriers to compliance: a cross-sectional survey-based study. J Dermatolog Treat 2020 Jun;31(4):403-405. [doi: 10.1080/09546634.2019.1587147] [Medline: 30889998]
15. Manne SL, Coups EJ, Kashy DA. Relationship factors and couples' engagement in sun protection. Health Educ Res 2016 Aug;31(4):542-554. [doi: 10.1093/her/cyw027] [Medline: 27247330]
16. Hay J, Ostroff J, Martin A, Serle N, Soma S, Mujumdar U, et al. Skin cancer risk discussions in melanoma-affected families. J Cancer Educ 2005;20(4):240-246. [doi: 10.1207/s15430154jce2004_13] [Medline: 16497137]
17. Loescher LJ, Crist JD, Siaki LA. Perceived intrafamily melanoma risk communication. Cancer Nurs 2009;32(3):203-210. [doi: 10.1097/NCC.0b013e31819ae11c] [Medline: 19295423]
18. Prestwich A, Conner MT, Lawton RJ, Ward JK, Ayres K, McEachan RR. Partner- and planning-based interventions to reduce fat consumption: randomized controlled trial. Br J Health Psychol 2014 Feb;19(1):132-148. [doi: 10.1111/bjhp.12047] [Medline: 23659492]
19. Voils CI, Coffman CJ, Yancy WS, Weinberger M, Jeffreys AS, Datta S, et al. A randomized controlled trial to evaluate the effectiveness of CouPLES: a spouse-assisted lifestyle change intervention to improve low-density lipoprotein cholesterol. Prev Med 2013 Jan;56(1):46-52. [doi: 10.1016/j.ypmed.2012.11.001] [Medline: 23146744]
20. Bowen DJ, Burke W, Hay JL, Meischke H, Harris JN. Effects of web-based intervention on risk reduction behaviors in melanoma survivors. J Cancer Surviv 2015 Jun;9(2):279-286 [FREE Full text] [doi: 10.1007/s11764-014-0412-0] [Medline: 25425205]
21. Butterfield RM, Lewis MA. Health-related social influence: a social ecological perspective on tactic use. J Soc Pers Relatsh 2016 Jun 30;19(4):505-526. [doi: 10.1177/0265407502019004050]
22. Lewis MA, Butterfield RM, Darbes LA, Johnston-Brooks C. The conceptualization and assessment of health-related social control. J Soc Pers Relatsh 2016 Jun 30;21(5):669-687. [doi: 10.1177/0265407504045893]
23. Lewis MA, Gladstone E, Schmal S, Darbes LA. Health-related social control and relationship interdependence among gay couples. Health Educ Res 2006 Aug;21(4):488-500. [doi: 10.1093/her/cyh075] [Medline: 16459342]

24. Glanz K, Lewis FM, Rimer BK. Health Behavior and Health Education. Med Sci Sports Exerc 1991;23(12):1404. [doi: 10.1249/00005768-199112000-00016]
25. Manne S, Day A, Coups EJ, Kashy D. Sun safe partners: a pilot and feasibility trial of a couple-focused intervention to improve sun protection practices. Prev Med Rep 2018 Dec;12:220-226 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.010] [Medline: 30370209]
26. How to Select a Sunscreen. American Academy of Dermatology. URL: https://www.aad.org/public/everyday-care/sun-protection/sunscreen/how-to-select-sunscreen [accessed 2020-01-22]
27. Sun-Protective Clothing: A Safe, Simple Way to Keep the Rays at Bay. The Skin Cancer Foundation. URL: http://www.skincancer.org/prevention/sun-protection/clothing [accessed 2020-01-22]
28. Glanz K, Yaroch AL, Dancel M, Saraiya M, Crane LA, Buller DB, et al. Measures of sun exposure and sun protection practices for behavioral and epidemiologic research. Arch Dermatol 2008 Feb;144(2):217-222. [doi: 10.1001/archdermatol.2007.46] [Medline: 18283179]
29. Hillhouse J, Turrisi R, Jaccard J, Robinson J. Accuracy of self-reported sun exposure and sun protection behavior. Prev Sci 2012 Oct;13(5):519-531 [FREE Full text] [doi: 10.1007/s11121-012-0278-1] [Medline: 22855253]
30. Glanz K, McCarty F, Nehl EJ, O'Riordan DL, Gies P, Bundy L, et al. Validity of self-reported sunscreen use by parents, children, and lifeguards. Am J Prev Med 2009 Jan;36(1):63-69 [FREE Full text] [doi: 10.1016/j.amepre.2008.09.012] [Medline: 18945582]
31. O'Riordan DL, Nehl E, Gies P, Bundy L, Burgess K, Davis E, et al. Validity of covering-up sun-protection habits: association of observations and self-report. J Am Acad Dermatol 2009 May;60(5):739-744 [FREE Full text] [doi: 10.1016/j.jaad.2008.12.015] [Medline: 19278750]
32. Janssen E, van Osch L, de Vries H, Lechner L. Measuring risk perceptions of skin cancer: reliability and validity of different operationalizations. Br J Health Psychol 2011 Feb;16(Pt 1):92-112. [doi: 10.1348/135910710X514120] [Medline: 21226786]
33. Manne S, Fasanella N, Connors J, Floyd B, Wang H, Lessin S. Sun protection and skin surveillance practices among relatives of patients with malignant melanoma: prevalence and predictors. Prev Med 2004 Jul;39(1):36-47. [doi: 10.1016/j.ypmed.2004.02.028] [Medline: 15207984]
34. Manne SL, Coups EJ, Jacobsen PB, Ming M, Heckman CJ, Lessin S. Sun protection and sunbathing practices among at-risk family members of patients with melanoma. BMC Public Health 2011 Feb 21;11:122 [FREE Full text] [doi: 10.1186/1471-2458-11-122] [Medline: 21338483]
35. Glanz K, Schoenfeld E, Weinstock MA, Layi G, Kidd J, Shigaki DM. Development and reliability of a brief skin cancer risk assessment tool. Cancer Detect Prev 2003;27(4):311-315. [doi: 10.1016/s0361-090x(03)00094-1] [Medline: 12893080]
36. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. Arch Gen Psychiatry 2006 May;63(5):484-489. [doi: 10.1001/archpsyc.63.5.484] [Medline: 16651505]
37. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
38. Lewis MA, McBride CM, Pollak KI, Puleo E, Butterfield RM, Emmons KM. Understanding health behavior change among couples: an interdependence and communal coping approach. Soc Sci Med 2006 Mar;62(6):1369-1380. [doi: 10.1016/j.socscimed.2005.08.006] [Medline: 16146666]

Abbreviations

SPF: sun protection factor

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by Y Wu, D Bowen; comments to author 25.03.20; revised version received 20.04.20; accepted 27.04.20; published 17.09.20.

Please cite as:
Manne S, Buller D, Devine K, Heckman C, Pagoto S, Frederick S, Mitarotondo A
Sun Safe Partners Online: Pilot Randomized Controlled Clinical Trial
J Med Internet Res 2020;22(9):e18037
URL: https://www.jmir.org/2020/9/e18037
doi:10.2196/18037
PMID:32673215

Original Paper

Effectiveness of a Web-Based SUpport PRogram (SUPR) for Hearing Aid Users Aged 50+: Two-Arm, Cluster Randomized Controlled Trial

Janine FJ Meijerink[1], MSc; Marieke Pronk[1], PhD; Birgit I Lissenberg-Witte[2], PhD; Vera Jansen[3], BSc; Sophia E Kramer[1], PhD

[1]Otolaryngology—Head and Neck Surgery, Ear and Hearing, Amsterdam Public Health research institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[3]Schoonenberg HoorSupport, Dordrecht, Netherlands

Corresponding Author:
Marieke Pronk, PhD
Otolaryngology—Head and Neck Surgery, Ear and Hearing
Amsterdam Public Health research institute
Amsterdam UMC, Vrije Universiteit Amsterdam
De Boelelaan 1117
Amsterdam
Netherlands
Phone: 1 +31 020 444 905
Email: m.pronk@amsterdamumc.nl

Abstract

Background: Hearing aid (HA) use is known to improve health outcomes for people with hearing loss. Despite that, HA use is suboptimal, and communication issues and hearing-related activity limitations and participation restrictions often remain. Web-based self-management communication programs may support people with hearing loss to effectively self-manage the impact of hearing loss in their daily lives.

Objective: The goal of the research is to examine the short- and long-term effects of a web-based self-management SUpport PRogram (SUPR) on communication strategy use (primary outcome) and a range of secondary outcomes for HA users aged 50 years and older.

Methods: Clients of 36 HA dispensing practices were randomized to SUPR (SUPR recipients; n=180 HA users) and 34 to care as usual (controls; n=163 HA users). SUPR recipients received a practical support booklet and online materials delivered via email over the course of their 6-month HA rehabilitation trajectory. They were encouraged to appoint a communication partner and were offered optional email contact with the HA dispensing practice. The online materials included 3 instruction videos on HA handling, 5 videos on communication strategies, and 3 testimonial videos. Care as usual included a HA fitting rehabilitation trajectory only. Measurements were carried out at baseline, immediately postintervention, 6 months postintervention, and 12 months postintervention. The primary outcome measure was self-reported use of communication strategies (3 subscales of the Communication Profile for the Hearing Impaired [CPHI]). Secondary outcome measures included self-reported personal adjustment to hearing loss (CPHI); use, satisfaction and benefit of HAs and SUPR (use questionnaire; International Outcome Inventory for Hearing Aids [IOI-HA], Alternative Interventions [IOI-AI]); recommendation of HA dispensing services; self-efficacy for HA handling (Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids [MARS-HA]); readiness to act on hearing loss (University of Rhode Island Change Assessment adapted for hearing loss [URICA-HL]); and hearing disability (Amsterdam Inventory for Auditory Disability and Handicap [AIADH]).

Results: Linear mixed model analyses (intention to treat) showed no significant differences between the SUPR and control group in the course of communication strategy use (CPHI). Immediately postintervention, SUPR recipients showed significantly higher self-efficacy for advanced HA handling than the controls, which was sustained at 12 months (MARS-HA; mean difference immediately postintervention: 5.3, 95% CI 0.3 to 10.4; *P*=.04). Also, SUPR recipients showed significantly greater HA satisfaction than controls immediately postintervention (IOI-HA; 0.3, 95% CI 0.09 to 0.5; *P*=.006), which was sustained at 12 months, and

significantly greater HA use than the controls immediately postintervention (IOI-HA; 0.3, 95% CI 0.02 to 0.5; *P*=.03), which was not sustained at 12 months.

Conclusions: This study provides ground to recommend adding SUPR to standard HA dispensing care, as long-term, modest improvements in HA outcomes were observed. Further research is needed to evaluate what adjustments to SUPR are needed to establish long-term effectiveness on outcomes in the psychosocial domain.

Trial Registration: ISRCTN77340339; http://www.isrctn.com/ISRCTN77340339

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2016-015012

(J Med Internet Res 2020;22(9):e17927) doi:10.2196/17927

KEYWORDS

hearing loss; hearing aids; auditory rehabilitation; self-management; communication programs; internet; hearing aid dispensing practice; randomized controlled trial

Introduction

Hearing loss is highly prevalent among older adults. Approximately one-third of adults aged 65 years or over have a disabling hearing loss [1]. Importantly, hearing loss is known to be associated with various negative health outcomes including falls [2], loneliness and depression [3-5], incident dementia [6,7], and even mortality [8,9]. The main clinical rehabilitation option for people with hearing loss is the provision of hearing aids (HAs) [10]. Although there is evidence that HA use can reverse negative effects [11-15], a substantial proportion of HA owners, estimated at 3% to 24%, never wear them [16-21]. Others still experience problems in their daily lives when wearing them [22]. Underlying reasons for such negative experiences include limited social support [23], low acceptance of hearing loss [24], problems with handling the devices [24-26], and poor use of supporting communication skills [22] (ie, strategies to enhance communication such as speechreading and reducing distance to the speaker). These reasons suggest that hearing rehabilitation should not be limited to HA fitting alone but also support users in addressing their residual hearing-related activity limitations and participation restrictions, while taking into account their personal and external contextual factors [27,28].

There are several educational communication programs providing support in these domains. They teach HA users knowledge and skills to effectively self-manage the multidimensional impact of hearing loss—for instance, by providing information about hearing, teaching communication strategies, or counseling to support better coping with the consequences of hearing loss [22,29-34]. While some programs proved to be effective in terms of HA benefit and use of communication strategies, their long-term effects are largely unknown [10]. Moreover, none of them were widely implemented in hearing health care practices [35], mainly because of limited resources [36] and high costs [28]. Delivery of communication programs via eHealth holds great promise because it potentially allows for providing services at the intensity that the patient prefers, automatized delivery with limited efforts for health care professionals, and wide reach, thereby ultimately improving (cost-) effectiveness and access to hearing care [37].

An example of a recently developed effective e-support program is the multimedia educational program c2HEAR [32]. This is based on reusable learning objects, which are short interactive videos covering information on HA handling, communication strategies, and adaptation to wearing HAs. They are delivered through DVD for television, over the internet, or on the PC [32]. The program was tested in a sample of first-time HA users attending the Nottingham Audiology Service (part of public hearing care in the United Kingdom) and appeared successful in improving HA use and practical HA skills. Another example is the effective web-based program by Malmberg et al [31], who evaluated it in Swedish general clinical practices among experienced HA users. Their program included online reading material combined with home training on hearing, HAs, and communication strategies and an online peer discussion forum. The program yielded improved communication skills [31]. It is crucial to test web-based support programs in these types of real-world, clinical care settings that are accessed by most of the adults with hearing loss seeking hearing care.

We recently developed a web-based self-management educational support program for adult HA users and their communication partners called SUPR (short for SUpport PRogram), to be offered within the HA dispensing care setting as an addition to a regular HA fitting trajectory. The home education program by Kramer et al [33], an intervention successful in improving communication strategy use and quality of life at 6-month follow-up, lies at the foundation of SUPR. The aim of our study was to evaluate the short term (ie, immediately postintervention) and the long term (ie, up to 12 months postintervention) effectiveness of SUPR on the use of communication strategies and a range of secondary outcomes compared with HA fitting alone. For that purpose, we performed a large-scale cluster randomized controlled trial (cRCT) including clients of one of the major HA dispenser chains in the Netherlands (Schoonenberg HoorSupport). The design of the cRCT has been described elsewhere [38]. The main hypothesis was that HA users receiving SUPR would improve their use of communication strategies (primary outcome) and improve several secondary outcomes (see Methods), while HA users receiving HA fitting without additional support would not. Furthermore, we hypothesized that the effects would be larger for first-time than for experienced HA users as was also observed for the home education program [33].

Methods

Study Design

The study was conducted and reported according to the Consolidated Standards of Reporting Trials (CONSORT) statement for cRCTs [39]. This study had a 2-arm cRCT design with the HA dispensing practice (henceforth: practice) as the unit of allocation. Cluster randomization was preferred over individual randomization to minimize the risk of contamination. The sample size calculation had indicated that 70 practices should participate in the study. Participating practices were prestratified by level of urbanization (located either in relatively rural or urban areas). Within both strata, a statistician randomized practices to either the intervention or control arm. The randomization sequence was generated by the statistical software R (R Foundation on Statistical Computing) with random permutation in blocks of size 4 with a fixed seed. Thirty-four practices were allocated to the control arm and 36 to the intervention arm.

Participants, HA dispensers (henceforth: dispensers), and researchers were aware of the practices' and participants' trial arm allocation. Participants and dispensers could not be blinded due to the nature of the intervention. Researchers could not be blinded because they administered all questionnaires to the participants, including the International Outcome Inventory (IOI), which revealed trial arm allocation. Additionally, researchers actively monitored the uptake of SUPR for the purposes of the process evaluation (submitted), also revealing arm allocation. The study was approved by the Dutch Institutional Review Board of the VU University Medical Center Amsterdam (IRB00002991; FWA number: FWA00017598) and registered [ISRCTN77340339].

Setting and Participants

Participants were recruited between February and September 2016 by the dispensers or their supporting staff of the participating practices. They informed each potentially eligible client about the study when clients were about to enter an HA trial period. Interested clients received an information package. The package contained information on the general study aims and group allocation (SUPR or control). No details about the SUPR program other than that it is a support program aimed at improving communication was provided at that point to ensure that clients would not ask or seek information about SUPR (thereby preventing contamination of the control group). Clients could sign up for the study via an online webpage and provided their consent there.

Eligible participants were Dutch speaking adults aged 50 years or over who had decided to purchase one or two HAs, had access to a device with internet connection, and were owner of an email account for the total duration of the study. These purchased HAs were the first ones for the first-time HA users and replacement HAs for the experienced users. We excluded participants who received additional care via an audiology clinic because this type of care might interfere with that provided in SUPR. We also excluded participants who purchased HAs to suppress tinnitus because tinnitus management was not included in SUPR.

Intervention

Control

Practices allocated to the control arm offered care as usual (HA fitting rehabilitation trajectory only). Briefly, the care usually provided included 4 face-to-face appointments with the dispenser. During the first appointment, a screening pure-tone audiogram was administered, and the client's goals and preferences were discussed. Clients were also asked to assign a communication partner and bring them along to future appointments. The communication partner could be any person the client communicated with on a regular basis (ie, a partner, child, neighbor, or caregiver). During the second appointment, additional hearing tests were performed. The HAs were then fitted immediately (if available) or at the third appointment. Clients were subsequently able to try out their HAs during a period of around 4 weeks in order to decide whether to purchase them. During a fourth appointment, the client's experiences and decision to purchase the HAs were discussed. Follow-up appointments were scheduled if needed. In addition, clients were able to visit the practice for small problems (ie, HA repairs) every working day during a service hour (4:00 to 5:00 pm) and contact the practice via telephone or an online form.

SUPR

The intervention group received the SUPR program as an addition to their HA fitting trajectory. A full description of the development of SUPR is reported elsewhere [40].

SUPR comprised 4 main elements: practical support booklet, emails, optional contact with the practice customer contact service, and involvement of a communication partner.

The practical support booklet, which clients received during their first appointment, contained tips and information on HAs, hearing loss, and communication strategies. Clients were instructed to write down their specific needs and goals they wished to reach with their HAs. These were discussed with the dispenser during follow-up appointments and were used for further refinement of the HA fitting.

A total of 17 emails were delivered over 6 months (approximately one email every 2 weeks). The first email was sent on the day the client started their HA trial period. The last email was sent approximately 6 months after the client had purchased the HA. Eleven emails contained links to educational videos, including 3 instruction videos on HA handling (eg, how to insert their specific type of HA), 5 videos on communication strategies and personal adjustment (eg, illustrating how to apply communication strategies in a birthday party setting), and 3 testimonial videos in which peers shared their experiences. Still photos from the different educational videos are shown in Figure 1. We are not able to provide a direct link to one of the videos in this manuscript since it is commercially sensitive information. The remaining 6 emails contained links to written communication strategy tips and information on HA maintenance. Two offered the client the option to contact the practice. The emails were delivered via a fully automated system. Access was free of charge. The intervention could be used ad libitum, and no reminders or prompts were used to encourage clients to watch the videos. The researchers were

able to check if participants had clicked on the email links to the educational videos.

Clients were invited to share their opinion regarding their HAs and SUPR by replying to emails 12 and 16 for optional contact with the practice customer contact service.

Figure 1. Stills of the SUPR educational videos: (1) testimonial video in which peers talk about their experiences with hearing loss; (2) instruction video on cleaning and maintenance of hearing aids; (3) video on how to apply effective communication strategies in a group setting; (4) instruction video on using assistive listening devices; (5) video on how to apply effective communication strategies in a one-to-one conversation; and (6) instruction video on how to insert hearing aids.

Clients were advised to assign a communication partner and involve them during the HA fitting trajectory and SUPR. The communication partner could sign up to receive the SUPR emails on behalf of their loved one if they did not own an email account. Clients were encouraged to watch the videos together with their communication partner. Communication partners were encouraged to write down their goals and experiences with the HAs of their loved ones in the practical support booklet.

Outcomes

All outcomes were collected via online questionnaires at baseline (t0: before the HA trial period), immediately postintervention (t1: 6 months after the HA purchase), 6 months postintervention (t2), and 12 months postintervention (t3). The outcomes were measured at all time points except for the ones that did not apply at t0 (ie, HA use—participants had not yet obtained their HAs).

Primary Outcome

The primary outcome was use of communication strategies as measured by 3 subscales of the Dutch Communication Profile for the Hearing Impaired (CPHI; maladaptive behaviors, verbal strategies, nonverbal strategies) [41,42]. Scores range from 1 to 5. The Dutch CPHI has a clear factor structure, and the subscales have good reported internal consistency (Cronbach alpha coefficients, henceforth alphas, between .81 and .86) [43].

Secondary Outcomes

Table 1 presents an overview of all secondary outcomes and the time points at which they were measured.

Table 1. Secondary outcomes measures (ranges) and time points.

Outcome measure	t0	t1	t2	t3
Personal adjustment to hearing loss (CPHI[a])				
Self-acceptance (1-5)	x	x	x	x
Acceptance of loss (1-5)	x	x	x	x
Stress and withdrawal (1-5)	x	x	x	x
Emotional response (HHDI[b]; 0-4)	x	x	x	x
Self-efficacy for HA[c] handling (MARS-HA[d])				
Basic (0-100)		x	x	x
Advanced (0-100)		x	x	x
HA use (IOI-HA[e]; 1-5)		x	x	x
HA pattern (use questionnaire; 1-5)		x	x	x
HA rehabilitation outcomes (IOI-HA)				
Satisfaction (1-5)		x	x	x
Quality of life (1-5)		x	x	x
HA rehabilitation outcomes and SUPR rehabilitation outcomes (IOI-HA and IOI-AI[f])				
Benefit (1-5)		x	x	x
Residual activity limitations (1-5)		x	x	x
Satisfaction (1-5)		x	x	x
Residual participation restrictions (1-5)		x	x	x
Impact on others (1-5)		x	x	x
Quality of life (1-5)		x	x	x
Recommendation of the services of the HA dispensing practice (1-10)	x	x	x	x
Readiness to act on hearing loss (URICA-HL[g])				
Precontemplation (problem denial; 1-5)	x	x	x	x
Contemplation (problem awareness; 1-5)	x	x	x	x
Preparation (information seeking and need for professional guidance; 1-5)	x	x	x	x
Action (healthy behavior acquisition or modification; 1-5)	x	x	x	x
Maintenance (sustained healthy behavior; 1-5)		x	x	x
Readiness (16-80)		x	x	x
Committed action (5-37)	x	x	x	x
Self-reported hearing disability (AIADH[h])				
Distinction of sounds (0-24)	x	x	x	x
Auditory localization (0-15)	x	x	x	x
Intelligibility in noise (0-15)	x	x	x	x
Intelligibility in quiet (0-15)	x	x	x	x
Detection of sounds (0-15)	x	x	x	x

[a]CPHI: Communication Profile for the Hearing Impaired.

[b]HHDI: Hearing Handicap and Disability Inventory.

[c]HA: hearing aids.

[d]MARS-HA: Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids.

[e]IOI-HA: International Outcome Inventory for Hearing Aids.

[f]IOI-AI: International Outcome Inventory for Alternative Interventions.

[g]URICA-HL: University of Rhode Island Change Assessment adapted for hearing loss.

[h]AIADH: Amsterdam Inventory for Auditory Disability and Handicap.

Psychosocial measures were assessed in two ways. First, personal adjustment to hearing loss was measured using 3 other subscales of the Dutch CPHI [41,42] (Table 1). The scales have good reported internal consistency (all alphas >.85) [43]. Second, the section emotional response of the Hearing Handicap and Disability Inventory (HHDI) was used [44]. As no psychometric information was available for this scale, we investigated the internal consistency in the current data set, which appeared good (alpha=.80).

Self-efficacy for HA handling was measured using the scales of the Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids (MARS-HA; Table 1) [45]. The MARS-HA has a clear factor structure, and the scales showed reasonable to good internal consistency (alphas .67 to .88) [45].

HA use was measured in two ways. First we used the use item of the validated IOI-HA scale [46,47]. Each item (including the use item) can be used separately [47]. Second, we used one item of the use questionnaire as developed by Laplante-Lévesque et al [48].

HA rehabilitation outcomes were measured using 2 items of the IOI-HA [46,47] for both the control and SUPR group (Table 1). The remaining rehabilitation outcomes were measured using 6 items of the IOI-HA for the control group and the IOI for Alternative Interventions (IOI-AI) for the SUPR group (Table 1) [49]. Comparing the item scores between the groups allowed us to compare rehabilitation outcomes (ie, comparing HA outcomes for the controls and SUPR rehabilitation outcomes for the SUPR group). The IOI-AI has good psychometric properties and individual items can be used [47,50] (and compared to their counterpart items of the IOI-HA) [49].

Recommendation of the services of the practice was measured using the question: “How likely is it that you would recommend the services of the practice to other people (family, friends, colleagues)?”

Self-reported hearing disability was measured using the 5 subscales of the Amsterdam Inventory for Auditory Disability and Handicap (AIADH) [51,52] (Table 1). The AIADH has a clear factor structure, and each subscale has good reported internal consistency (alphas .75 to .91) [52].

The Dutch version of the University of Rhode Island Change Assessment adapted for hearing loss (URICA-HL) [53] was used to assess participant readiness on 5 stages of change (Table 1). In addition to the stage scores, the composite readiness score (contemplation + action + maintenance scores – precontemplation) and the committed action score (action-contemplation) were calculated [54]. Laplante-Lévesque et al [54] found a clear factor structure for the stages and reported good internal consistency (all alphas >.76).

Statistical Analysis

Sample Size

The required sample size was calculated separately for first-time and experienced HA users. To detect a clinically meaningful effect of 0.67 [55] on the communication strategies CPHI subscale maladaptive behaviors (ie, the subscale with the smallest clinically meaningful effect) in first-time HA users (power 80%, significance level 5%, intracluster correlation coefficient [ICC] .01), two first-time HA users from each of the 70 practices would be needed in the analysis. To detect a clinically meaningful effect of 0.4 in experienced HA users (power 80%, significance level 5%, ICC .01), 3 clients from each practice would be needed in the analysis. As we anticipated a 20% loss to follow-up and a 30% recruitment rate, 4 first-time and 5 experienced clients were aimed to be recruited per practice.

Data Analysis

A statistical analysis plan was written and agreed upon before data analysis was started. Note that although we originally planned to use the overall summed score of the AIADH [38], it was agreed that the 5 subscale scores would provide a more detailed insight. In addition, measuring HA use objectively via HA data logging was not feasible as it turned out that these data were not collected as part of standard procedures in the practices.

Independent samples *t* tests, Mann-Whitney *U* tests, and chi-square tests were used to determine whether client characteristics were similarly distributed across the experimental conditions. Linear mixed models with fixed effects for group, time, and their 2-way interaction (ie, time*group), and random intercepts for subjects and practices were used to test differences in the course of the outcomes (t0 to t3 or t1 to t3) between the groups. Post hoc analyses based on the estimated fixed effects were carried out in case significant group differences in the course in outcomes were found to assess at which time points these occurred. For outcomes assessed at t1 to t3, we also examined whether there was a significant difference between the groups at t1 to determine the immediate posttreatment effect. To illustrate, in case of a statistically significant group difference at t1, a nonstatistically significant interaction term would indicate that this group difference was maintained at t2 and t3. In contrast, a statistically significant interaction term would indicate that this group difference changed (ie, either disappeared or worsened). Nonnormality was checked for all outcomes and data were transformed when necessary. Potential confounders (client characteristics) were examined for all outcomes and added as (fixed) covariates to the model in case they were differently distributed (P<.05) across groups at t0. Subgroup differences, using 3-way (time*group*type of client) interaction terms, were performed to check whether any intervention effects differed between first-time and experienced clients.

Main analyses were performed on the principle of intention to treat (ITT). A per protocol analysis was additionally performed including SUPR recipients who had clicked through to a video on communication strategies and personal adjustment from at least two emails and controls who did not receive any SUPR emails. Bonferroni corrections for multiple testing were applied for the primary outcome (3 subscales) and for the post hoc analyses (3 follow-up measurements), such that a P<.016 (0.05/3) was considered to indicate a statistically significant group difference. For all secondary outcomes, a P value of <.05

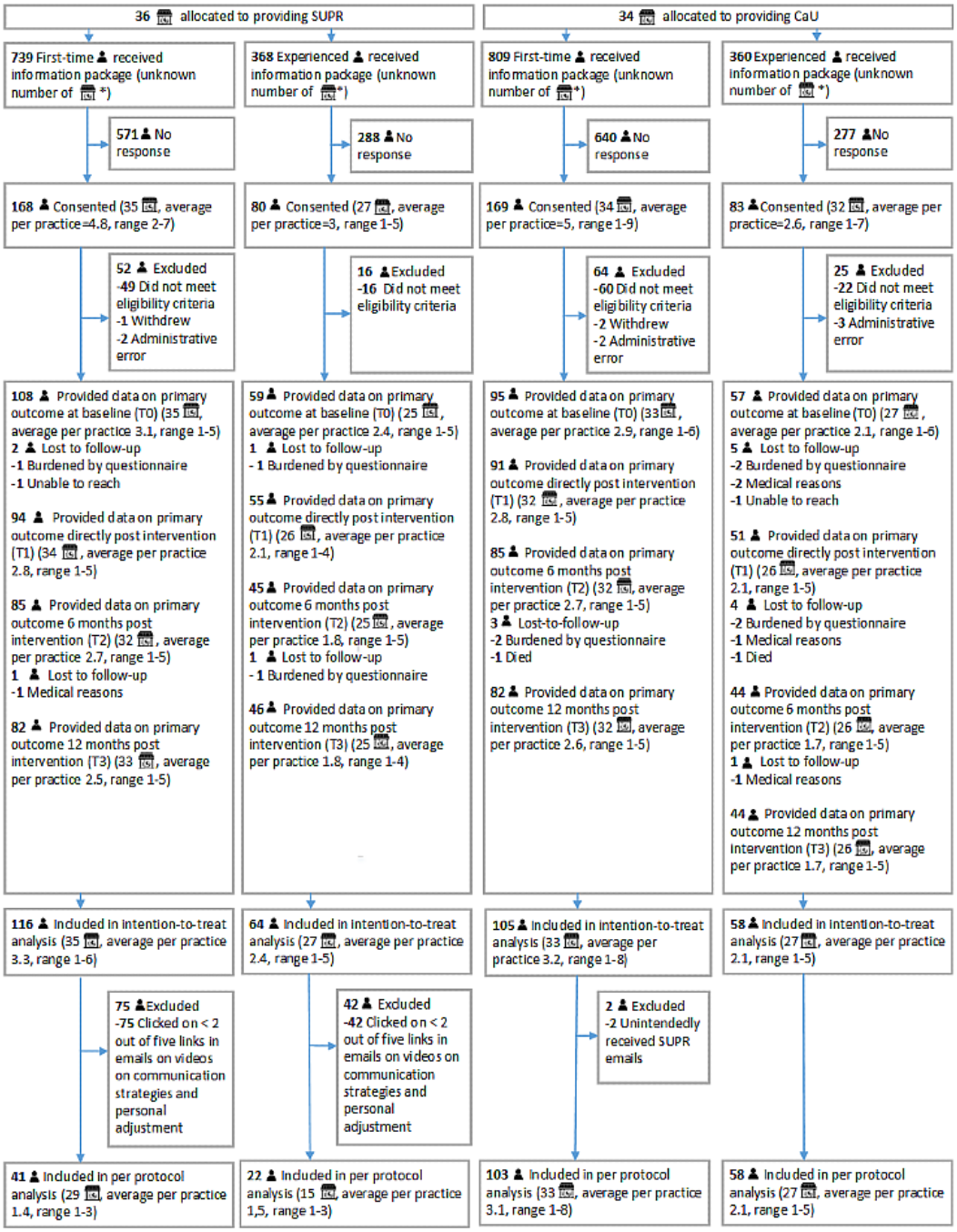
was considered statistically significant. In case significant group differences were found at t1, t2, or t3, mean differences between the mean values in the SUPR and control group were reported along with their 95% confidence interval and *P* value. Because a linear mixed model gives unbiased results in the presence of missing data, imputation of missing outcomes was not considered. Thus, clients were included in the analyses if they had provided data at one time point at least. Analyses were carried out using SPSS Statistics version 26 (IBM Corp).

Results

Study Population

Figure 2 displays the flow of first-time and experienced clients through the study. Between February and September 2016, 1107 (739 first-time and 368 experienced) clients in the intervention arm and 1169 (809 first time and 360 experienced) clients in the control arm were invited to participate in the study. The number of invited clients is an estimation based on the reported numbers of invited clients of the practices complying with the protocol to report this.

Figure 2. Flow of participants through the study. *Per practice, an unknown number of respondents was invited to participate due to some dispensers' noncompliance with the protocol to report this. The number of invited clients is an estimation based on the reported numbers of invited clients of the practices complying with the protocol to report this.



In total, 248 clients (across 35 practices in the SUPR arm) and 252 clients (34 practices in the control arm) enrolled for the study and consented to participate. Of these, 180 and 163 clients were included in the ITT analysis, respectively. Loss to follow-up was 2.8% (5/180) in the SUPR arm and 8.0% in the control arm (13/163). Sixty-three clients in the SUPR arm and 161 controls were included in the per protocol analysis.

Table 2 presents the baseline characteristics of the participants. The number of participants in each practice ranged from 1 to 11 (mean 5). Of the total sample, 60% (206/343) were male (mean age 68.1 [SD 8.5] years), and mean pure-tone hearing loss was 43.7 (SD 11.1) decibels Hearing Level. The characteristics for SUPR and control participants were similar, as were the outcomes at baseline (P>.05), indicating that correction in the analyses due to significant group differences was not required.

Table 2. Baseline characteristics.

Characteristics	SUPR group	Control group
Sex	n=180	n=163
Male, n (%)	108 (60)	98 (60)
Age in years, mean (SD)	68.1 (8.4)	68.2 (8.7)
Marital status, n (%)	n=177	n=158
Married	130 (73)	111 (70)
Cohabiting	9 (5)	8 (5)
Widowed	24 (14)	14 (9)
Divorced	7 (4)	16 (10)
Single, never married	7 (4)	9 (6)
Living situation, n (%)	n=177	n=158
Living together with other people	144 (81)	122 (77)
Living alone	33 (19)	36 (23)
Educational level, n (%)	n=177	n=157
Low	38 (22)	28 (18)
Middle	123 (70)	110 (70)
High	16 (9)	19 (12)
Paid job, n (%)	n=177	n=158
Yes	39 (22)	38 (24)
No	138 (78)	120 (76)
Country of birth, n (%)	n=177	n=158
The Netherlands	162 (92)	149 (94)
Other	15 (9)	9 (6)
Better ear average hearing loss in dB HL[a], mean (SD)	43 (11.7)	44.5 (10.5)
Bilaterally fitted hearing aids (ie, two ears)[b]	132 (89)	126 (89)

[a]dB HL: decibels hearing level averaged across 1, 2, and 4 kilohertz.

[b]Measured via the t1 questionnaire (n=291).

Primary Outcomes

Intention to Treat and Per Protocol Analysis

The ITT analysis showed no statistically significant (P≥.016) group differences in the course of communication strategy use (Table 3). The per protocol analysis showed no significant (P≥.016) group differences either (Table 3). There were no differences in effects between first-time and experienced clients.

Table 3. Descriptive statistics and results of the linear mixed models on communication strategy use (Communication Profile for the Hearing Impaired; primary outcome).

Communication strategy use subscales and group	T0		T1		T2		T3		*P* value[a]
	n	mean (SD)	n	mean (SD)	n	mean (SD)	n	mean (SD)	
Intention to treat									
Maladaptive behaviors									.84
SUPR group	167	4.4 (0.5)	149	4.6 (0.4)	130	4.6 (0.4)	128	4.6 (0.4)	
Control group	152	4.4 (0.6)	142	4.5 (0.5)	129	4.5 (0.6)	126	4.6 (0.5)	
Verbal strategies									.09
SUPR group	167	2.2 (0.6)	149	2.3 (0.7)	130	2.3 (0.7)	128	2.3 (0.7)	
Control group	152	2.3 (0.7)	142	2.2 (0.7)	129	2.2 (0.7)	126	2.3 (0.7)	
Nonverbal strategies									.09
SUPR group	167	2.9 (0.8)	149	3.0 (0.9)	130	2.9 (0.9)	128	2.9 (0.9)	
Control group	152	3.0 (1.0)	142	2.9 (0.9)	129	2.9 (0.9)	126	2.9 (0.9)	
Per protocol									
Maladaptive behaviors									.92
SUPR group	60	4.4 (0.5)	56	4.6 (0.4)	54	4.5 (0.5)	46	4.6 (0.4)	
Control group	151	4.4 (0.6)	141	4.5 (0.5)	128	4.5 (0.6)	125	4.6 (0.5)	
Verbal strategies									.11
SUPR group	60	2.2 (0.6)	56	2.3 (0.8)	54	2.4 (0.9)	46	2.3 (0.8)	
Control group	151	2.3 (0.7)	141	2.2 (0.7)	128	2.2 (0.7)	125	2.3 (0.7)	
Nonverbal strategies									.06
SUPR group	60	2.8 (0.9)	56	2.9 (0.9)	54	2.9 (1.0)	46	2.9 (0.9)	
Control group	151	3.1 (1.0)	141	2.9 (0.9)	128	2.9 (0.9)	125	2.9 (0.9)	

[a]*P* value for difference in the course of the outcomes between groups (interaction term time*group). A *P* value of <.016 was considered statistically significant.

Secondary Outcomes

Intention to Treat Analysis

There were no differences in secondary outcomes between first-time and experienced clients unless stated otherwise. The results for the psychosocial outcomes are presented in Table 4. There were no statistically significant group differences in the course of these outcomes (*P*≥.05).

The results on self-efficacy for HA handling are presented in Multimedia Appendix 1. A statistically significant (*P*<.05) group difference in self-efficacy for basic HA handling over time (time*group: *P*=.01) was observed. However, post hoc analyses showed no statistically significant (*P*≥.016) differences at the follow-up measurements. Immediately postintervention, the self-efficacy for advanced HA handling scores were significantly (*P*<.05) higher for the SUPR group than for the control group (5.3, 95% CI 0.3-10.4; *P*=.04). This effect was sustained at 6- and 12-month follow-up (time*group *P*=.56, so *P*≥.05).

Multimedia Appendix 2 shows the results for HA use. Immediately postintervention, the SUPR group had significantly (*P*<.05) greater HA use compared with the controls (mean difference 0.3, 95% CI 0.02 to 0.5; *P*=.03). This group difference was not maintained at 6- and 12-month follow-up (time*group *P*=.008, so *P*<.05). There was a statistically significant group difference neither immediately postintervention nor in the course of the outcome HA use pattern.

The results of the IOI-HA and IOI-AI item scores are also presented in Multimedia Appendix 2. Note that the IOI-HA items on satisfaction and quality of life were assessed both for SUPR and control groups. Immediately postintervention, HA satisfaction was significantly (*P*<.05) greater in the SUPR group than in the controls (mean difference 0.3, 95% CI 0.09 to 0.5; *P*=.006). This group difference was maintained at 6- and 12-month follow-up (ie, time*group *P*=.05, so *P*≥.05). There was no significant group difference in the course of quality of life. When examining the IOI outcomes in which HAs were directly compared with SUPR, there were significant (*P*<.05) group differences in favor of the controls immediately postintervention on satisfaction with the intervention (mean difference –0.6, 95% CI –0.9 to –0.4; *P*<.001), benefit experienced from the intervention (mean difference –1.0, 95% CI –1.3 to –0.8; *P*<.001), and quality of life (mean difference –0.4, 95% CI –0.7 to –0.2; *P*<.001). These effects were maintained at 6- and 12-month follow-up (time*group *P*=.30, time*group *P*=.24, and time*group *P*=.41, respectively, so

P≥.05). In other words, controls experienced greater levels of satisfaction, benefit, and quality of life resulting from their HAs than the SUPR recipients did with the SUPR program. There were no significant group differences when comparing the IOI-HA and IOI-AI item residual activity limitations, residual participation restrictions, and impact on others.

Table 4. Descriptive statistics and results of the linear mixed models on personal adjustment (Communication Profile for the Hearing Impaired) and emotional response (Hearing Handicap and Disability Inventory; secondary outcomes).

Psychosocial outcomes and group	T0		T1		T2		T3		*P* value[a]
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	
Personal adjustment to hearing loss									
Self-acceptance									.63
SUPR group	167	4.2 (0.7)	149	4.4 (0.6)	129	4.4 (0.7)	128	4.5 (0.5)	
Control group	152	4.2 (0.8)	142	4.4 (0.7)	128	4.4 (0.6)	126	4.4 (0.7)	
Acceptance of loss									.12
SUPR group	167	3.6 (0.7)	149	3.9 (0.8)	129	3.8 (0.8)	128	3.8 (0.8)	
Control group	152	3.5 (0.8)	142	3.9 (0.8)	128	3.8 (0.9)	126	4.0 (0.8)	
Stress and withdrawal									.89
SUPR group	167	3.5 (0.7)	149	3.9 (0.7)	129	3.8 (0.8)	128	3.8 (0.8)	
Control group	152	3.5 (0.9)	142	3.9 (0.8)	128	3.9 (0.8)	126	3.9 (0.8)	
Emotional response									.36
SUPR group	167	1.3 (0.7)	149	0.9 (0.7)	127	1.0 (0.7)	128	1.0 (0.7)	
Control group	151	1.4 (0.8)	140	1.0 (0.8)	126	1.0 (0.7)	124	1.0 (0.7)	

[a]*P* value for difference in the course of the outcomes between groups (interaction term time*group). A *P* value of <.05 was considered statistically significant.

Multimedia Appendix 3 displays the results on recommendation of the practice services, readiness to act on hearing loss, and self-reported hearing disability. A significant (*P*<.05) group difference over the course of the URICA-HL action score (time*group: *P*=.01) was found. However, the post hoc analyses indicated no statistically significant (*P*≥.016) group differences at the follow-up measurements. A significant (*P*<.05) interaction for type of HA client was found for the URICA-HL committed action score (*P*=.03). A significant (*P*<.05) group difference was found for experienced clients (time*group *P*=.001), while for first time clients there was no difference (*P*=.46). However, post hoc analyses showed no statistically significant (*P*≥.016) group differences for the experienced clients on the committed action scores at the follow-up measurements. No significant group differences were observed for the other URICA scales or self-reported hearing disability.

Per Protocol Analysis

In total, 35.0% (63/180) of clients in the SUPR group were included in the per protocol analysis as they had clicked through to a video on communication strategies and personal adjustment from at least two emails. Almost all control clients (98.8%) (161/163) were included in the per protocol analysis; only 2 controls had received SUPR (due to an administrative error). Differences between ITT and per protocol analysis were observed for some outcomes. First, contrary to the significantly better self-efficacy for advanced HA handling and HA use observed for the SUPR group in the ITT analysis, the per protocol analysis did not reveal any significant (*P*≥.05) group differences immediately postintervention (*P*=.23 and *P*=.052, respectively). Second, contrary to the ITT analysis, there was no significant (*P*≥.05) difference in the course of the action scores (time*group *P*=.21). Next, contrary to the absence of a difference on HA use pattern found in the ITT analysis, the per protocol analysis showed a significant (*P*<.05) group difference immediately postintervention, such that the SUPR group had a more stable pattern in HA use than controls (–0.4, 95% CI –0.7 to 0.04; *P*=.03; a lower score indicated a more stable pattern in HA use). This effect was sustained at 6- and 12-month follow-up (time*group *P*=.92, so *P*≥.05). Fourth, contrary to the ITT analysis, there was no significant (*P*≥.05) group difference in quality of life resulting from using the intervention immediately postintervention (*P*=.052). Last, similar to the ITT analysis, the significant group difference on satisfaction with the intervention and benefit experienced from the intervention was also found in the per protocol analysis. However, contrary to the ITT analysis, the per protocol analysis showed statistically significant (*P*<.05) different effects for first-time and experienced clients. For both outcomes, a long-term effect was observed for experienced clients (similar to the ITT analysis). For new clients, the effect was only observed at 6- and 12-month follow-up and not immediately postintervention.

Discussion

Principal Findings

Using a cRCT design, we evaluated the effectiveness of SUPR, a web-based self-management support program provided as an

addition to regular HA fitting to HA users aged 50 years and over in order to improve the self-management of hearing difficulties and HA use. The study showed that SUPR did not lead to more frequent use of communication strategies (primary outcome) compared with care as usual. Nonetheless, SUPR significantly improved clients' self-efficacy for advanced HA handling and HA satisfaction at 12 months, as well as HA use immediately postintervention. No group differences were observed for any of the remaining secondary outcomes.

It is encouraging to see that SUPR was able to significantly increase HA use immediately postintervention. Although the effect seems small (mean difference of 0.3 on a scale from 1 to 5), the fact that SUPR was able to improve HA use can be considered valuable because a recently published systematic review indicated that there is no evidence of interventions showing any improvements on HA use on the short, medium, or long term [10]. The evidence was judged as limited because the majority of previous studies had nonpowered small sample sizes (limiting the occurrence of statistically significant differences) or were carried out in nonrepresentative samples (ie, military veteran populations). The SUPR study can thus be regarded as an important addition to the existing body of evidence. Moreover, the per protocol analysis revealed a significant positive effect of SUPR on HA use stability in the long term, indicating a more stable HA use pattern (ie, the same number of hours of HA use every day) among SUPR participants than among controls. The absence of a long-term improvement in HA use may imply that without support like SUPR, HA clients may tend to fall back and be left to full self-management. This underlines the importance of having follow-up support after clients have completed SUPR to increase the likelihood of extending the effects to the long term.

It has been argued that more frequent or more stable HA use does not automatically imply more satisfaction with HAs [56] and that HA use (in hours) alone cannot be viewed as an indicator of successful HA use [57]. From this perspective, it is interesting to see that we not only found increased HA use in SUPR recipients, but also significantly greater HA satisfaction than the controls postintervention. This shows that the increase in hours of use indeed coincided with greater HA satisfaction.

This study demonstrated that the SUPR group had a significantly higher score of 5.3 points (95% CI 0.3 to 10.4; scored on a scale from 0 to 100) on the advanced HA handling self-efficacy scale directly postintervention than controls that lasted up to 12 months later. Both groups seemed to perform at ceiling on the basic HA handling self-efficacy scale, suggesting that during appointments dispensers already ensured clients became skilled in basic HA handling (ie, HA and battery insertion and removal, HA cleaning and maintenance). This was also stated by Ferguson et al [32], who found similar ceiling effects. Although the instruction videos mostly focused on basic rather than advanced HA handling skills, watching them might have increased clients' confidence in their ability to also handle more advanced skills, like troubleshooting or naming the model of a particular HA. We are uncertain about the clinical meaning of a difference of 5.3 points, however. Further research should address this.

Bennett et al [58] found that most problems HA owners experienced were related to HA management including HA use, handling, and ongoing care and that these had the greatest impact on HA success. SUPR improving HA clients' confidence in their ability to manage their HAs can thus be considered an encouraging finding. Other work by Bennett et al [59] showed that HA management can be classified according to two overarching themes: the device and the person. The results of our study suggest that SUPR is successful in improving HA management concepts related to the device (HA maintenance and repairs, daily HA use, advanced HA knowledge) but not for HA management related to the person (learning to come to terms with HAs, communication strategies, working with a clinician). This is further discussed below.

Previous studies have shown that both individual and group auditory rehabilitation interventions can effectively increase communication strategy use [22,31,33,60] both in the short- and long term. A plausible explanation for the absence of an effect in this study may be related to the video-watching rates as reported in our process evaluation study (submitted). Whereas up to 37% of the participants in the intervention arm had watched the instruction videos, only 7% to 16% had watched the videos on communication strategies and personal adjustment. Not engaging with web-based interventions as was intended is a well-known problem in intervention effectiveness research [61,62]. The per protocol analysis still showed no effects on communication strategy use, but it must be noted that the sensitivity analysis was most likely underpowered (SUPR n=60, care as usual n=151). We intended to perform the per protocol analysis with clients who had clicked on more than two video links, but this resulted in samples too small to allow a meaningful statistical analysis.

An alternative explanation for the absence of a communication strategy use or personal adjustment effect may be related to the setting in which SUPR was provided. Given that the HA dispensing setting is a primary care one, mainly focusing on HA fitting and dispensing, clients may not have expected or been ready for receiving educational videos. Such type of rehabilitation may better fit in specialized (ie secondary or tertiary) hearing health care settings. Kramer et al [33] chose a tertiary setting to provide their home education program, resulting in improved communication strategy skills.

Control group participants reported significantly greater benefit, satisfaction, and quality of life because of their HAs (IOI-HA) than SUPR participants reported for SUPR (IOI-AI). This suggests that HAs were viewed as more impactful than SUPR. In a way, this is not surprising, since HAs can be considered as a basis. They amplify sounds and may thus improve listening ability and thereby quality of life [63]. It is the combination of the two (HAs and additional support) that may be most beneficial [10]. Also, it is important to consider that HAs have a long history (more than a century) of development while the development of (web-based) support programs is still in its infancy. There is much to learn still to further refine communication programs to ensure they fit clients' needs and have a larger impact.

The URICA questionnaire outcomes conflicted with what was expected both in the direction of the effect and changes over time and were therefore difficult to interpret. We expected the SUPR group to show an increase in the action scores, while in fact the follow-up scores were lower than baseline in both groups. Differences in interpretation of what taking action meant to someone may have occurred between participants but also within participants over time (as the intervention might have influenced what taking action was) causing invalid measurements. This is further discussed in Meijerink et al [40].

Strengths and Limitations

This was the first audiological rehabilitation study implementing and evaluating a web-based self-management support program on such a large scale and in a real-life HA dispensing setting. The large sample size, use of a robust RCT design, and outcome assessment at the short, medium, and long term can be considered unique in our research field. Given these strengths, the study tackles most of the limitations mentioned in a recent meta-analysis on intervention studies to improve HA use [10]. Including 70 clusters across the Netherlands and purposefully sampling for spread in degree of rural/urban areas minimized imbalance across treatment groups and increased the generalizability of the findings to the Dutch real-world practice. Nevertheless, there are a number of limitations.

First, while we did reach our targeted sample size for the first-time clients, we did not for the experienced clients due to recruitment difficulties. The limited number of experienced clients may have resulted in nonsignificant differences in effects between first-time and experienced clients. A second limitation is that clients had to provide consent for study participation while knowing their group allocation. This may have affected clients' willingness to participate in the study and introduced selection bias. Unfortunately, it was not possible to prevent this because randomizing the practices after obtaining clients' consent would have delayed the intervention period by 7 months until clients had provided their consent. This was deemed unacceptable given the real-life character of the study. A third limitation was that participating clients, dispensers and researchers could not be blinded. This may have introduced bias. Possibly, clients who were aware of receiving SUPR may have responded more favorably compared to controls, while controls being aware of receiving standard care only may have sought alternative treatments (which would have increased the likelihood of contamination). We attempted, however, to prevent controls from seeking access to SUPR by reducing the amount of information given about SUPR content and by offering controls SUPR after study completion. A fourth limitation is that we initially aimed to measure HA use using data logging in order to measure HA use objectively, but this appeared unfeasible. Hence, all outcomes were self-reported, possibly resulting in overreporting of HA use [18,48]. Finally, behavior change was expected at multiple levels, and therefore many outcomes were evaluated. This increased the likelihood of finding statistically significant results by chance. We therefore applied Bonferroni corrections for the primary outcomes and the post hoc analyses but not for the secondary outcomes. It should be noted that there is debate among statisticians as to when multiple outcomes should be corrected for [64,65]. Using a Bonferroni correction for all outcomes is concerned too strict by some as it would increase the chance for false negatives [66].

Conclusions

While the popularity of web-based platforms to complement HA fitting is rising [30-32], high-quality evidence (ie, assuring external validity and power) to show the long-term benefits of eHealth in HA rehabilitation is still lacking [37]. This study is a valuable addition to the existing evidence for such platforms in hearing health care. While SUPR did not significantly enhance the use of communication strategies, this study provides evidence for the effectiveness of SUPR to improve self-efficacy for HA handling and HA satisfaction in the long term and HA use in the short term. Given that the effects were mainly found in the HA handling domain, this study indicates that an intervention offering web-based instructions is a promising addition to the services provided by dispensers. Further research is needed to evaluate if adjustments to SUPR will lead to a higher adherence of clients in following the intervention to improve the (long-term) effectiveness of communication strategy use and other psychosocial outcomes.

Acknowledgments

We thank all participants for taking part in the SUPR study as well as the hearing aid dispensers of the participating hearing aid dispensing practices; Schoonenberg HoorSupport (AudioNova International BV) for their contribution to the study design and setup and support in writing this paper; Bregje van der Wouden for her role in the early phase of this project; Liselotte Vuijk and Nancy van der Mark for their contribution to the data collection; Josine Stuber for her advice on the statistical analyses; and Bernadette Paulissen for her vision in founding this project. This study is funded by AudioNova International BV and sponsored by the VU University Medical Centre. Employees of AudioNova International BV were involved in the study design and data collection and were given the opportunity to comment on the manuscript but had no role in the data analysis and/or the interpretation of the data, approval of the manuscript, and the decision to submit the article for publication.

Authors' Contributions

SEK obtained funding for this research. SEK and MP designed the study, assisted by VJ. JFJM and MP developed the analysis plan, supported by BLW. JFJM collected the data, assisted by VJ and MP and supervised by MP and SEK. JFJM performed statistical analyses, supported by BLW. JFJM wrote the article and input was provided by all coauthors. All authors approved the final manuscript.

Conflicts of Interest

Most of JFJM's appointment at the Amsterdam University Medical Center as a PhD student on the SUPR project (including completing the tasks related to the submitted work) and the design and implementation of the SUPR study were facilitated through a research grant sponsored by AudioNova International BV. MP was employed as a researcher at Schoonenberg HoorSupport (daughter company of AudioNova International BV) for a 6-month period on other research work, received a (cofunding) research grant from Sonova AG (mother company of AudioNova International BV) for other research work, and has been paid for delivering a one-off scientific presentation for Sonova AG. VJ is an employee at Schoonenberg HoorSupport and SEK has been paid for delivering a presentation for Sonova AG; no other relationships or activities that could appear to have influenced the submitted work can be reported.

Multimedia Appendix 1
Descriptive statistics and results of the linear mixed models on self-efficacy for hearing aid handling (Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids; secondary outcome).
[DOCX File , 16 KB - jmir_v22i9e17927_app1.docx]

Multimedia Appendix 2
Descriptive statistics and results of the linear mixed models on hearing aid use (International Outcome Inventory for Hearing Aids) and hearing aid pattern (use questionnaire; secondary outcomes).
[DOCX File , 19 KB - jmir_v22i9e17927_app2.docx]

Multimedia Appendix 3
Descriptive statistics and results of the linear mixed models on recommendation of the services of the hearing aid dispensing practice, readiness to act on hearing loss (University of Rhode Island Change Assessment adapted for hearing loss), and self-reported hearing disability (Amsterdam Inventory for Auditory Disability and Handicap; secondary outcomes).
[DOCX File , 26 KB - jmir_v22i9e17927_app3.docx]

References

1. Hearing loss in persons 65 years and older based on WHO global estimates on prevalence of hearing loss. World Health Organization. 2012. URL: https://www.who.int/pbd/deafness/news/GE_65years.pdf [accessed 2020-08-10]
2. Jiam NT, Li C, Agrawal Y. Hearing loss and falls: A systematic review and meta-analysis. Laryngoscope 2016 Nov;126(11):2587-2596. [doi: 10.1002/lary.25927] [Medline: 27010669]
3. Pronk M, Deeg DJH, Smits C, van Tilburg TG, Kuik DJ, Festen JM, et al. Prospective effects of hearing status on loneliness and depression in older persons: identification of subgroups. Int J Audiol 2011 Dec;50(12):887-896. [doi: 10.3109/14992027.2011.599871] [Medline: 21929374]
4. Saito H, Nishiwaki Y, Michikawa T, Kikuchi Y, Mizutari K, Takebayashi T, et al. Hearing handicap predicts the development of depressive symptoms after 3 years in older community-dwelling Japanese. J Am Geriatr Soc 2010 Jan;58(1):93-97. [doi: 10.1111/j.1532-5415.2009.02615.x] [Medline: 20002512]
5. Strawbridge WJ, Wallhagen MI, Shema SJ, Kaplan GA. Negative consequences of hearing impairment in old age: a longitudinal analysis. Gerontologist 2000 Jun;40(3):320-326. [doi: 10.1093/geront/40.3.320] [Medline: 10853526]
6. Thomson RS, Auduong P, Miller AT, Gurgel RK. Hearing loss as a risk factor for dementia: a systematic review. Laryngoscope Investig Otolaryngol 2017 Apr;2(2):69-79 [FREE Full text] [doi: 10.1002/lio2.65] [Medline: 28894825]
7. Lin FR, Metter EJ, O'Brien RJ, Resnick SM, Zonderman AB, Ferrucci L. Hearing loss and incident dementia. Arch Neurol 2011 Feb;68(2):214-220 [FREE Full text] [doi: 10.1001/archneurol.2010.362] [Medline: 21320988]
8. Gopinath B, Schneider J, McMahon CM, Burlutsky G, Leeder SR, Mitchell P. Dual sensory impairment in older adults increases the risk of mortality: a population-based study. PLoS One 2013;8(3):e55054 [FREE Full text] [doi: 10.1371/journal.pone.0055054] [Medline: 23469161]
9. Miyawaki A, Kobayashi Y, Kawachi I. Self-reported hearing/visual loss and mortality in middle-aged and older adults: findings from the Komo-Ise Cohort, Japan. J Epidemiol 2020 Feb 05;30(2):67-73 [FREE Full text] [doi: 10.2188/jea.JE20180198] [Medline: 30662042]
10. Barker F, Mackenzie E, Elliott L, Jones S, de Lusignan S. Interventions to improve hearing aid use in adult auditory rehabilitation. Cochrane Database Syst Rev 2016 Aug 18(8):CD010342 [FREE Full text] [doi: 10.1002/14651858.CD010342.pub3] [Medline: 27537242]
11. Rumalla K, Karim AM, Hullar TE. The effect of hearing aids on postural stability. Laryngoscope 2015 Mar;125(3):720-723. [doi: 10.1002/lary.24974] [Medline: 25346316]
12. Dawes P, Emsley R, Cruickshanks KJ, Moore DR, Fortnum H, Edmondson-Jones M, et al. Hearing loss and cognition: the role of hearing AIDS, social isolation and depression. PLoS One 2015;10(3):e0119616 [FREE Full text] [doi: 10.1371/journal.pone.0119616] [Medline: 25760329]

13. Choi JS, Betz J, Li L, Blake CR, Sung YK, Contrera KJ, et al. Association of using hearing aids or cochlear implants with changes in depressive symptoms in older adults. JAMA Otolaryngol Head Neck Surg 2016 Jul 01;142(7):652-657. [doi: 10.1001/jamaoto.2016.0700] [Medline: 27258813]
14. Amieva H, Ouvrard C, Giulioli C, Meillon C, Rullier L, Dartigues J. Self-reported hearing loss, hearing aids, and cognitive decline in elderly adults: a 25-year study. J Am Geriatr Soc 2015 Oct;63(10):2099-2104. [doi: 10.1111/jgs.13649] [Medline: 26480972]
15. Castiglione A, Benatti A, Velardita C, Favaro D, Padoan E, Severi D, et al. Aging, cognitive decline and hearing loss: effects of auditory rehabilitation and training with hearing aids and cochlear implants on cognitive function and depression among older adults. Audiol Neurootol 2016;21 Suppl 1:21-28 [FREE Full text] [doi: 10.1159/000448350] [Medline: 27806352]
16. Bertoli S, Staehelin K, Zemp E, Schindler C, Bodmer D, Probst R. Survey on hearing aid use and satisfaction in Switzerland and their determinants. Int J Audiol 2009 Apr;48(4):183-195. [doi: 10.1080/14992020802572627] [Medline: 19363719]
17. Abrams H, Kihm J. An introduction to MarkeTrak IX: a new baseline for the hearing aid market. 2015. URL: https://www.hearingreview.com/practice-building/marketing/introduction-marketrak-ix-new-baseline-hearing-aid-market [accessed 2020-08-10]
18. Solheim J, Hickson L. Hearing aid use in the elderly as measured by datalogging and self-report. Int J Audiol 2017 Jul;56(7):472-479. [doi: 10.1080/14992027.2017.1303201] [Medline: 28332420]
19. Kaplan-Neeman R, Muchnik C, Hildesheimer M, Henkin Y. Hearing aid satisfaction and use in the advanced digital era. Laryngoscope 2012 Sep;122(9):2029-2036. [doi: 10.1002/lary.23404] [Medline: 22752928]
20. Aazh H, Prasher D, Nanchahal K, Moore BCJ. Hearing-aid use and its determinants in the UK National Health Service: a cross-sectional study at the Royal Surrey County Hospital. Int J Audiol 2015 Mar;54(3):152-161. [doi: 10.3109/14992027.2014.967367] [Medline: 25395258]
21. Hartley D, Rochtchina E, Newall P, Golding M, Mitchell P. Use of hearing AIDS and assistive listening devices in an older Australian population. J Am Acad Audiol 2010;21(10):642-653. [doi: 10.3766/jaaa.21.10.4] [Medline: 21376005]
22. Chisolm TH, Abrams HB, McArdle R. Short- and long-term outcomes of adult audiological rehabilitation. Ear Hear 2004 Oct;25(5):464-477. [doi: 10.1097/01.aud.0000145114.24651.4e] [Medline: 15599193]
23. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. Int J Audiol 2014 Feb;53 Suppl 1:S18-S27. [doi: 10.3109/14992027.2013.860488] [Medline: 24447233]
24. Knudsen LV, Oberg M, Nielsen C, Naylor G, Kramer SE. Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: a review of the literature. Trends Amplif 2010 Sep;14(3):127-154 [FREE Full text] [doi: 10.1177/1084713810385712] [Medline: 21109549]
25. Solheim J, Gay C, Hickson L. Older adults' experiences and issues with hearing aids in the first six months after hearing aid fitting. Int J Audiol 2018 Jan;57(1):31-39. [doi: 10.1080/14992027.2017.1380849] [Medline: 28952841]
26. McCormack A, Fortnum H. Why do people fitted with hearing aids not wear them? Int J Audiol 2013 May;52(5):360-368 [FREE Full text] [doi: 10.3109/14992027.2013.769066] [Medline: 23473329]
27. Hickson L, Scarinci N. Older adults with acquired hearing impairment: applying the ICF in rehabilitation. Semin Speech Lang 2007 Nov;28(4):283-290. [doi: 10.1055/s-2007-986525] [Medline: 17935013]
28. Boothroyd A. Adult aural rehabilitation: what is it and does it work? Trends Amplif 2007 Jun;11(2):63-71 [FREE Full text] [doi: 10.1177/1084713807301073] [Medline: 17494873]
29. Hickson L, Worrall L, Scarinci N. A randomized controlled trial evaluating the active communication education program for older people with hearing impairment. Ear Hear 2007 Apr;28(2):212-230. [doi: 10.1097/AUD.0b013e31803126c8] [Medline: 17496672]
30. Thorén ES, Oberg M, Wänström G, Andersson G, Lunner T. A randomized controlled trial evaluating the effects of online rehabilitative intervention for adult hearing-aid users. Int J Audiol 2014 Jul;53(7):452-461. [doi: 10.3109/14992027.2014.892643] [Medline: 24749664]
31. Malmberg M, Lunner T, Kähäri K, Andersson G. Evaluating the short-term and long-term effects of an internet-based aural rehabilitation programme for hearing aid users in general clinical practice: a randomised controlled trial. BMJ Open 2017 Jun 06;7(5):e013047 [FREE Full text] [doi: 10.1136/bmjopen-2016-013047] [Medline: 28592571]
32. Ferguson M, Brandreth M, Brassington W, Leighton P, Wharrad H. A randomized controlled trial to evaluate the benefits of a multimedia educational program for first-time hearing aid users. Ear Hear 2016;37(2):123-136 [FREE Full text] [doi: 10.1097/AUD.0000000000000237] [Medline: 26565785]
33. Kramer SE, Allessie GHM, Dondorp AW, Zekveld AA, Kapteyn TS. A home education program for older adults with hearing impairment and their significant others: a randomized trial evaluating short- and long-term effects. Int J Audiol 2005 May;44(5):255-264. [doi: 10.1080/14992020500060453] [Medline: 16028788]
34. Preminger JE, Yoo JK. Do group audiologic rehabilitation activities influence psychosocial outcomes? Am J Audiol 2010 Dec;19(2):109-125 [FREE Full text] [doi: 10.1044/1059-0889(2010/09-0027)] [Medline: 20595224]
35. Laplante-Lévesque A, Hickson L, Worrall L. Rehabilitation of older adults with hearing impairment: a critical review. J Aging Health 2010 Mar;22(2):143-153. [doi: 10.1177/0898264309352731] [Medline: 20056812]

36. Swanepoel DW, Clark JL, Koekemoer D, Hall JW, Krumm M, Ferrari DV, et al. Telehealth in audiology: the need and potential to reach underserved communities. Int J Audiol 2010 Mar;49(3):195-202. [doi: 10.3109/14992020903470783] [Medline: 20151929]
37. Paglialonga A, Cleveland Nielsen A, Ingo E, Barr C, Laplante-Lévesque A. eHealth and the hearing aid adult patient journey: a state-of-the-art review. Biomed Eng Online 2018 Jul 31;17(1):101 [FREE Full text] [doi: 10.1186/s12938-018-0531-3] [Medline: 30064497]
38. Meijerink JF, Pronk M, Paulissen B, Witte BI, Wouden BVD, Jansen V, et al. Effectiveness of an online SUpport PRogramme (SUPR) for older hearing aid users: study protocol for a cluster randomised controlled trial. BMJ Open 2017 Jun 20;7(5):e015012 [FREE Full text] [doi: 10.1136/bmjopen-2016-015012] [Medline: 28634259]
39. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
40. Meijerink JF, Pronk M, Kramer SE. Experiences With and Lessons Learned From Developing, Implementing, and Evaluating a Support Program for Older Hearing Aid Users and Their Communication Partners in the Hearing Aid Dispensing Setting. Am J Audiol 2020 Sep 18;29(3S):638-647. [doi: 10.1044/2020_AJA-19-00072] [Medline: 32946251]
41. Demorest ME, Erdman SA. Development of the communication profile for the hearing impaired. J Speech Hear Disord 1987 May;52(2):129-143. [doi: 10.1044/jshd.5202.129] [Medline: 3573744]
42. Mokkink LB, Knol DL, van Nispen RMA, Kramer SE. Improving the quality and applicability of the Dutch scales of the Communication Profile for the Hearing Impaired using item response theory. J Speech Lang Hear Res 2010 Jun;53(3):556-571. [doi: 10.1044/1092-4388(2010/09-0035)] [Medline: 20530379]
43. Mokkink LB, Knol DL, Zekveld AA, Goverts ST, Kramer SE. Factor structure and reliability of the Dutch version of seven scales of the Communication Profile for the Hearing Impaired (CPHI). J Speech Lang Hear Res 2009 Apr;52(2):454-464. [doi: 10.1044/1092-4388(2008/08-0036)] [Medline: 18952858]
44. Van den Brink RHS. Attitude and illness behaviour in hearing impaired elderly [Dissertation]. Groningen: University of Groningen; 1995.
45. West RL, Smith SL. Development of a hearing aid self-efficacy questionnaire. Int J Audiol 2007 Dec;46(12):759-771. [doi: 10.1080/14992020701545898] [Medline: 18049965]
46. Cox RM, Alexander GC. The International Outcome Inventory for Hearing Aids (IOI-HA): psychometric properties of the English version. Int J Audiol 2002 Jan;41(1):30-35. [doi: 10.3109/14992020209101309] [Medline: 12467367]
47. Kramer SE, Goverts ST, Dreschler WA, Boymans M, Festen JM. International Outcome Inventory for Hearing Aids (IOI-HA): results from The Netherlands. Int J Audiol 2002 Jan;41(1):36-41. [doi: 10.3109/14992020209101310] [Medline: 12467368]
48. Laplante-Lévesque A, Nielsen C, Jensen LD, Naylor G. Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport. J Am Acad Audiol 2014 Feb;25(2):187-198. [doi: 10.3766/jaaa.25.2.7] [Medline: 24828219]
49. Noble W. Extending the IOI to significant others and to non-hearing-aid-based interventions. Int J Audiol 2002 Jan;41(1):27-29. [doi: 10.3109/14992020209101308] [Medline: 12467366]
50. Hickson L, Worrall L, Scarinci N. Measuring outcomes of a communication program for older people with hearing impairment using the International Outcome Inventory. Int J Audiol 2006 Apr;45(4):238-246. [doi: 10.1080/14992020500429625] [Medline: 16684705]
51. Kramer SE, Kapteyn TS, Festen JM, Tobi H. Factors in subjective hearing disability. Audiology 1995;34(6):311-320. [doi: 10.3109/00206099509071921] [Medline: 8833311]
52. Meijer AGW, Wit HP, TenVergert EM, Albers FWJ, Muller Kobold JEP. Reliability and validity of the (modified) Amsterdam Inventory for Auditory Disability and Handicap. Int J Audiol 2003 Jun;42(4):220-226. [doi: 10.3109/14992020309101317] [Medline: 12790347]
53. Pronk M, Deeg DJH, Versfeld NJ, Heymans MW, Naylor G, Kramer SE. Predictors of entering a hearing aid evaluation period: a prospective study in older hearing-help seekers. Trends Hear 2017;21:2331216517744915 [FREE Full text] [doi: 10.1177/2331216517744915] [Medline: 29237333]
54. Laplante-Lévesque A, Hickson L, Worrall L. Stages of change in adults with acquired hearing impairment seeking help for the first time: application of the transtheoretical model in audiologic rehabilitation. Ear Hear 2013;34(4):447-457. [doi: 10.1097/AUD.0b013e3182772c49] [Medline: 23364333]
55. Demorest ME, Erdman SA. Retest stability of the communication profile for the hearing impaired. Ear Hear 1988 Oct;9(5):237-242. [doi: 10.1097/00003446-198810000-00002] [Medline: 3224770]
56. Wong LLN, Hickson L, McPherson B. Hearing aid satisfaction: what does research from the past 20 years say? Trends Amplif 2003;7(4):117-161 [FREE Full text] [doi: 10.1177/108471380300700402] [Medline: 15004650]
57. Laplante-Lévesque A, Jensen LD, Dawes P, Nielsen C. Optimal hearing aid use: focus groups with hearing aid clients and audiologists. Ear Hear 2013;34(2):193-202. [doi: 10.1097/AUD.0b013e31826a8ecd] [Medline: 23183046]
58. Bennett RJ, Laplante-Lévesque A, Meyer CJ, Eikelboom RH. Exploring hearing aid problems: perspectives of hearing aid owners and clinicians. Ear Hear 2018;39(1):172-187. [doi: 10.1097/AUD.0000000000000477] [Medline: 28787315]

59. Bennett RJ, Meyer CJ, Eikelboom RH, Atlas MD. Investigating the knowledge, skills, and tasks required for hearing aid management: perspectives of clinicians and hearing aid owners. Am J Audiol 2018 Mar 08;8(27):67-84. [doi: 10.1044/2017_AJA-17-0059] [Medline: 29305614]
60. Öberg M. A multicentre study evaluating the effects of the Swedish ACE programme. Int J Audiol 2017 Nov;56(11):876-886. [doi: 10.1080/14992027.2017.1346304] [Medline: 28685633]
61. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
62. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
63. Ferguson MA, Kitterick PT, Chong LY, Edmondson-Jones M, Barker F, Hoare DJ. Hearing aids for mild to moderate hearing loss in adults. Cochrane Database Syst Rev 2017 Sep 25;9:CD012023 [FREE Full text] [doi: 10.1002/14651858.CD012023.pub2] [Medline: 28944461]
64. Wason JMS, Stecher L, Mander AP. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? Trials 2014 Sep 17;17(15):364 [FREE Full text] [doi: 10.1186/1745-6215-15-364] [Medline: 25230772]
65. Vasilopoulos T, Morey TE, Dhatariya K, Rice MJ. Limitations of significance testing in clinical research: a review of multiple comparison corrections and effect size calculations with correlated measures. Anesth Analg 2016 Mar;122(3):825-830. [doi: 10.1213/ANE.0000000000001107] [Medline: 26891394]
66. Rice T, Schork N, Rao D. Methods for handling multiple testing. Adv Genet 2008;60:293-308. [doi: 10.1016/s0065-2660(07)00412-9] [Medline: 18358325]

Abbreviations

AIADH: Amsterdam Inventory for Auditory Disability and Handicap
CONSORT: Consolidated Standards of Reporting Trials
cRCT: cluster randomized controlled trial
CPHI: Communication Profile for the Hearing Impaired
HA: hearing aid
HHDI: Hearing Handicap and Disability Inventory
ICC: intracluster correlation coefficient
IOI: International Outcome Inventory
IOI-AI: International Outcome Inventory for Alternative Interventions
IOI-HA: International Outcome Inventory for Hearing Aids
ITT: intention to treat
MARS-HA: Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids
URICA-HL: University of Rhode Island Change Assessment adapted for hearing loss

Edited by G Eysenbach; submitted 22.01.20; peer-reviewed by C Bentley, C Meyer; comments to author 10.04.20; revised version received 26.05.20; accepted 14.06.20; published 22.09.20.

Please cite as:
Meijerink JFJ, Pronk M, Lissenberg-Witte BI, Jansen V, Kramer SE
Effectiveness of a Web-Based SUpport PRogram (SUPR) for Hearing Aid Users Aged 50+: Two-Arm, Cluster Randomized Controlled Trial
J Med Internet Res 2020;22(9):e17927
URL: http://www.jmir.org/2020/9/e17927/
doi:10.2196/17927
PMID:32960175

Original Paper

Understanding the Steps Toward Mobile Early Intervention for Mothers and Their Infants Exiting the Neonatal Intensive Care Unit: Descriptive Examination

Kathleen M Baggett[1], PhD; Betsy Davis[2], PhD; Susan H Landry[3], PhD; Edward G Feil[2], PhD; Anna Whaley[4], PhD; Alana Schnitz[5], PhD; Craig Leve[2], MS

[1]Mark Chaffin Center for Healthy Development, School of Public Health, Georgia State Universtiy, Atlanta, GA, United States

[2]Oregon Research Institute, Eugene, OR, United States

[3]University of Texas Health Sciences Center- Houston, Houston, TX, United States

[4]University of Kansas, Kansas City, KS, United States

[5]Juniper Gardens Children's Project, University of Kansas, Kansas City, KS, United States

Corresponding Author:
Kathleen M Baggett, PhD
Mark Chaffin Center for Healthy Development
School of Public Health
Georgia State Universtiy
14th Marietta Street NW, Suite 100&200
GA 30303
Atlanta, GA, 30303
United States
Phone: 1 4044131571
Email: kbaggett@gsu.edu

Abstract

Background: Neonatal intensive care unit (NICU) history, combined with systemic inequities for mothers of nondominant cultures and mothers who are socioeconomically disadvantaged, places infants at an extraordinary risk for poor developmental outcomes throughout life. Although receipt of early intervention (EI) is the best single predictor of developmental outcomes among children with and at risk for early developmental delays, mothers and infants with the greatest needs are least likely to receive EI. Mobile internet-based interventions afford substantial advantages for overcoming logistical challenges that often prevent mothers who are economically disadvantaged from accessing EI. However, the bridge from the NICU to a mobile internet intervention has been virtually unexplored.

Objective: This study aims to examine progression flow from NICU exit referral to an early mobile internet intervention to increase EI access and promote parent mediation of infant social-emotional and communication development.

Methods: Three NICUs serving the urban poor in a Midwestern city were provided support in establishing an electronic NICU exit referral mechanism into a randomized controlled trial of a mobile internet intervention for mothers and their infants. Measurement domains to reflect the bridge to service included each crucial gateway required for navigating the path into Part C EI, including referral, screening, assessment, and intervention access. An iterative process was used and documented to facilitate each NICU in establishing an individualized accountability plan for sharing referral materials with mothers before their NICU exit. Subsequent to the referral, progression flow was documented on the basis of a real-time electronic recording of service receipt and contact records. Mother and infant risk characteristics were also assessed. Descriptive analyses were conducted to summarize and characterize each measurement domain.

Results: NICU referral rates for EI were 3 to 4 times higher for open-shared versus closed-single gatekeeper referral processes. Of 86 referred dyads, 67 (78%) were screened, and of those screened, 51 (76%) were eligible for assessment. Of the 51 assessment-eligible mothers and infants, 35 dyads (69%) completed the assessment and 31 (89%) went on to complete at least one remote coaching intervention session. The dyads who accessed and engaged in intervention were racially and ethnically diverse and experiencing substantial adversity.

Conclusions: The transition from the NICU to home was fraught with missed opportunities for an EI referral. Beyond the referral, the most prominent reason for not participating in screening was that mothers could not be located after exiting the NICU. Stronger NICU referral mechanisms for EI are needed. It may be essential to initiate mobile interventions before exiting the NICU for maintaining post-NICU contact with some mothers. In contrast to a closed, single point of referral gatekeeper systems in NICUs, open, shared referral gating systems may be less stymied by individual service provider biases and disruptions.

(J Med Internet Res 2020;22(9):e18519) doi:10.2196/18519

KEYWORDS

early intervention; equity; NICU; low birthweight; transition; mobile internet intervention; infants; mothers

Introduction

Background

Low infant birth weight, requiring neonatal intensive care unit (NICU) treatment, places infants at high risk for a host of detrimental outcomes, including cognitive, language, and social delays and disabilities, which often persist into adulthood [1-3]. The discharge from a NICU is a stressful transition in which the responsibilities for around-the-clock care of a fragile infant shift from a NICU medical team to parents. This transition is particularly burdensome for mothers who often experience the bulk of responsibility for infant care [4]. A central challenge of this transition is to connect families with early intervention (EI) services.

Parent participation in EI services is the single best predictor of developmental outcomes for children with and at risk for developmental disabilities identified during the first year of life [5,6]. EI can support parents in recognizing and responding sensitively to subtle cues of infants with developmentally immature social signaling systems, which is foundational for establishing social interaction, feeding, and sleeping routines that promote infant regulation and social communication competency development [7]. By and large, however, the reach of EI services that target parent practices to support early social-emotional and communication competencies is extremely limited relative to societal needs [6,8,9]. Moreover, substantial inequities persist in EI access. Infants whose mothers are socioeconomically disadvantaged and of nondominant cultures are disproportionately over-represented in NICUs in the United States [10] because of historically driven systemic and structural inequities [11]. However, Part C service systems disproportionately serve White, middle-income, and upper-middle-income families [12].

In contrast to White infants in the United States, African American infants with special needs are five to eight times less likely to be referred for EI services [12]. They are also more likely to receive lower-quality care in NICUs because of both structural and interpersonal racism [11]. Structural barriers such as low-paying, unstable work with unpredictable hours without paid leave or quality childcare can deplete mothers' physical, psychological, and social resources for parenting a newborn in general and for engaging in the NICU in particular [13,14]. Implicit bias faced by these mothers within the medical system [15] can exacerbate maternal stress, shown to continue long after the NICU experience [16], and undermine mothers as central change agents of their infants' development and well-being. As social determinants of health, these inequities threaten infant development by compromising maternal functioning and parenting practices. Moreover, it is possible that these inequities lessen the willingness of mothers to connect with future intervention services [17], delay intervention access, and decrease mothers' opportunities to access and benefit from EI to improve infant developmental outcomes.

Recognizing the substantial impact of systematic gaps in access to timely intervention, Child Find efforts emphasize the need to identify, locate, and assess all infants with developmental delays, particularly those who are poor and of nondominant culture [6]. Unfortunately, published research on Child Find efforts, which are crucial for receiving EI services, tends to reflect a striking absence of representative samples of these families [6]. Published studies that systematically examine crucial junctures at which mothers either progress toward or fall off the pathway from NICU referral to EI access are also lacking. To obtain EI subsequent to NICU exit, mothers must successfully navigate crucial gateways that bridge the NICU experience to EI service receipt. These gateways include referral, screening, assessment, and intervention access [18], each of which must be navigated successfully to obtain EI. Early systemic barriers to intervention include failure to provide supported referrals, lack of routine developmental monitoring and screening, and insufficient reach of public awareness campaigns about the relevance of EI for infants and toddlers and their families [6,18]. The cost of home visiting intervention programs is particularly prohibitive because of state budget crises that often result in the rationing of state-funded home visiting services [9,19]. When home visiting programs are available, the barriers to parent engagement include unpredictable work schedules; shift work outside the 9 to 5 workday; transient housing; and living with relatives, friends, or landlords who are gatekeepers to the home and unamenable to home visits [20].

Mobile internet interventions, particularly those with remote coaching, afford substantial advantages for overcoming logistical challenges that often prevent mothers who are poor and of nondominant culture from accessing EI [20]. The advantages of mobile internet interventions include around-the-clock accessibility to program content, greater ease and flexibility in scheduling and rescheduling remote visits, less stigma, and greater parental autonomy to select and share video-recorded interactions at any time of day for the purpose of obtaining EI support [20]. Although advantages for mobile interventions are known to exist, the bridge from a NICU exit referral to a mobile internet intervention has been virtually unexamined in published studies.

Objectives

To increase equitable EI access, the purpose of this paper is to examine progression flow from a NICU exit referral to an early mobile internet intervention to promote parent mediation of infant social-emotional and communication competency development.

The progression flow on the bridge from the NICU to EI access is viewed within a randomized controlled pilot study of a mobile internet intervention with remote coaching designed to strengthen parent practices, which scaffold infant social-emotional and communication competencies. To illuminate the junctures at which mothers connect with or disconnect from progressing from the NICU to EI, we address the following questions: (1) What NICU referral structures impede or facilitate referral to intervention? For example, the diffusion literature suggests that when responsibility for action is shared across multiple members of a group, it can result in reduced outcome monitoring, a reduced sense of individual agency, and diffusion of responsibility and action [21]. (2) Among mothers referred for intervention, what is the screening rate, and what are the identified reasons for failure to screen? (3) Among mothers screened, what is the assessment completion rate, and what are the identified reasons for failure to assess? (4) Among mothers and infants assessed, what is the rate of intervention initiation and completion of the core intervention? and (5) What are the demographic and risk characteristics of mothers who traverse the bridge from NICU referral to intervention access? We hypothesize that when mothers are supported through each gateway on the bridge from the NICU to EI, the resultant internet-based intervention sample will be diverse with regard to demographic and risk characteristics.

Methods

Procedures

Our mobile internet intervention study procedures, from a NICU exit referral to a mobile intervention, provide a unique framework for examining progression flow through the crucial junctures that mirror the Part C EI system gateways: referral, screening, assessment, and intervention [18]. After institutional review board approval, recruitment efforts focused on 3 Level 3 NICUs serving the urban poor in a Midwestern city. These NICUs were selected because they were part of a medical conglomerate with similar characteristics that included a centralized geographic location in the urban core within 2 square miles of one another, similar annual admission rates, and a racially and ethnically diverse patient population, including those who lack insurance and the ability to pay. Through an iterative process, the research team conducted a series of meetings with each NICU team to generate a referral accountability plan, which was documented by the research team and provided to the NICU team for review and revision until the NICU team confirmed that their plan was complete and accurate. Each NICU-generated plan specified the NICU personnel who would share referral information with mothers, collect cards that mothers signed indicating their interest in being contacted by the study team, send electronic referrals to the research team, and respond to a biweekly prompt to provide an electronic referral update. Electronic referral update reports included the number of eligible mother-infant dyads in the NICU during the most recently completed referral period, the number of mothers with whom referral information was discussed, the outcome of each referral discussion, barriers to referral, and identified solutions. The research team provided referral materials to each NICU, which included service provider posters with mother and infant eligibility criteria to remind and prompt providers to refer all eligible mothers, a mobile intervention study letter to be shared by providers with mothers, a mother interest card for mothers to grant permission for study team follow-up, and a script for providers to use when sharing referral materials and collecting mother interest cards. An electronic NICU referral mechanism was established for NICU service providers to connect mothers and their low-birth-weight infants to a randomized controlled trial of a mobile internet intervention.

Referral criteria included biological or adoptive mothers, living in the metropolitan area of the NICU, who spoke English and whose infants at birth weighed <2500 g, were at least 24 weeks' gestational age, were no more than 5 months corrected gestational age at NICU exit, and who were not diagnosed with hydrocephalus, bronchopulmonary dysplasia, or beyond a grade 3 intraventricular hemorrhage. Referral criteria were established to avoid potential study burden for mothers of infants who were experiencing acute medical crises, including a high risk for NICU return or intensive care unit (ICU) entry. NICU teams were encouraged to refer all eligible mothers and infants in addition to any and all other service referrals such that all referred mothers were free to participate in existing community service referrals as usual without exclusion.

On receipt of each electronic referral, the research staff recorded the date of referral, referral source, and referral contact information into a project database. Research staff mailed consent forms to referred mothers and contacted them by phone to (1) confirm referral eligibility criteria, (2) review and discuss the consent form, and (3) determine whether mothers viewed themselves as able and willing to engage in the intervention study. Mothers who could not be reached by phone because of a disconnected number or failure to connect after at least five attempts were sent a letter encouraging mothers to contact the study team if interested in the program. Mothers' perceived ability to participate in the study was determined on the basis of their negative responses to a brief structured interview question in which mothers were first informed of personal situations that should be prioritized over intervention study participation, such as homelessness, shelter residence, inpatient mental health or substance abuse treatment, or a major physical or mental illness requiring intensive treatment such as schizophrenia, cancer, or HIV/AIDS. Mothers were then asked whether they were experiencing one or more of these situations or any other situation that could interfere with their ability to participate in the intervention study. An affirmative response was exclusionary and met the criteria for intervention study ineligibility. For mothers who were screened eligible and agreed to participate in the study, an in-home assessment visit was scheduled. All contact attempts, the outcome of each contact attempt, the eligibility screening outcome, and the scheduled assessment date were recorded in the project database.

Assessment

Informed consent was obtained at the onset of a 2-hour, in-home assessment visit. Electronic questionnaires were completed by mothers on the web via Qualtrics entry on an iPad (Apple Inc) to provide information about demographics and maternal and infant risk characteristics. The *Measurement Domains and Measures* Section provides a full description of the measurement domains and measures. Assessments were conducted by research assistants who had obtained at least a bachelor's degree in education, human development, or psychology and had at least 2 years of intervention research experience conducting in-home assessments and mobile intervention protocols with mothers and infants. Assessors were trained and observed to implement the assessment protocol with fidelity before data collection. The assessment details are also provided in the *Measurement Domains and Measures* section.

Mobile Intervention

Following assessment, mothers were randomized to 1 of 2 mobile internet interventions with identical structures. For both groups, the number of sessions and structural components of each session included (1) a web-based self-directed learning program through video-based teaching with check-in questions and provision of immediate corrective feedback, (2) an action plan outlining daily activity practice (homework) based on session content, (3) parent-recorded video and secure upload of session skill practice during interaction with her infant, and (4) a video-based coach call to coview the parent-recorded video of interaction with her infant [20]. For both intervention groups, meaningful access to a mobile internet intervention was operationalized as (1) mothers' completion of an in-home intervention session in which mothers were fully guided and scaffolded to interact with each mobile intervention component (ie, video modeling content, review questions, action plan, video creation, and coach call) and (2) mothers' completion of each of the above content components of the remote intervention session with on-demand scaffolding provided through messaging, phone, or video call.

At the in-home intervention orientation visit, all mothers were given an iPhone with unlimited data, text, and call plan. They were granted entry into a 12-session mobile internet intervention. Coaches used a demonstration video to introduce mothers to the mobile intervention, use the mobile phone features, and navigate through the first mobile intervention session. Coaches verbally scaffolded mothers' use of each session component by providing the phone and materials to the mother and serving as a guide on the side when mothers navigated through the entire first session, including the coach call procedures.

After the first session with coach guidance and full scaffolding, mothers autonomously completed the second intervention session with on-demand remote coach scaffolding between and during coach calls. The demand context for the coach response included (1) questions from the mother and (2) coach electronic monitoring of mothers' progress or nonprogress through intervention session components and feedback to celebrate mothers' successes and address barriers to progress. We expected that this meaningful access support in the first 2 sessions would increase the probability of mothers' continued progress in completing the remaining 10 remote sessions.

Measurement Domains and Measures

The measurement domains included the following: (1) challenges and solutions to referral of mothers and infants from the NICU, (2) mother and infant progression flow from the point of referral through the point of intervention access, and (3) maternal and infant demographics and risk characteristics. Measures pertaining to each domain are identified below.

Challenges and Solutions

Challenges and solutions to referral of mothers and infants from the NICU were documented by research assistants based on a review of biweekly NICU referral reports, follow-up discussion of barriers and solutions with NICU teams, and recorded events observed by the research team, such as changes in the NICU personnel, NICU referral strategy changes, and reported beliefs of the NICU personnel about referral.

Progression Flow Data

Progression flow data included records of electronic referrals and documentation of attempted screening calls, completed screening calls, screening outcomes, assessments scheduled, electronic recorded time stamp of assessment completion, intervention session completion, and recorded time of coach call completion.

Demographics

Demographic information included data on mother and infant age (years and months, respectively), ethnicity, race (based on federal reporting categories), mothers' educational level (multiple categories ranging from less than high school to postgraduate), no significant other relationship status, annual household income (6 categorical ranges), and number of children and adults in the household.

Maternal and Infant Characteristics

Maternal and infant risk characteristics included maternal financial strain and depression, infant time in the NICU, birth weight in ounces for calculation of very low birth weight status, months premature, corrected gestational age, infant social-emotional development concerns, and social-emotional behavior challenges (see Measures section below). Financial strain was measured using a 9-item questionnaire with a 5-point Likert-type scale for difficulty paying bills, money left over after paying bills, and money availability for necessities and other activities [22]. Maternal risk for postpartum depression was assessed using the Postpartum Depression Screening Scale (PDSS) [23]. The PDSS is a 35-item Likert-type self-report instrument that demonstrates strong sensitivity and specificity for postpartum depression in the 15 months after childbirth [24]. It has adequate psychometrics for mothers of infants in NICUs [25].

Infant social-emotional development was assessed using Ages & Stages Questionnaires: Social Emotional (ASQ-SE) screening tool [26]. This brief screener of social emotional functioning demonstrates high internal consistency with an overall α of .82 [27].

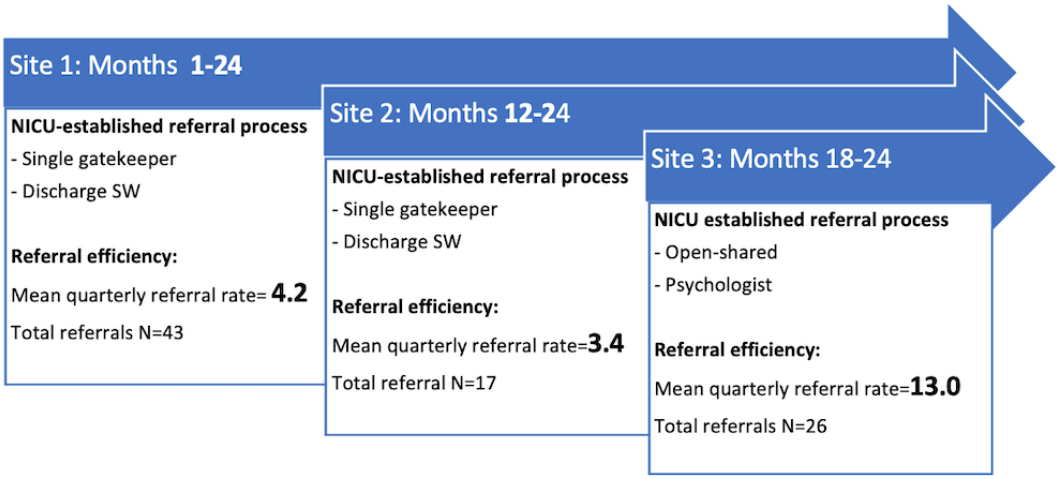
Infant social-emotional behavior concerns were assessed with the Devereux Early Childhood Assessment for Infants (DECA-I) scale [28]. The DECA-I is a 33-item behavior rating scale that assesses child protective factors central to the social and emotional health and resilience of infants from 1 to 18 months. This norm-referenced measure, with demonstrated adequate reliability and validity [29], provides a cutoff score for social-emotional behaviors in the concern range.

Results

To address the first research question, "What NICU referral approaches impede or facilitate referral from the NICU to a mobile EI?" we provide a brief description of the NICU site referral processes and referral rates as well as a summary of identified factors that impeded or facilitated referral. The established referral processes for each NICU site, as noted above, were identical with regard to the content of printed referral material for NICU personnel and mothers, inclusion and exclusion criteria for referral, electronic referral mechanism, and biweekly NICU referral prompting and reporting. However, there were systematic differences in the ways that each NICU self-selected to engage NICU personnel in their referral approach and to adhere to their established referral plan. Sites 1 and 2 established a single gatekeeper as a point of referral from the NICU to the intervention. In both cases, the gatekeeper was the social worker responsible for patient discharge. This individual also responded to biweekly prompts for site referral accountability reporting. Hence, we refer to Sites 1 and 2 as a closed, single gatekeeper approach. In contrast, Site 3 established an open, shared referral approach in which the NICU psychologist shared referral materials with all nursing and social work staff and encouraged conversations between staff and mothers about the referral materials, including whether mothers had seen the materials, what questions they had about the referral, if they had already expressed interest in referral to learn about the intervention, or would like to be referred to learn about the opportunity. The NICU psychologist also engaged in conversations with mothers about study referral and served as the contact for responding to biweekly accountability prompts and reporting.

The 3 NICU sites referred to a combined total of 86 mothers and their infants for mobile intervention, with Site 1 contributing 43 referrals (50% of the referral sample) within a 24-month period, Site 2 contributing 17 referrals (20% of the referral sample) within a 12-month period, and Site 3 contributing 26 referrals (30% of the referral sample) within a 6-month period. Sites were added sequentially such that the referral window length varied with a 24-month referral window for Site 1, a 12-month window for Site 2, and a 6-month referral window for Site 3. As noted above, the annual census of Level 3 NICUs was similar. Examination of mean quarterly referral rates in Figure 1 shows that Sites 1 and 2 had a much lower mean quarterly rate of 4.2 and 3.4 referrals per quarter, respectively, in contrast to Site 3 with 13 referrals per quarter.

Figure 1. Quarterly referral rates by NICU referral site. NICU: neonatal intensive care unit; SW: software.

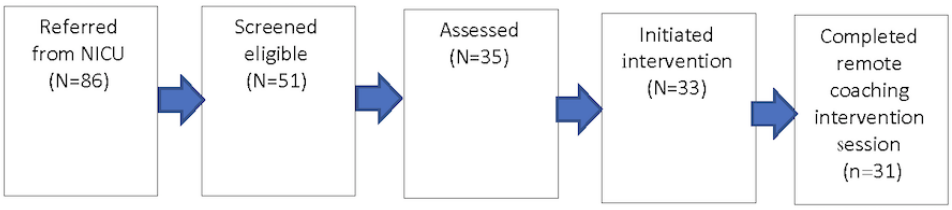


A review of the referral documentation yielded several factors associated with lower versus higher quarterly referral rates. These included the NICU self-selected referral approach and adherence to site-generated referral plans. In contrast to the open, shared referral structure, the closed single gatekeeper referral sites demonstrated the lowest quarterly referral rates and less adherence to their site-generated referral plan. Examples of challenges to adherence included the inability to refer as planned because of personnel issues such as illness, unpredictable staffing patterns, and staff turnover. Another challenge to adherence was provider belief that referral is best governed by provider clinical judgment rather than the principle of universal referral of all eligible infants and mothers. In contrast, the open, shared referral site demonstrated stronger adherence to the referral plan, which yielded the highest referral efficiency rates of the 3 sites.

Herein, we address research questions 2 to 4 pertaining to the junctures at which mothers and their infants either fall away from or progress through the crucial sequential gateways of screening, assessment, and intervention access on the path from referral from the NICU to early mobile intervention engagement. As displayed in Figure 2, 86 mothers referred from the NICU for mobile intervention, 67 (78%) were screened. Seven mothers (8%) declined to be screened and 12 mothers (14%) could not

be reached for complete phone screening. The most common reasons for failure to contact included incomplete or inaccurate referral contact information, phone disconnection, and returned mail. Of the 67 mothers screened, 51 (76%) were eligible for assessment. Of the 51 screened eligible, 10 mothers (20%) could either not be contacted to schedule an assessment or could not be assessed because of a family move or infant complications requiring ICU or other hospitalizations, 6 (12%) declined assessment, and 35 mothers (69%) and their infants initiated assessment.

Figure 2. Mean quarterly referral rate.



Of the 35 mothers who initiated the assessment, 1 mother died and another moved before initiating intervention. The remaining 33 mothers initiated intervention, with 31 mothers completing at least one remote coaching session. After initiating the intervention, reasons for failure to complete at least 1 remote session included a maternal death and 1 mother moving to an undisclosed location to escape neighborhood and partner violence. Of the 35 mothers assessed, 31 (89%) meaningfully accessed intervention by completing at least one intervention session with a remote coaching call. Examination of intervention access and engagement patterns showed that there were no significant differences between the intervention groups. On average, mothers completed 9.18 intervention sessions (all core content). The modal number of sessions completed was 12, which constituted all possible sessions.

Finally, we describe the demographic and risk characteristics of mothers and their infants who accessed the intervention. It was anticipated that when mothers were supported at each gateway, the resultant intervention sample would be ethnically and racially diverse. In addition, we expected the sample to reflect a high level of need relative to socioeconomic stressors and distress. We first examined demographic and risk characteristics by intervention group, which resulted in the finding of no significant between-group differences in any of the mother-infant demographic or risk characteristics. Hence, sample characteristics are presented for the combined intervention groups in Table 1. The sample was racially and ethnically diverse, with 55% (18/33) infants identified by their mothers as Black and 21% (7/33) identified as Latinx. The sample that accessed and engaged in the intervention was highly vulnerable. The majority of mothers were experiencing significant psychosocial stressors.

For the majority (81%), income was at <300% of the Federal Poverty Guideline, with 50% of the sample at or below 100% of the Federal Poverty Guideline. Relative to financial strain, 92% of mothers reported “not enough or barely enough” money left at the end of the month after paying bills. The majority did not have a college degree, and nearly half (43%) had not graduated from high school. More than half of the mothers were experiencing significant symptoms of depression and reported no significant other relationship. More than one-third of infants obtained scores classified in the concern range for social-emotional development and behavior on the ASQ-SE and the DECA-I Toddler, respectively.

Table 1. Sample demographics and risk characteristics of the mobile intervention access sample.

Variable	Value
Maternal	
Age (years), mean (SD); range	27.03 (5.49); 17.67 to 38.00
Race/ethnicity, %	
Black	52
Latinx	21
Maternal education (<college degree), %	88
Income, %	
≤100% federal poverty level	50
101%-300% federal poverty level	31
>300% federal poverty level	19
Relationship status (no significant other)	52
Significant depressive symptoms[a] (>PDSS[b] clinical cutoff)	58
Infant	
Gender (female), %	49
Race/ethnicity, %	
Black	55
Latinx	21
Birth weight	
Mean (SD); range, g	1859.04 (988.20); 510.29 to 4053.98
Extremely low, %	18
Very low, %	24
Low, %	39
Non-LBW[c] complication, %	18
Prematurity level (months), mean (SD); range	1.60 (1.13); −0.72 to 3.48
Time in NICU[d] (months), mean (SD); range	1.36 (1.15); 0.16 to 4.70
Chronological age at pre (months), mean (SD); range	4.01 (2.33); 0.45 to 10.32
Gestational age at pre (months), mean (SD); range	2.43 (2.26); −0.72 to 8.74
Social-emotional functioning[a]	39% ASQ-SE[e] developmental concern; 36% DECA[f] behavioral concern

[a]To establish significant symptoms of maternal depression, infant social-emotional developmental functioning concern, and behavioral concern, established clinical cutoff scores were used for the PDSS total depression score, ASQ-SE, and DECA, respectively.

[b]PDSS: Postpartum Depression Screening Scale.

[c]LBW: low birth weight.

[d]NICU: neonatal intensive care unit.

[e]ASQ-SE: Ages & Stages Questionnaires: Social Emotional.

[f]DECA: Devereux Early Childhood Assessment.

Discussion

The 3 referring Level 3 NICUs of similar annual census size and located within 2 square miles of one another demonstrated different levels of referral efficiency. In contrast to the closed, single gatekeeper referral approach of 2 NICUs, the shared, open referral approach of the third NICU resulted in higher referral efficiency. In addition, the shared, open NICU referral approach was associated with fewer reported disruptions in implementing their plan to discuss mobile intervention referral information with all eligible mothers. Within the closed single gatekeeper referral approach, the transition from the NICU to home was fraught with missed opportunities for EI referral. Not only were single gatekeeper NICUs more likely to report that their plans for referral were more often disrupted because of external factors such as unanticipated changes in staffing plans, but they also more often reported that they did not discuss referral information with mothers who they did not think would

be interested in referral. This suggests that having a closed single gatekeeper referral system may be more susceptible to the bias of a single person's judgment, which leads to missed opportunities for referral. In contrast, an open shared process involving multiple potential points of referral may afford more protection against individual bias that disrupts referral to EI. It is of interest to note that our approach to identify a NICU point of contact to interact with the study team and maintain oversight for the transfer of all internal NICU referrals was informed by the dissemination and implementation literature. This indicates the crucial role of identifying and engaging *champions* to support the establishment of implementation procedures [30]. Although each of our NICU points of contact self-identified as a *champion* of NICU referral into EI, only 1 of the 3 NICU points of contact provided an operational demonstration of *championing* as evidenced by actionable activities such as (1) communicating a shared responsibility of all NICU team members to engage in conversations with mothers about the importance of EI and (2) encouraging repeated and redundant opportunities for mothers to consider their own readiness to act on a referral for EI. It is likely that this type of operationalization of champions may play a crucial role in protection against the diffusion of responsibility within open, shared referral gating systems.

After referral, the most common reason for mothers and their infants to fall off the path toward intervention was that they could not be contacted after leaving the NICU. Hence, for some mothers, it may be important to conduct screening and assessment in the NICU to establish mobile intervention contact before the transition home. Several factors likely contributed to the inability to contact mothers after their transition home. In addition to the most commonly documented reasons, which included incomplete referral contact information and family mobility, another factor likely to have interfered with contact was an exacerbation of maternal distress that may have interfered with their ability to respond to contact attempts. The transition from the NICU to home is well documented as a time of heightened distress above and beyond the notable stress of NICU experience for many mothers [4]. To promote engagement in EI, it may be important for some mothers to establish supportive intervention contact, which can buffer against transition stress before the transition from the NICU to home.

Most screened eligible mothers and their infants (69%) selected to participate in and engaged in assessment, and 89% of those assessed went on to meaningfully engage in the mobile intervention. The fact that mothers completed, on average, at least 9 sessions, constituting all the core content, and that the modal pattern was the completion of all 12 content sessions is noteworthy. In contrast, home visiting studies of parenting interventions have consistently documented the concerning finding that, on average, approximately half of intended mothers receive any intervention and that the average amount of intervention received is, on average, only 25% of what was intended [29]. Intervention initiation and engagement in our mobile intervention sample were substantially higher. Moreover, this mobile intervention sample was racially and ethnically diverse and experienced significant psychosocial stressors. Hence, it is possible to engage mothers of nondominant culture and their infants who are experiencing a host of psychosocial stressors in a mobile EI program. However, there's a need to establish stronger NICU referral mechanisms to EI.

Limitations

The limitations of this research include a small convenience sample restricted to descriptive methods. In this small convenience sample, documented barriers to referral pertained to NICU referral characteristics. We did not have access to the NICU-level data about potential infant factors that could have influenced NICU health provider referral and/or parental acceptance of referral to EI. Our examination of maternal responses, in terms of moving toward or falling away from the path to intervention access, was conducted within an intervention study wherein resources, including ongoing training and support, were consistently applied to reinforce intervention research staff to prioritize and sustain outreach efforts in the face of substantial maternal and infant adverse experiences. Part C EI programs, especially with regard to resources for ongoing staff training and support, are often strained. Hence, the transferability of effective outreach strategies to facilitate maternal movement from referral to intervention access must take into account resource differences and work toward efforts to increase training and support resources within Part C EI programs if we are to succeed in reaching mothers and infants most in need.

Implications for Future Research

Future studies should include NICU patient population-level data to examine infant characteristics that may be associated with referral provision and referral acceptance. To elucidate solutions for overcoming referral barriers within the NICU, future research needs to be conducted within NICUs to determine what factors in an open-gating system are associated with higher rates of referral such that these can be experimentally implemented and studied to increase effective and efficient referral practices. Subsequent research needs to be conducted with larger samples of NICUs to explicate characteristics of intervention referral champions and their operational execution of engaging NICU teams in processes that promote universal referral, characterized by broad, repeated, and redundant contact opportunities for referral. Dissemination and implementation of best practices identified from such research are crucial for improving equitable referrals such that all parents with infants in the NICU are provided opportunities to enter the first gateway on the path to accessing needed intervention, regardless of race, ethnicity, and income. Beyond NICU referral optimization, the resource infrastructure within Part C EI programs warrants closer examination with regard to the mechanisms that optimize or jeopardize family engagement at every crucial juncture on the pathway from referral to intervention access.

Acknowledgments

The authors would like to acknowledge the support provided for this research by the Health Resources and Services Administration (HRSA) of the US Department of Health and Human Services (HHS) and the R40 Maternal and Child Health Field-initiated Innovative Research Studies Program under grant number R40MC26822. The content and conclusions herein are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, or the US government.

Conflicts of Interest

SL developed the Play and Learning Strategies intervention program. KB, BD, EF, and SL are the developers of the InfantNet program, the original intervention platform on which the ePALS Mom and Baby Net for the NICU program is based.

References

1. Ritchie K, Bora S, Woodward L. Peer relationship outcomes of school-age children born very preterm. J Pediatr 2018 Oct;201:238-244. [doi: 10.1016/j.jpeds.2018.05.034] [Medline: 29958672]
2. Choi KH, Martinson ML. The relationship between low birthweight and childhood health: disparities by race, ethnicity, and national origin. Ann Epidemiol 2018 Oct;28(10):704-9.e4. [doi: 10.1016/j.annepidem.2018.08.001] [Medline: 30172559]
3. Aarnoudse-Moens C, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. Pediatrics 2009 Aug;124(2):717-728. [doi: 10.1542/peds.2008-2816] [Medline: 19651588]
4. Enlow E, Faherty L, Wallace-Keeshen S, Martin A, Shea J, Lorch S. Perspectives of low socioeconomic status mothers of premature infants. Pediatrics 2017 Mar;139(3) [FREE Full text] [doi: 10.1542/peds.2016-2310] [Medline: 28223372]
5. Landa RJ. Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. Int Rev Psychiatry 2018 Feb;30(1):25-39 [FREE Full text] [doi: 10.1080/09540261.2018.1432574] [Medline: 29537331]
6. McManus BM, Richardson Z, Schenkman M, Murphy N, Morrato EH. Timing and intensity of early intervention service use and outcomes among a safety-net population of children. JAMA Netw Open 2019 Jan 4;2(1):e187529 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7529] [Medline: 30681716]
7. Landry SH, Smith KE, Swank PR, Guttentag C. A responsive parenting intervention: the optimal timing across early childhood for impacting maternal behaviors and child outcomes. Dev Psychol 2008 Sep;44(5):1335-1353 [FREE Full text] [doi: 10.1037/a0013030] [Medline: 18793067]
8. Tang BG, Feldman HM, Huffman LC, Kagawa KJ, Gould JB. Missed opportunities in the referral of high-risk infants to early intervention. Pediatrics 2012 Jun;129(6):1027-1034 [FREE Full text] [doi: 10.1542/peds.2011-2720] [Medline: 22614772]
9. Rosenberg S, Robinson C, Shaw E, Ellison M. Part C early intervention for infants and toddlers: percentage eligible versus served. Pediatrics 2013 Jan;131(1):38-46. [doi: 10.1542/peds.2012-1662] [Medline: 23266922]
10. Giurgescu C, Engeland CG, Zenk SN, Kavanaugh K. Stress, inflammation and preterm birth in African American women. Newborn Infant Nurs Rev 2013 Dec;13(4):171-177. [doi: 10.1053/j.nainr.2013.09.004]
11. Collins JW, David RJ, Handler A, Wall S, Andes S. Very low birthweight in African American infants: the role of maternal exposure to interpersonal racial discrimination. Am J Public Health 2004 Dec;94(12):2132-2138. [doi: 10.2105/ajph.94.12.2132] [Medline: 15569965]
12. Feinberg E, Silverstein M, Donahue S, Bliss R. The impact of race on participation in part C early intervention services. J Dev Behav Pediatr 2011 May;32(4):284-291 [FREE Full text] [doi: 10.1097/DBP.0b013e3182142fbd] [Medline: 21389865]
13. Horbar JD, Edwards EM, Greenberg LT, Profit J, Draper D, Helkey D, et al. Racial segregation and inequality in the neonatal intensive care unit for very low-birth-weight and very preterm infants. JAMA Pediatr 2019 May 1;173(5):455-461 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.0241] [Medline: 30907924]
14. Holditch-Davis D, Santos H, Levy J, White-Traut R, O'Shea TM, Geraldo V, et al. Patterns of psychological distress in mothers of preterm infants. Infant Behav Dev 2015 Nov;41:154-163 [FREE Full text] [doi: 10.1016/j.infbeh.2015.10.004] [Medline: 26495909]
15. Penner A, Dovidio F, West V, Gaertner L, Albrecht L, Dailey K, et al. Aversive racism and medical interactions with black patients: a field study. J Exp Soc Psychol 2010 Mar 1;46(2):436-440 [FREE Full text] [doi: 10.1016/j.jesp.2009.11.004] [Medline: 20228874]
16. Garel M, Dardennes M, Blondel B. Mothers' psychological distress 1 year after very preterm childbirth. Results of the EPIPAGE qualitative study. Child Care Health Dev 2007 Mar;33(2):137-143. [doi: 10.1111/j.1365-2214.2006.00663.x] [Medline: 17291317]
17. Blair K, Lee I, Cho S, Dunlap G. Positive behavior support through family–school collaboration for young children with autism. Top Early Child Spec Educ 2010 Aug 6;31(1):22-36 [FREE Full text] [doi: 10.1177/0271121410377510]
18. Barger B, Rice C, Simmons CA, Wolf R. A systematic review of part c early identification studies. Topics Early Child Spec Educ 2018 May;38(1):4-16 [FREE Full text] [doi: 10.1177/0271121416678664] [Medline: 28479651]

19. Roberts H, Needelman H, Jackson B, McMorris C, Munyon A. Effect of community size on eligibility for early intervention for infants with a neonatal intensive care experience. J Rural Health 2014;30(3):259-264. [doi: 10.1111/jrh.12055] [Medline: 24330261]
20. Feil E, Baggett K, Davis B, Landry S, Sheeber L, Leve C, et al. Randomized control trial of an internet-based parenting intervention for mothers of infants. Early Child Res Q 2020;50:36-44 [FREE Full text] [doi: 10.1016/j.ecresq.2018.11.003] [Medline: 32863565]
21. Beyer F, Sidarus N, Bonicalzi S, Haggard P. Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring. Soc Cogn Affect Neurosci 2017 Jan 1;12(1):138-145 [FREE Full text] [doi: 10.1093/scan/nsw160] [Medline: 27803288]
22. Conger R, Elder JG. Families in troubled times: the Iowa youth and families project. In: Families in Troubled Times: Adapting to Change in Rural America. New York, USA: Aldine de Gruyter; 1994:3-19.
23. Beck C, Gable R. Postpartum depression screening scale (PDSS). Los Angeles: Western Psychological Services 2002. [doi: 10.1037/t42726-000]
24. Vogeli M, Hooker A, Everhart D, Kaplan S. Psychometric properties of the postpartum depression screening scale beyond the postpartum period. Res Nurs Health 2018 Apr;41(2):185-194. [doi: 10.1002/nur.21861] [Medline: 29603768]
25. McCabe K, Blucker R, Gillaspy JJ, Cherry A, Mignogna M, Roddenberry A, et al. Reliability of the postpartum depression screening scale in the neonatal intensive care unit. Nurs Res 2012;61(6):441-445. [doi: 10.1097/NNR.0b013e318268d06c] [Medline: 22960585]
26. Squires J, Bricker D, Twombly E. Parent-completed screening for social emotional problems in young children: the effects of risk/disability status and gender on performance. Infant Ment Health J 2003 Jan;25(1):62-73 [FREE Full text] [doi: 10.1002/imhj.10084]
27. Yovanoff P, Squires J. Determining cutoff scores on a developmental screening measure: use of receiver operating characteristics and item response theory. J Early Interv 2016 Jul 25;29(1):48-62 [FREE Full text] [doi: 10.1177/105381510602900104]
28. MIECHV Issue Brief on Family Enrollment and Engagement. Maternal and Child Health Bureau - HRSA. URL: https://mchb.hrsa.gov/sites/default/files/mchb/MaternalChildHealthInitiatives/HomeVisiting/tafiles/enrollmentandengagement.pdf [accessed 2020-07-07]
29. Miech E, Rattray N, Flanagan M, Damschroder L, Schmid A, Damush T. Inside help: an integrative review of champions in healthcare-related implementation. SAGE Open Med 2018;6:2050312118773261 [FREE Full text] [doi: 10.1177/2050312118773261] [Medline: 29796266]
30. Powell G, Mackrain M, LeBuffe P, Lewisville N. Devereux Early Childhood Assessment for Infants and Toddlers Technical Manual. Lewisville, NC: Kaplan Early Learning Corporation; 2007.

Abbreviations

ASQ-SE: Ages & Stages Questionnaires: Social Emotional
DECA-I: Devereux Early Childhood Assessment for Infants
EI: early intervention
HHS: Health and Human Services
HRSA: Health Resources and Services Administration
NICU: neonatal intensive care unit
PDSS: Postpartum Depression Screening Scale

Edited by G Eysenbach; submitted 03.03.20; peer-reviewed by Y Lee, Y Hirshler ; comments to author 27.04.20; revised version received 13.07.20; accepted 26.07.20; published 22.09.20.

Please cite as:
Baggett KM, Davis B, Landry SH, Feil EG, Whaley A, Schnitz A, Leve C
Understanding the Steps Toward Mobile Early Intervention for Mothers and Their Infants Exiting the Neonatal Intensive Care Unit: Descriptive Examination
J Med Internet Res 2020;22(9):e18519
URL: http://www.jmir.org/2020/9/e18519/
doi:10.2196/18519
PMID:32960178

Original Paper

Exploring Participants' Experiences of a Web-Based Program for Bulimia and Binge Eating Disorder: Qualitative Study

See Heng Yim[1], BSc; Emma Bailey[1], BSc, MSc; Gemma Gordon[1], BA, MSc; Nina Grant[2], BSc, MSc, PhD, DClinPsy; Peter Musiat[1], PhD; Ulrike Schmidt[1,2], FRCPsych, PhD, MD

[1]King's College London, London, United Kingdom

[2]South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
See Heng Yim, BSc
King's College London
Institute of Psychiatry, Psychology and Neuroscience
PO Box 59
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 2078480160
Email: see.yim@hmc.ox.ac.uk

Abstract

Background: Guided cognitive behavioral self-help is a recommended first-line treatment for eating disorders (EDs) such as bulimia nervosa (BN) or binge eating disorder (BED). Online versions of such self-help programs are increasingly being studied in randomized controlled trials (RCTs), with some evidence that they can reduce ED symptoms, although intervention dropout is variable across interventions. However, in-depth research into participants' experiences and views on the acceptability of web-based interventions is limited.

Objective: This is a qualitative process study of participants' experiences of everyBody Plus, a web-based cognitive behavioral intervention, integrated into a large RCT to aid the interpretation of the main trial's results. To our knowledge, this is the first such study in digital intervention for EDs research to include real-time feedback into the qualitative analysis. This study aims to build upon the emerging literature by qualitatively exploring participants' experiences of a web-based intervention for BN and BED.

Methods: Participants were those who took part in the UK arm of a larger RCT investigating the efficacy of the everyBody Plus intervention. Reflexive thematic analysis was completed on 2 sources of data from the online platform: real-time feedback quotes provided at the end of completing a module on the platform (N=104) and semistructured telephone interview transcripts (n=12).

Results: Four main themes were identified. The first theme identified positive and negative user experiences, with a desire for a more customized and personalized intervention. Another theme positively reflected on how flexible and easy the intervention was to embed into daily life, compared with the silo of face-to-face therapy. The third theme identified how the intervention had a holistic impact cognitively, emotionally, interpersonally, and behaviorally. The final theme was related to how the intervention was not a one size fits all and how the perceived usefulness and relevance were often dependent on participants' demographic and clinical characteristics.

Conclusions: Overall, participants reported positive experiences with the use of the everyBody Plus web-based intervention, including flexibility of use and the potential to holistically impact people's lives. The participants also provided valuable suggestions for how similar future web-based interventions could be improved and, in the context of EDs, how programs can be designed to be more inclusive of people by encompassing different demographic and clinical characteristics.

(J Med Internet Res 2020;22(9):e17880) doi:10.2196/17880

KEYWORDS

eHealth; self-help; eating disorders; bulimia; binge eating disorder; internet-based intervention; qualitative research

Introduction

Background

In many countries, the digitalization of health care services is a key strategic objective. For example, in England, the National Health Service (NHS) Long Term Plan emphasizes the need to make better use of data and digital technology in the NHS and to improve access to digital tools and services [1]. Recommendations to enable NHS staff to make the most of such innovative technologies are made in a recent independent report and a supplementary report on the *Digital Future of Mental Healthcare* [2]. In this context, it was argued that digital therapies could provide evidence-based stand-alone self-help or combined mental health interventions for service users.

In eating disorders (EDs), a growing number of studies have assessed the efficacy of eHealth and mobile health (mHealth) interventions [3-5], especially the use of structured cognitive behavioral online self-help interventions for individuals with bulimia nervosa (BN) or binge eating disorder (BED) [6]. There is some evidence that such interventions are able to reduce ED symptoms compared with waiting-list control, but comparisons with more traditional book-based self-help or face-to-face therapy are as yet relatively rare [7].

To understand people's experiences of utilizing such web-based interventions more fully, qualitative data can be used to inform the design of complex interventions [8]. However, to date, few such studies are available. A systematic review and metasynthesis of self-help interventions for EDs [9] identified only four studies that used qualitative methodology to understand people's experience of web-based interventions. Through meta-ethnography, six concepts related to users' experiences of the programs were synthesized. Intervention-related factors included anonymity and privacy, accessibility and flexibility, and guidance. User-related factors included agency/autonomy, self-motivation, and expectation/attitude. These revealed some unique advantages of computer-based interventions, namely, the neutrality and the *machine-like* properties of the computer that shield the participants from other users and their online therapists or coaches. This is in contrast to potentially emotionally *hotter* face-to-face therapy, where patients might feel judged. There was a sense of increased fluidity as to where and when users could access the intervention, which required greater motivation. Health care professionals were seen as a guide, coach, or facilitator rather than a therapist. Some users viewed web-based interventions as a first step toward recovery, which might need to be supplemented with face-to-face therapy. An additional study on the views of people with EDs on online self-help interventions agrees with this point (Yim et al, unpublished data, 2020).

The Technology Acceptance Model is a framework to help understand users' adoption and acceptance of information technology. This model posits that a potential user's intention to use a technology and their actual usage behavior is based on the perceived usefulness and ease of use [10]. These factors have not been examined in depth in most clinical trials that focus on the efficacy of web-based interventions in EDs. Qualitative studies, especially if integrated into large-scale randomized controlled trials (RCTs), can enhance our understanding of any contextual factors as well as facilitators and barriers that influence an intervention's acceptability, efficacy, and scalability [11]. However, in many RCTs that report on qualitative process data, these are published only after study outcomes are known, which has the potential for confirmation bias when interpreting the process data [12].

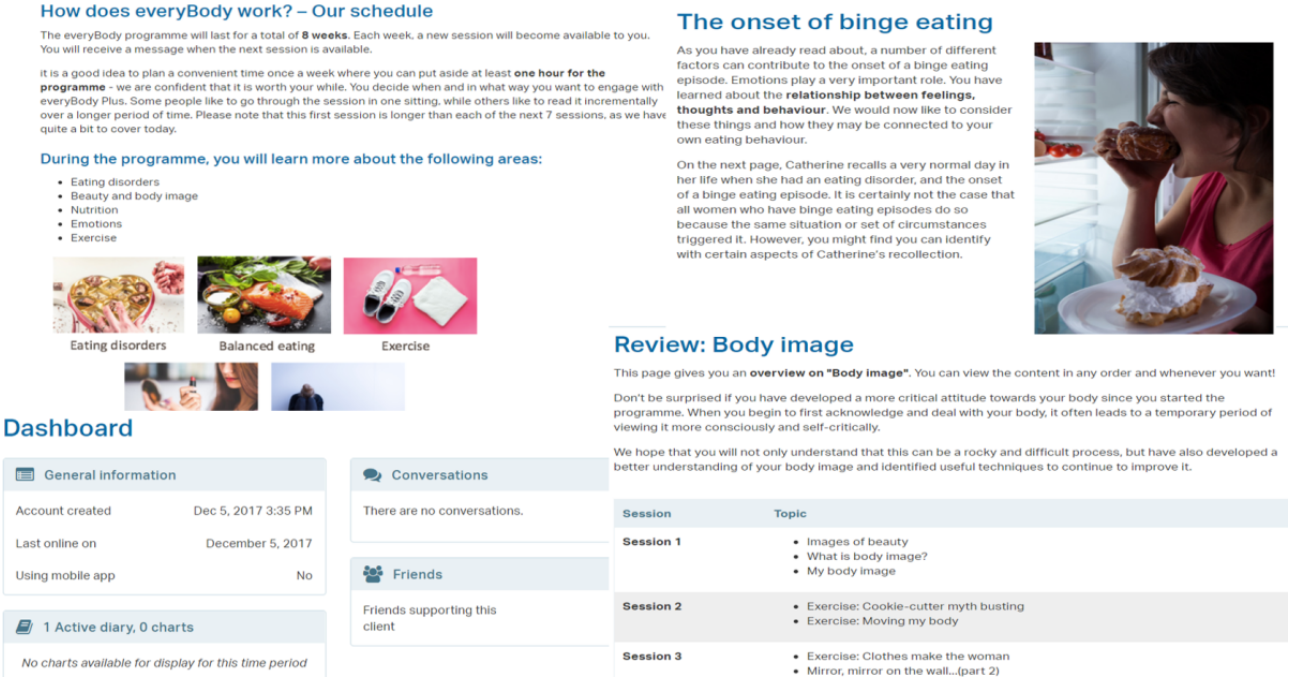
Objectives

This study is a qualitative process evaluation of an ongoing pragmatic two-country (Germany and the United Kingdom), multisite RCT that examines the efficacy of an 8-session, guided, internet-based cognitive behavioral intervention (everyBody Plus) in adult women with BN, BED, and other specified feeding or eating disorder with binge eating (International Standard Randomised Controlled Trial Number [ISRCTN] Registry number: 12608780). Figure 1 shows a collage consisting of screenshots of the intervention. The intervention was administered to individuals who were seeking treatment or currently waiting for face-to-face outpatient treatment, aiming to bridge the waiting time. The UK arm of the intervention was developed to closely match the German program equivalent, which had been adapted from Student Bodies before the translation and editing of the English trial intervention content. Student Bodies was previously shown to be effective in reducing ED symptoms in RCTs of young women with subthreshold eating disorders [13,14]. The high usability of Student Bodies for EDs had previously been demonstrated in a mixed methods study of 9 users with usability ratings of 83.1 out of 100 [15]. In the UK trial, the main adaptations included updates on the layout of the program, replacing lengthy text passages with explanatory videos, and including written and audio testimonials of fictitious participants. Each session took approximately 1 hour to complete. Further details on the intervention and the trial can be found in the protocol [16]. The expected follow-up completion was in May 2020, followed by quantitative analysis of the RCT. Importantly, the process data are reported before completion, as recommended by Oakley et al [12].

The aims of this study are as follows:

- Explore participants' experiences of the everyBody Plus web-based intervention
- Add to and complement the future quantitative RCT findings
- Add to the emerging literature on people's experiences of web-based interventions for BN and BED

Figure 1. A collage of screenshots of the everyBody Plus intervention.



Methods

The reporting of this study followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) [17].

Participants and Procedure

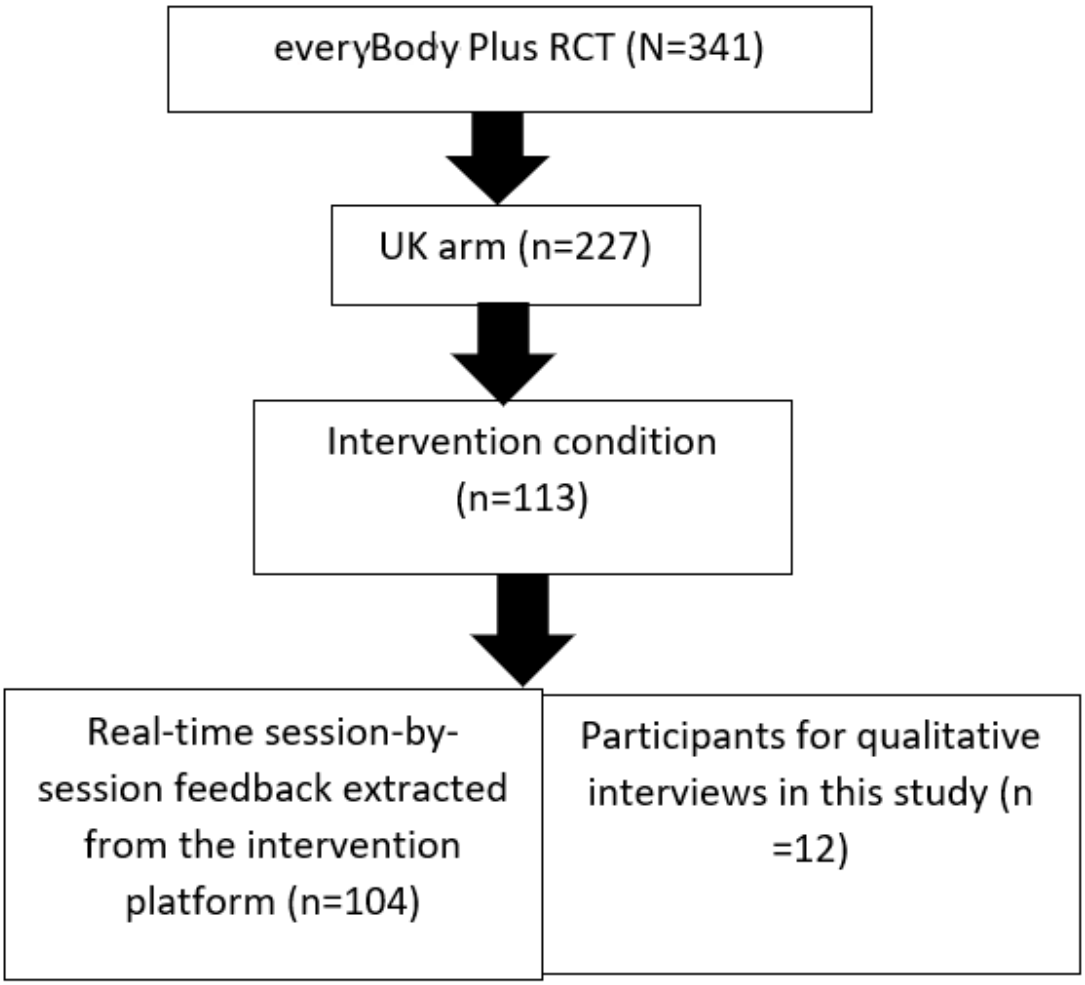
A total of 341 adult females were recruited into the larger RCT (ISRCTN Registry number: 12608780) from June 2016 to May 2019. Of these, 227 were from the United Kingdom. The mean age of these participants was 30.7 years (SD 10.8), and their mean self-reported BMI was 31.4 kg/m^2 (SD 13.8). The UK participants were recruited through 15 NHS Foundation Trusts, through national and regional ED charities (Beat, South Yorkshire Eating Disorders Association, and National Centre for Eating Disorders), and through King's College London email circulars and social media and word of mouth (for details of the RCT inclusion and exclusion criteria, please refer to the study by Vollert et al [16]). All the UK participants (N=113) allocated to the intervention condition were eligible to participate in this study in 2 different ways. First, if they had completed at least one module on the platform (n=104), their real-time feedback quotes were extracted from the platform. At the end of each module, participants were asked "Did you find this session useful?" and "We would love to hear your opinion of this session, both positive and negative comments!" where they could provide feedback. A further avenue for feedback was the group forum under the thread *Session 8: Review.* Informed consent was obtained before the commencement of their participation in the RCT. Ethical approval for the RCT was obtained from the UK Health Research Authority (reference no. 16/NW/0888).

In addition, all 113 UK participants allocated to the intervention were invited to take part in a telephone interview to discuss their experiences about the intervention via email and messaging on the online platform, where further informed consent was obtained from those who decided to participate. The analysis involved UK participants only, as the language differences hindered the ability to conduct an integrated analysis.

A total of 12 participants (aged between 21 and 60 years; mean 42.2, SD 13.7; mean BMI=32.7, range 19.9-50.6) consented and took part between June and August 2018. Those who did not take part did not provide any reason. Moreover, 9 interview participants completed all 8 intervention modules, 1 completed 7 modules, and the remaining 2 completed 5 and 6 modules. For clarity, Figure 2 shows the participant breakdown of this study within the larger RCT.

A semistructured interview guide was devised for the purpose of the study, which included questions on expectations, acceptability, content, and user experiences (UX; Multimedia Appendix 1). The topic guide was devised with reference to similar studies [18] and the Technology Acceptance Model [10]. Interviews were conducted over the phone by the first author SY, and the duration ranged from 20 to 60 min. No incentive was given for interview participation. Interviews were audio recorded with the participants' permission. Field notes were made during the conversations to include the interviewer's thoughts but were not used in the analysis.

Figure 2. Participant and data breakdown in the current study. RCT: randomized controlled trial.



Data Analysis

Real-time participants' quotes on the intervention platform and interview transcripts were transcribed verbatim by the first author SY. The two sets of data were combined and analyzed together and not linked by participants. The data were entered into NVivo software (QSR International, version 12) for analysis. Although the real-time feedback offered a naturalistic perspective of participants' experiences of the intervention with minimal influence of the study team, the interview transcripts allowed a more in-depth, elaborate perspective for analysis.

Reflexive thematic analysis (TA) was chosen for analysis [19] because it is theoretically flexible and is useful for a relatively large set of data. The analysis broadly followed 6-step process by Braun and Clarke (2013) [20]: SY led the data extraction and transcription. First, SY familiarized herself with the data. The data were coded inductively with a mix of latent and semantic coding. Initial themes were generated after coding all transcripts, following which the codes were refined. A thematic framework was then developed and discussed with author US with illustrative examples for clarification and refinement. A critical realist approach was adopted [19], based on the idea that a real world exists but meanings are constructed and influenced by context. Although both in-session feedback and interviews rendered a relatively large sample for qualitative research, we did not aim for data saturation or complete analysis of data but instead focused on conceptual rigor [21]. We also did not include frequencies or percentages in our analysis following the guidance from Braun and Clarke [22], as the frequency of similar responses might not determine the value or relevance to the research question. Similarly, the frequency count might not be appropriate, as we could not assume the meaning of the absence of certain themes among some participants' responses. As this approach is regarded as a *big Q* philosophically and procedurally, in contrast with *small q* versions of TA [23], multiple coders are not necessary in this approach.

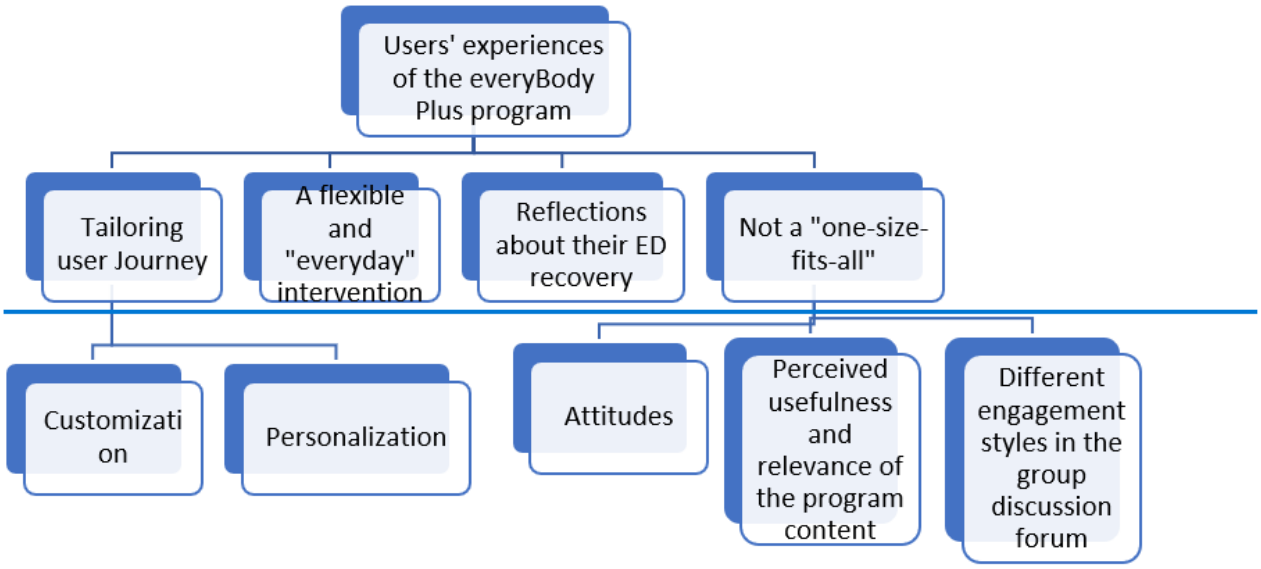
Service User Involvement

All 113 participants were invited to review the paper through the intervention platform. Two participants expressed interest in reviewing. The paper was sent via encrypted email, and the feedback was positive in general. They indicated that the paper was easy to read. The results and limitations sections were subsequently modified in response to the feedback. For example, the timing of the process study was noted as a limitation for not being able to capture how the intervention facilitated or hindered the subsequent face-to-face therapy.

Reflexivity

The first author, SY, was aware of her multiple roles in the trial as a study coordinator, as one of the online therapists, and as an interviewer. Although this might help develop trust and closeness, as some participants had electronic contact with her during the trial, this relationship might bring bias when interviewing participants. To reduce bias, possible themes and preconceptions were discussed with the wider team. Author EB conducted interviews with the participants for whom SY was assigned as their online therapist.

Figure 3. Coding tree. ED: eating disorder.



Results

Overview of Themes

Four themes were identified, including (1) tailoring user journey, (2) a flexible and *everyday* intervention, (3) reflections about their ED recovery, and (4) *not a one size fits all* (Figure 3). We describe these themes using the illustrative samples below. Please see Multimedia Appendix 2 for a detailed and comprehensive set of quotes.

Tailoring User Journey

This theme was closely related to the design and UX of web-based interventions in which participants expressed a strong desire for a more tailored, personalized user journey. A varied usage pattern was shown: although some participants used the program whenever they had time, on the go, or whenever they received a reminder, other participants went through the program at dedicated times.

The participants had a mixed UX. Some considered the interface easy to navigate and technical issues were minimal, whereas others pointed out the less user-friendly aspects. For example, one participant said on the group discussion forum:

> *I also found the UX really bad. It's like eating a lovely biscuit that has sand in it. It doesn't matter how lovely the biscuit is, the sand ruins it.* [Group forum, GF]

Sessions were seen as too lengthy (each session could take around an hour to finish), especially for those with poor concentration. For some, self-identified perfectionist tendencies meant that the tasks became "stressful to try to absorb" or even overwhelming to cope with.

Two subthemes were identified: (1) customization and (2) personalization. Customization allows users to adapt features according to their preferences. One of the features was notifications or reminders. In general, participants welcomed the email notifications reminding them, for example, to complete self-monitoring diaries or be notified when a new message was received. Nevertheless, participants seemed to prefer to be given the option to adjust the frequency (eg, daily or weekly) and the type of notifications (eg, text or email) to suit their needs:

> *(The) programme should fit around the user more, rather than the user having to fit around the programme.* [GF]

Currently, the program allows the adjustment of the frequency of the reminders and has the option to switch off the notifications. However, there was no SMS notification, and participants could not adjust the receiving time of the notifications.

Personalization was another common factor influencing the UX. Participants appreciated the use of multimedia such as text, pictures, and audio to suit different learning styles. In addition, many participants wished for a bookmark or personal route functionality, where they could immediately access the page they last visited, as the sessions were long and too cognitively demanding to finish in one go. A participant contended:

> *I don't like about the user-friendliness is that if I close the page partway through, even if I do an intermediate save, next time I open the session in a new window I must start from the beginning and click through all the pages to where I previously was.* [In-session feedback, SF]

Participants showed different preferences regarding the type of technical device used. Different motivational factors were at play when using mobile phones or laptops. Mobile phones seemed to be suited for doing the program while *on the go*:

> *When I am out, I don't take my laptop with me...my phone, I will just bring my phone...that I can just log in and do the main thing, like the diary, symptom tracker, get hold of people, so it does definitely help having an app.* [P9, interview, I]

Using the intervention on a mobile phone might also be more private, especially in households where a laptop was shared among the family members. However, an advantage of using a laptop was to allow a space to reflect, as users may easily get distracted when doing the program on the go. For example, a user said:

> *...going onto a laptop and doing some kind of regular, some sort of like a class or something, actually made it, gave it a bit more weight, made it feel like I was taking it a bit more seriously, so I personally found doing it via a laptop would probably be better in terms of my level of engagement.* [P9, I]

Furthermore, an issue highlighted by participants was that the mobile version did not include all the intervention materials:

> *I don't think you could access it on your mobile in the same way, it was just stripped. You have to go onto the computer to do some of it.* [P7, I]

A Flexible and Everyday Intervention

The online nature of the program made it easy to embed the intervention into their everyday lives:

> *I found sometimes during therapy though, it became like a silo, separate from your normal life. Whereas this, because it became part of your normal life, you could think about it, think about reflecting on aspects of your life, whilst you are living it almost.* [P7, I]

They were able to flexibly go over the notes whenever and as many times as they wanted to (eg, "I often struggle to remember things discussed during therapy sessions"). The idea of engaging with the program on an ongoing basis through the program content and diary logging, reflecting on oneself regularly, also created a sense of continuity of self-improvement. For example, as participant 11 put it:

> *And I think, what I like about it, because you're doing these diaries weekly, you are working on yourself all the time.*

However, despite the advantages of the guided self-help intervention being online, additional downloadable or offline materials were still preferred by some participants. In this way, they could go through the materials at their own pace; otherwise, the process might become "too quick" and "difficult" (P5, I).

Reflections About Their ED Recovery

Many participants described the helpfulness of the program (eg, "Lots of info. A lot to take in. Helpful reflection and realisation.") in relation to 4 broad areas: (1) cognitive, (2) emotional, (3) interpersonal, and (4) behavioral aspects.

Cognitively, the program helped them gain new understanding and insights about their ED. One participant said:

> *I didn't realise an ED was related to black and white thinking, to me that was part of my other diagnosis (borderline personality disorder), I didn't realise it impacts my eating at all.* [P9, I]

Some mentioned specific program elements as helpful, for instance:

> *I can see that I have changed my attitude to exercise...now it is about contributing to more mental health and physical wellbeing...* [P6, I]

Nevertheless, some thought that the program did not offer any new information. Others thought it was still important to be reminded even though they were aware of the information:

> *I have seen and done some of these things before but they are equally valuable and need to be returned to again and again. I realise how obsessed I have been and still am to a degree with my body and how shame and disgust still rule me even though I have spent most my life with eating issues.* [IF]

Emotionally, participants mentioned being better at regulating their difficult emotions and gaining self-confidence and acceptance. One person described "feeling comfortable in her own skin now, which was an odd feeling" (P9, I). In relation to interpersonal relationships, one participant thought she had learned to deal with people's comments about her instead of resorting to binge eating, and some mentioned being able to participate in social situations that involved food, such as family meals.

The practical techniques introduced in the program, such as the use of a reflection diary, mindfulness, and carrying out behavioral experiments, were seen as useful by some participants in affecting behavioral changes. Specific targeted behaviors included having fewer binge eating episodes as well as examining and reducing avoidance behaviors. For example, P3 said:

> *I ended up having like 6 weeks which I haven't binged at all, which was a massive achievement to me.*

Participants appreciated the action-oriented nature of the program:

> *I found that quite helpful, because you are putting it into practice. You are not just reading, you are doing something.* [P1, I]

Not a One Size Fits All

Participants' attitudes to and views of perceived usefulness and relevance of the program content were mixed. They also showed

different engagement styles in the group discussion forum. As P2 noted:

> *People have different pasts and some have been struggling for years and years, and for some of them it's a wake-up call...knowing that this isn't best for all.*

Participants were in different stages of recovery, and their EDs were maintained by different factors. It could be difficult for a program to be "geared around different people, different levels, identifying the balance" (P3, I).

Participants held a spectrum of attitudes toward engaging with the program as well as communicating with their online therapist. Several participants pointed out that the delay in receiving help from the NHS prompted them to try other forms of support:

> *Yeh, I'm one of the lucky ones, because if not I'd still be waiting, you know its 14 months, I waited for any kind of help and you know and that's a long time without help and I can see why people with eating disorders whichever one it is, end up, you know, doing silly stuff, self-harming or committing suicide, it's because you've just got no one to help.* [P12, I]

To some participants, the intervention felt like a "God-sent," to "help kickstart things again" (P4, I). Some held a curious stance, wanting to try something new:

> *I just really wanted to do it, you know, it was something I hadn't heard about before and just really wanted to do it.* [P11]

Others held a more ambivalent or skeptical view. Some participants hoped for more intensive and face-to-face help, especially those who considered themselves as "very, very, seriously unwell." P10 also argued that as the program was intended for people who were on the waiting list for therapy, knowing that she would be receiving therapy soon, she "didn't invest in the programme a lot," but instead she was just "reading, ticking the box, that kind of thing."

Divergent attitudes were also held regarding the therapist-user interaction. It was acknowledged that although this form of intervention helped reduce the power imbalance between the user and the therapist and kept users accountable and on track, the flip side was that the remoteness of the therapist-patient interaction diminished the trust in and the sense of *human-ness* of the therapist. To illustrate, some participants mentioned that they did not feel like a patient when accessing the program:

> *I think that power imbalance when you are in therapy, almost keeps you with that eating disorder, whereas this doesn't because I found it quite empowering, because I was in control, I could have clicked on this module or not, take the information or not, but it was different for me.* [P7, I]

On the other hand, other participants found the online therapist difficult to relate to as they did not know who they were. Although some expected to receive some *expert help* even though it was not *an in-depth therapy*, they were doubtful about the *therapist's* qualifications and wondered if:

> *It was a very clever algorithm because the responses were very packed and formulaic...Maybe it's something about online that it's very hard to kind of set up that relationship...I admit it must as well be it's been hard for me to get past my own prejudices.* [P2, I]

In relation to the program content, the difference in the perceived relevance and usefulness of the program was further complicated by issues around intersectionality, such as *gender, discrimination, prejudice, and disability*. P2 felt the program did not capture the complexity of the issue, and the program needed to "acknowledge that society impacts on women in different ways, and different people experience it in different ways." Comorbidity also influenced participants' ability to work on the program. For example, P1 said that it is not that easy for her to take the learnings onboard when her mood was low, and she felt depressed.

In particular, participants who were older (over 50 years) and those living in larger bodies felt they were not well served by the program, as reflected in the examples, case studies, and exercises. One user commented:

> *I feel the course has been totally dismissive of any problems that come with obesity. It is not represented in any of the examples given, any questions asked or any exercises offered...A lot of it seems based on the idea that we should just accept our body. Well, I'm currently walking past posters from Cancer Research telling me that OBESITY is one of the largest causes of cancer, and I'm meant to ignore that in favour of feeling good about my body, am I?"* [SF]

The other example was shown in the case studies, where people with BED did not find that the case studies resonated with them or saw the photos used as being *glamorized* and not *authentic* but *imaginary.* Participants who were older felt that the program seemed to target a younger audience, a participant expressed that "there was no elderly women in the pictures in the program" For example, P7 expressed the following:

> *It seems slightly more geared towards younger women than older women, say those little videos and everything...Getting old with an eating disorder is a different challenge in itself...I think I hear things about women my age with an eating disorder, it always feels a little bit like it's been written off, like there's no hope after a certain age if you have it for a certain amount of time that it will be with you for life.*

The series of mirror exposure exercises in the program exemplified the issue about the relevance of the program to some extent. Some mentioned that those exercises were clearly geared toward people with normal weight. That said, most participants found the mirror exercises challenging, although some found them useful in boosting self-confidence and self-acceptance. A user (P9, I) mentioned "at the end of the day, anything that's worthwhile isn't gonna be easy."

Furthermore, some participants felt that attractiveness was overemphasized in the program. Users disagreed with the focus

and felt they did not "give a damn about the media portray(al) of what an ideal body shape or weight is" (SF). Others did find the focus on external factors useful. For example, a user mentioned that it made her realize how her "distorted body image was in many ways the by-product of a distorted beauty ideal" and that recognizing that this was part of a long socialization process gave her a degree of comfort.

Participants who found the program unrelatable and not very useful shared reflections on its potential to exacerbate unhelpful or distressing thoughts and behaviors. A user posted on the platform:

> *I had become very depressed and had begun to think about self-harming as a way of relieving the tension and anxiety I was feeling. I realised a lot of that was coming from the exercises we had been doing as part of the programme, and my frustrations with the programme.* [GF]

One interviewee decided to stop doing the program. However, there were also participants who found the program comprehensive and "didn't appear to have anything lacking" (P4, I) and found the case studies engaging and useful.

The inability to relate to the program was also shown in the group discussion forum, acting as a barrier to engagement. Some participants did not feel included or that they fitted in the forum due to having different perceptions of their issues or not wanting to upset others. A user said:

> *I am not a big girl, I am a normal size, I am trying to join in a conversation about how you feel about how you look when you know that there are people out there, that battling harder in a way, because they are so large. I just couldn't feel I could do it, so it's difficult when there's being such different from other people using the forum.* [P4, I]

A myriad of engagement styles on the group forum was displayed. Some people used the forum to support others and be supported, which helped them to feel less alone (P11, I). Other users described being a *lurker* in the group, reading others' messages as opposed to being actively engaged. For instance, one of them mentioned that this was not for her and that she never wanted to participate in anything by writing or asking for support (P10, I).

Discussion

Principal Findings

In line with the stated aims of our study, namely, to understand participants' experience of the everyBody Plus web-based intervention, to add to and complement future quantitative RCT findings, and to add to the emerging literature on people's experience of web-based interventions for BN and BED, our TA generated 4 themes:

1. Tailoring the user journey: users had mixed feedback regarding the UX and the design of the program and expressed that the web-based intervention should be both customized and personalized.
2. A flexible and *everyday* intervention: the program's online nature favored flexibility and continuity to embed the program into daily lives.
3. Reflections about their ED recovery: participants reported positive impacts of the program from a cognitive, emotional, interpersonal, and behavioral perspective.
4. Not a *one size fits all*: attitudes regarding the program content (case studies and the group forum included), usefulness, and relevance were divergent. In particular, participants who were older and had a larger BMI held more negative views.

Many of the comments encompassed within these themes speak to more than 1 of our 3 study aims. We discuss the findings from each of the themes in the following paragraphs.

In relation to theme 1, the UX of the program, concerns over lengthy content alongside a lack of a bookmark function to track users' progress, reduces the perceived ease of use, as postulated in the Technology Acceptance Model. These findings highlight the need for using in-depth service user involvement in co-designing interventions before embarking on clinical RCTs. For example, Graham et al [24] argued for the importance of design research methods (eg, *think-aloud* protocol, ethnography, and user testing) in the development of web-based interventions for EDs. In this study, some of the problems identified here with UX could perhaps be prevented if a formal UX study had been carried out before the RCT; yet, due to aligning with the German version of the intervention and time constraints, a prior study was not done. Although we did not conduct a formal usability study here, our participants' comments suggest a much more mixed usability experience than that found by Nitsch et al [15]. This may be because in the Nitsch et al [15] study, participants were women aged 18-25 years with an interest in improving body image and reducing disordered eating behaviors, who were recruited via the internet or social media and who were reimbursed for participating in the study. Thus, their views may not fully reflect those of people with clinical EDs included here. A limitation here was that the mobile version did not include all the intervention materials that were accessible on the laptop or desktop version; thus, we could not firmly draw any conclusion on the perceived relative merits of mobile or web versions. However, the analysis reflects that rather than designing an intervention as either *mobile-first* or on the web, a hybrid approach could be adopted. Participants pointed out the different nature and usage patterns of mobile phones and laptops. The findings of this study echo a previous study by Morrison et al [25]—they found that the mobile app content was typically used on-the-go and browsed for shorter periods. Different elements of the program could be augmented on mobile or laptop versions, and the choice should be promoted, which also includes having an option to download the materials or have an offline reading. Participants' feedback about the laptop version, making them take the intervention more seriously and allow them more space for reflection, matches the findings of Dennison et al [26]. These authors reported that apps were seen as more *disposable* rather than a long-term commitment. Although convenience is an important factor for adoption in digital health intervention [15], it seems that for sustained engagement, the benefits of using a laptop need to be considered,

especially when the program materials are complex and require active reflection.

In relation to theme 2, the flexibility of the web-based program transcends geographical and time restrictions, indicating an advantage of this form of intervention delivery, as previously noted in other qualitative studies (refer to the review by Yim and Schmidt [9]). A further advantage is the ease with which the intervention is integrated into people's daily lives. As the *therapy time* is not scheduled, flexibility allows participants to access the intervention at any time, including when they are struggling with their symptoms. Completing a weekly symptom diary also gives a momentum to work on making psychological and behavioral changes. This advantage has not been mentioned in previous studies on web-based interventions for EDs. Andreassen et al [27] described a changing spatial configuration of intervention delivery from clinical spaces into domestic spaces in eHealth and postulated that this might make the difficulty more pronounced at home. In the current context, perhaps the change in therapy space is beneficial as users could reflect and access support while *living it.* This supports the concept of agency synthesized in the meta-ethnography by Yim and Schmidt [9], where users have more control over the *therapy* and hence their recovery.

Findings from theme 3, namely, participants' reflections on their EDs are broadly consistent with meta-analysis by de Vos et al [28] regarding the criteria for ED recovery. In addition to ED symptom reduction, participants mentioned psychological, behavioral, and interpersonal changes. These reflections seem to be more comprehensive than themes identified in previous studies such as *improvements in bulimic symptoms* [29] and *changes in ED symptoms* [30], thereby increasing the perceived usefulness of the program. Indeed, the impact of EDs goes beyond cognitions and behaviors, with individuals with EDs experiencing interpersonal difficulties [31] and difficulties in emotional regulation [32]. Hence, the results of this study suggest that this web-based program has the potential to encompass a more comprehensive approach and to enable people to make holistic changes and improvements in their lives.

In relation to the fourth and last theme, similar to previous qualitative studies (refer to the review by Yim and Schmidt [9]), very heterogeneous attitudes and feedback were found. The dialectic of *anonymity* and *safety* versus *remoteness* and the *robotic* quality of web-based interventions highlights the idea that the computer can be perceived as a shield and buffer, yet could also be hard for participants to relate to. This was compounded by the complexity of the diverse engagement patterns in the group forum. Our previous study on people with EDs' and their carers' views on online self-help discussed the importance of interaction among participants and the community formed in the group discussion forum to break the isolation (Yim et al, unpublished data, 2020). However, this study reveals a more complex picture. We show the risk of further pushing people away due to not being able to identify with others in the group or fear of receiving unhelpful comments. Such findings are similar to qualitative studies on group cognitive behavioral therapy for EDs, whereby participants also commented on the composition of groups and how feeling different from others can interfere with how supported they feel [33]. The division was particularly felt among users of different body sizes and the extent to which they engaged in compensatory behaviors. This raises questions around how a supportive community could be formed as well as how group forums could be moderated to foster inclusivity of people with different characteristics.

Strengths and Limitations

To our knowledge, this qualitative study is the first process study within a larger RCT that included both in-session feedback and interviews in the analysis of a digital intervention for EDs. The incorporation of real-time feedback from all participants (N=104) was a key strength, ensuring representativeness and generalizability of our findings. However, the sample size of the interview component of the study (n=12) was small, and the interviewees were not fully representative of the whole sample in some aspects, as they were older and a high proportion had completed the full intervention when compared with the 36.3% (41/113) intervention completion rate of the UK sample in the RCT. When considering its limitations, themes relating to how the program facilitated or hindered face-to-face therapy were not generated. It would be useful to understand how the program might be synergistic or in conflict with the therapists' approaches when participants were off the waiting list, as the intervention was intended to *bridge* the waiting time between referral and face-to-face treatment. One should be cautious that the results of this study were based on data from the UK participants in the everyBody Plus trial only and may not fully represent the experiences of people from other countries using the everyBody Plus intervention in different cultural or health care contexts or those of people using other web-based interventions for EDs. Additionally, this program was designed for adult females, and the scope of feedback excludes the experiences and views of male users of web-based interventions. Nonetheless, this study raises several important implications for the RCT and for research into eHealth for EDs in general.

Implications for Future Research and Practice

Implications for Digital Interventions for EDs in General

Research on intervention efficacy and effectiveness needs to consider the program layout and usability, and design research methods [24] should be adopted before any large-scale RCT to minimize technical issues affecting intervention uptake and adoption. Participants' different preferences for mobile or laptop use suggest that instead of designing for either eHealth or mHealth, programs should be designed as a hybrid, giving users a choice. Some features that are more routine and can easily be done *on the go*, such as self-monitoring diaries, could be designed as *mobile-first*.

Clinically, this, like other web-based interventions that may be used while people are waiting for face-to-face therapy, shows potential for increasing access and preparing people for therapy through teaching cognitive behavioral principles and helping them reflect on their thinking and behaviors. Clinicians need to be aware of and monitor people's motivations, explaining the purpose of the program as a treatment intervention in its own right, to avoid people treating this as a *tick-box* exercise knowing that they have already secured face-to-face therapy sessions.

In addition, our findings cast doubts on the possibility of designing a program that is inclusive of and speaks to the full range of people with binge eating–related issues. Another issue any web-based program needs to consider is the potential harm group forums may pose to people who do not identify with others in the groups, as found here. Future research and design of web-based programs may want to consider using more tailored group forums, depending on key clinical characteristics. For instance, having separate forums for those who do and do not use compensatory behaviors, such as self-induced vomiting, could allow participants to feel more included and make the forums more relevant to their specific needs.

Implications Specific to everyBody Plus and the RCT

For the everyBody Plus intervention specifically, the results reveal the need for future iterations to improve the UX. The Technology Acceptance Model identified *ease of use* as a factor for technology adoption. Improving the UX will be paramount to increase adoption and adherence. In the version used in this study, there was no way for participants to return to the page they last visited or to bookmark information. As the modules were perceived as long, the bookmarking function was important. The mobile version needs to be improved as some of the training materials are accessible only via a laptop or desktop computer.

At the time of writing this paper, the main findings from the trial in relation to the clinical outcomes are yet unknown. Although the main analyses of the RCT (from which the present data are drawn) are prespecified in the trial protocol [16], qualitative data such as those presented here may offer ideas for additional, more exploratory (moderator) analyses and offer a nuanced understanding of reasons explaining the attrition rate in the main intervention.

Conclusions

In conclusion, this study has highlighted what participants find helpful and positive about this web-based intervention and provided suggestions for improvements to the future design of such interventions. This will help inform and complement the upcoming analysis of the RCT to allow us to better draw conclusions on the next steps for the web-based intervention. Ultimately, these findings also generate insights for interested parties when designing and evaluating complex digital health interventions.

Acknowledgments

The authors would like to thank the following National Health Service (NHS) Trusts who assisted with the recruitment and online support of the participants (in alphabetical order): Avon and Wiltshire Mental Health Partnership NHS Trust; Birmingham and Solihull Mental Health NHS Foundation Trust; Black Country Partnership NHS Foundation Trust; Bradford District Care NHS Foundation Trust; Cambridgeshire and Peterborough NHS Foundation Trust; Mersey Care NHS Foundation Trust; North West Boroughs Healthcare NHS Foundation Trust; Northamptonshire Healthcare NHS Foundation Trust; Rotherham Doncaster and South Humber NHS Foundation Trust; Sheffield Health and Social Care NHS Foundation Trust; Somerset Partnership NHS Trust; South London and Maudsley NHS Foundation Trust; South Staffordshire and Shropshire Healthcare NHS Foundation Trust; the Tees, Esk and Wear Valleys NHS Foundation Trust; the Lancashire Care NHS Foundation Trust; and the 2gether NHS Foundation Trust. This study was supported by the ICare (Integrated Technology in Mental Health Care Delivery in Europe) Consortium under the European Union's Horizon 2020 research and innovation program under grant no 634757. US was supported by a Senior Investigator Award from the National Institute for Health Research (NIHR) and received salary support from the NIHR Biomedical Research Centre for Mental Health, South London and the Maudsley NHS Foundation Trust and the Institute of Psychiatry, Psychology and Neuroscience, King’s College London. The views expressed herein are not those of NIHR or the NHS.

Authors' Contributions

SY and US conceived the study and its design. SY and EB worked on data collection and transcription, and data were analyzed by SY under the supervision of US. GG, NG, and PS were involved in the design of the web-based intervention and the RCT. All authors read, commented on, and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview schedule/questions guide.
[DOCX File , 25 KB - jmir_v22i9e17880_app1.docx]

Multimedia Appendix 2
Themes and quotes.
[DOCX File , 34 KB - jmir_v22i9e17880_app2.docx]

References

1. The NHS Long Term Plan. NHS England. 2019. URL: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2020-06-25]
2. Foley T, Woollard J. The Digital Future of Mental Healthcare and Its Workforce: A Report on a Mental Health Stakeholder Engagement to Inform the Topol Review. The Topol Review — NHS Health Education England. 2019. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-Mental-health-paper.pdf [accessed 2020-06-25]
3. Traviss-Turner GD, West RM, Hill AJ. Guided self-help for eating disorders: a systematic review and metaregression. Eur Eat Disord Rev 2017 May;25(3):148-164. [doi: 10.1002/erv.2507] [Medline: 28276171]
4. Anastasiadou D, Folkvord F, Lupiañez-Villanueva F. A systematic review of mHealth interventions for the support of eating disorders. Eur Eat Disord Rev 2018 Sep;26(5):394-416. [doi: 10.1002/erv.2609] [Medline: 29927004]
5. Loucas CE, Fairburn CG, Whittington C, Pennant ME, Stockton S, Kendall T. E-therapy in the treatment and prevention of eating disorders: a systematic review and meta-analysis. Behav Res Ther 2014 Dec;63:122-131 [FREE Full text] [doi: 10.1016/j.brat.2014.09.011] [Medline: 25461787]
6. Pittock A, Hodges L, Lawrie SM. The effectiveness of internet-delivered cognitive behavioural therapy for those with bulimic symptoms: a systematic review: a review of iCBT treatment for bulimic symptoms. BMC Res Notes 2018 Oct 22;11(1):748 [FREE Full text] [doi: 10.1186/s13104-018-3843-2] [Medline: 30348226]
7. Yim SH, Schmidt U. Self-help treatment of eating disorders. Psychiatr Clin North Am 2019 Jun;42(2):231-241. [doi: 10.1016/j.psc.2019.01.006] [Medline: 31046925]
8. Ludvigsen MS, Meyer G, Hall E, Fegran L, Aagaard H, Uhrenfeldt L. Development of clinically meaningful complex interventions - the contribution of qualitative research. Pflege 2013 Jun;26(3):207-214. [doi: 10.1024/1012-5302/a000292] [Medline: 23732316]
9. Yim SH, Schmidt U. Experiences of computer-based and conventional self-help interventions for eating disorders: a systematic review and meta-synthesis of qualitative research. Int J Eat Disord 2019 Oct;52(10):1108-1124. [doi: 10.1002/eat.23142] [Medline: 31343088]
10. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
11. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. Br Med J 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
12. Oakley A, Strange V, Bonell C, Allen E, Stephenson J, RIPPLE Study Team. Process evaluation in randomised controlled trials of complex interventions. Br Med J 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]
13. Jacobi C, Völker U, Trockel MT, Taylor CB. Effects of an internet-based intervention for subthreshold eating disorders: a randomized controlled trial. Behav Res Ther 2012 Feb;50(2):93-99. [doi: 10.1016/j.brat.2011.09.013] [Medline: 22137366]
14. Saekow J, Jones M, Gibbs E, Jacobi C, Fitzsimmons-Craft EE, Wilfley D, et al. StudentBodies-eating disorders: a randomized controlled trial of a coached online intervention for subclinical eating disorders. Internet Interv 2015 Nov;2(4):419-428. [doi: 10.1016/j.invent.2015.10.004]
15. Nitsch M, Dimopoulos CN, Flaschberger E, Saffran K, Kruger JF, Garlock L, et al. A guided online and mobile self-help program for individuals with eating disorders: an iterative engagement and usability study. J Med Internet Res 2016 Jan 11;18(1):e7 [FREE Full text] [doi: 10.2196/jmir.4972] [Medline: 26753539]
16. Vollert B, Beintner I, Musiat P, Gordon G, Görlich D, Nacke B, et al. Using internet-based self-help to bridge waiting time for face-to-face outpatient treatment for Bulimia Nervosa, Binge Eating Disorder and related disorders: study protocol of a randomized controlled trial. Internet Interv 2019 Apr;16:26-34 [FREE Full text] [doi: 10.1016/j.invent.2018.02.010] [Medline: 30775262]
17. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
18. Pretorius N, Arcelus J, Beecham J, Dawson H, Doherty F, Eisler I, et al. Cognitive-behavioural therapy for adolescents with bulimic symptomatology: the acceptability and effectiveness of internet-based delivery. Behav Res Ther 2009 Sep;47(9):729-736. [doi: 10.1016/j.brat.2009.05.006] [Medline: 19515360]
19. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
20. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. London: Sage Publications; 2013.
21. O'Reilly M, Parker N. 'Unsatisfactory saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. Qual Res 2012 May 17;13(2):190-197. [doi: 10.1177/1468794112446106]
22. Answers to Frequently Asked Questions About Thematic Analysis. The University of Auckland. 2019. URL: https://cdn.auckland.ac.nz/assets/psych/about/our-research/documents/Answers%20to%20frequently%20asked%20questions%20about%20thematic%20analysis%20April%202019.pdf [accessed 2020-06-25]

23. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.
24. Graham AK, Wildes JE, Reddy M, Munson SA, Barr Taylor C, Mohr DC. User-centered design for technology-enabled services for eating disorders. Int J Eat Disord 2019 Oct;52(10):1095-1107. [doi: 10.1002/eat.23130] [Medline: 31313370]
25. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014 Oct 22;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
26. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
27. Andreassen HK, Dyb K, May CR, Pope CJ, Warth LL. Digitized patient-provider interaction: how does it matter? A qualitative meta-synthesis. Soc Sci Med 2018 Oct;215:36-44. [doi: 10.1016/j.socscimed.2018.08.036] [Medline: 30205277]
28. de Vos JA, LaMarre A, Radstaak M, Bijkerk CA, Bohlmeijer ET, Westerhof GJ. Identifying fundamental criteria for eating disorder recovery: a systematic review and qualitative meta-analysis. J Eat Disord 2017;5:34 [FREE Full text] [doi: 10.1186/s40337-017-0164-0] [Medline: 29118983]
29. McClay C, Waters L, Schmidt U, Williams C. A survey of attitudes towards computerized self-help for eating disorders within a community-based sample. Behav Cogn Psychother 2014 Nov 28;44(1):65-78. [doi: 10.1017/s1352465814000484]
30. Sánchez-Ortiz VC, House J, Munro C, Treasure J, Startup H, Williams C, et al. 'A computer isn't gonna judge you': a qualitative study of users' views of an internet-based cognitive behavioural guided self-care treatment package for bulimia nervosa and related disorders. Eat Weight Disord 2011 Jun;16(2):e93-101. [doi: 10.1007/BF03325314] [Medline: 21989103]
31. Arcelus J, Haslam M, Farrow C, Meyer C. The role of interpersonal functioning in the maintenance of eating psychopathology: a systematic review and testable model. Clin Psychol Rev 2013 Feb;33(1):156-167. [doi: 10.1016/j.cpr.2012.10.009] [Medline: 23195616]
32. Ruscitti C, Rufino K, Goodwin N, Wagner R. Difficulties in emotion regulation in patients with eating disorders. Borderline Personal Disord Emot Dysregul 2016;3:3 [FREE Full text] [doi: 10.1186/s40479-016-0037-1] [Medline: 27252857]
33. Laberg S, Törnkvist A, Andersson G. Experiences of patients in cognitive behavioural group therapy: a qualitative study of eating disorders. Scand J Behav Ther 2001 Jan;30(4):161-178. [doi: 10.1080/02845710127474]

Abbreviations

BED: binge eating disorder
BN: bulimia nervosa
COREQ: Consolidated Criteria for Reporting Qualitative Research
ED: eating disorder
ISRCTN: International Standard Randomised Controlled Trial Number
mHealth: mobile health
NHS: National Health Service
NIHR: National Institute for Health Research
RCT: randomized controlled trial
TA: thematic analysis
UX: user experiences

Edited by G Eysenbach; submitted 18.01.20; peer-reviewed by M Moessner, M Sadeghi, L McCann; comments to author 16.03.20; revised version received 26.06.20; accepted 19.07.20; published 23.09.20.

Please cite as:
Yim SH, Bailey E, Gordon G, Grant N, Musiat P, Schmidt U
Exploring Participants' Experiences of a Web-Based Program for Bulimia and Binge Eating Disorder: Qualitative Study
J Med Internet Res 2020;22(9):e17880
URL: http://www.jmir.org/2020/9/e17880/
doi:10.2196/17880
PMID:32965235

Original Paper

Self-Practice of Stabilizing and Guided Imagery Techniques for Traumatized Refugees via Digital Audio Files: Qualitative Study

Catharina Zehetmair[1*], MSc; Ede Nagy[1*], Dr sc hum; Carla Leetz[1*], BSc; Anna Cranz[1*], Dr sc hum; David Kindermann[1*], Dr med; Luise Reddemann[2*], Prof Dr med; Christoph Nikendei[1*], (apl) Prof Dr med, MME

[1]Department of General Internal Medicine and Psychosomatics, Center for Psychosocial Medicine, University Hospital of Heidelberg, Heidelberg, Germany

[2]Department of Clinical Psychology, Psychotherapy and Psychoanalysis, University of Klagenfurt, Klagenfurt, Austria

[*]all authors contributed equally

Corresponding Author:
Catharina Zehetmair, MSc
Department of General Internal Medicine and Psychosomatics
Center for Psychosocial Medicine
University Hospital of Heidelberg
Thibautstraße 4
Heidelberg, 69115
Germany
Phone: 49 6221568373
Fax: 49 6221 568612
Email: Catharina.Zehetmair@med.uni-heidelberg.de

Abstract

Background: Refugees have an increased risk of developing mental health problems. There are insufficient psychosocial care structures to meet the resulting need for support. Stabilizing and guided imagery techniques have shown promising results in increasing traumatized refugees' emotional stabilization. If delivered via audio files, the techniques can be practiced autonomously and independent of time, space, and human resources or stable treatment settings.

Objective: This study aimed to evaluate the self-practice of stabilizing and guided imagery techniques via digital audio files for traumatized refugees living in a reception and registration center in Germany.

Methods: From May 2018 to February 2019, 42 traumatized refugees participated in our study. At T1, patients received digital audio files in English, French, Arabic, Farsi, Turkish, or Serbian for self-practice. Nine days later, at T2, a face-to-face interview was conducted. Two months after T2, a follow-up interview took place via telephone.

Results: At T2, about half of the patients reported the daily practice of stabilizing and guided imagery techniques. At follow-up, the average frequency of practice was once weekly or more for those experiencing worse symptoms. No technical difficulties were reported. According to T2 and follow-up statements, the techniques helped the patients dealing with arousal, concentration, sleep, mood, thoughts, empowerment, and tension. The guided imagery technique "The Inner Safe Place" was the most popular. Self-practice was impeded by postmigratory distress factors, like overcrowded accommodations.

Conclusions: The results show that self-practice of stabilizing and guided imagery techniques via digital audio files was helpful to and well accepted by the assessed refugees. Even though postmigratory distress factors hampered self-practice, "The Inner Safe Place" technique was particularly well received. Overall, the self-practiced audio-based stabilizing and guided imagery techniques showed promising results among the highly vulnerable group of newly arrived traumatized refugees.

(J Med Internet Res 2020;22(9):e17906) doi:10.2196/17906

KEYWORDS

stabilizing techniques; guided imagery; refugees; qualitative analyses; posttraumatic stress disorder; mental health; PTSD; audio; therapy

Introduction

With a prevalence rate of approximately 40%, mental health is a major problem for refugees in their host country [1]. Posttraumatic stress disorder (PTSD) is one of the most commonly reported mental health issues [1,2]. Multiple studies support trauma exposure approaches like narrative exposure therapy, trauma-focused cognitive behavioral therapy (CBT), and eye movement desensitization and reprocessing to address PTSD symptoms in the refugee population [1,3,4]. However, displaced people face many obstacles limiting preconditions for trauma exposure therapy, such as frequent reallocation, uncertainty regarding their asylum application outcome as well as financial, intercultural, and language barriers upon arrival [5]. Furthermore, a premature trauma confrontation should be avoided [6]. Hence, providing traumatized refugees with initial stabilizing treatment may be very helpful until the refugees' surroundings are sufficiently stable for trauma confronting treatment approaches.

Evidence is growing that stabilization techniques positively affect refugees' mental health [7-9]. In this context, stabilizing and guided imagery techniques in line with Reddemann [10] are promising treatment approaches for adult and minor refugees with PTSD in individual and group therapy approaches [11-14]. The use of stabilizing and guided imagery techniques seems helpful in securing an initial emotional stabilization even under unstable conditions [13,14]. However, all face-to-face therapeutic interventions with refugees face problems of language heterogeneity and the need for space, time, and human resources. Given the fact that online self-help interventions are almost as effective as face-to-face interventions [15,16], modern media with app- and web-based alternatives offer possibilities to bridge these barriers impeding traditional face-to-face psychotherapy.

Web- and app-based studies in health-related areas have increased significantly in recent years. Systematic reviews show positive results from web- and app-based interventions [17-21]. Especially for patients with trauma-related disorders, web-based CBT [22-28], coping strategy programs [29,30], as well as computer games to re-consolidate traumatic intrusive memories [31,32], have been shown to have promising effects in reducing trauma symptoms [33]. Further, app-based self-help interventions using image- and audio-based formats show encouraging effects on distress, PTSD, and depression symptoms [16,34-37]. Therefore, digital psychosocial interventions are a promising approach for asylum seekers and refugees.

This study builds on earlier studies using stabilizing and guided imagery techniques in newly arrived refugees [13,14] to evaluate traumatized refugees' self-practice of stabilizing and guided imagery techniques via digital audio files using in-depth qualitative analysis. Our research questions were the following: (1) What kind of practicing behavior is shown? (2) Which technical difficulties are reported? (3) What clinical effects do the patients achieve through stabilizing and guided imagery techniques via digital audio files? (4) What difficulties are reported in the self-practice of the techniques via audio files? In previous studies, the guided imagery technique called "The Inner Safe Place" has played a key role [11,12], yet little is known about the experiences with this technique. Hence, we have additionally focused on the question of (5) experiences specific to the guided imagery technique, "The Inner Safe Place."

Methods

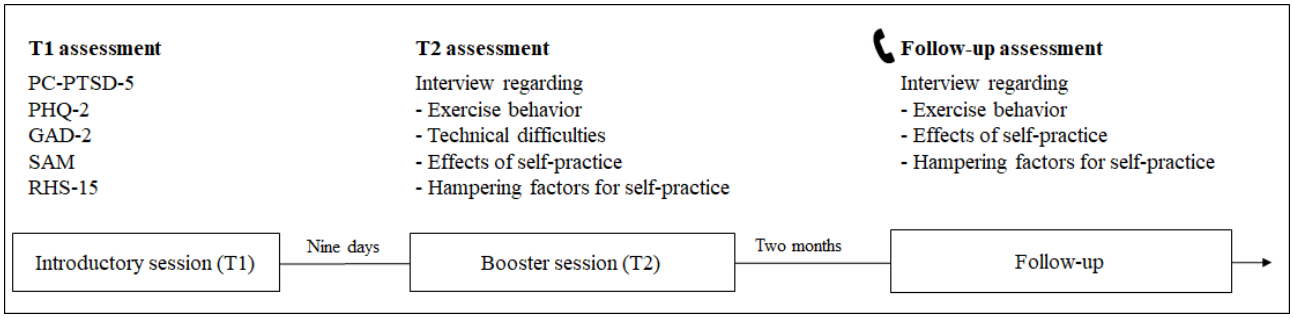
Participants and Study Design

We conducted a prospective, descriptive study using qualitative semistructured interviews. The setting was the refugee state registration and reception center "Patrick Henry Village" (PHV), Heidelberg-Kirchheim, Germany. At the PHV, the University Hospital of Heidelberg, in cooperation with physicians in private practice, operates a medical and psychosocial walk-in clinic [38,39]. Between the end of May and early December 2018, refugees who sought help in the psychosocial walk-in clinic [38] and met our study's inclusion criteria were referred to treatment with the audio-based stabilizing and guided imagery techniques. Due to follow-up interviews conducted two months later, the overall study period (recruitment and follow-up) was from May 2018 until February 2019. Inclusion criteria were a diagnosis of PTSD, access to a personal smartphone for digital audio file delivery, and fluency in speaking and understanding of one of the following languages: English, French, Arabic, Farsi, Turkish, or Serbian. Exclusion criteria were substance addiction, current psychosis, and age under 18 years. All participants had applied for asylum in Germany or were in the process of applying during the intervention.

Procedure and Ethics

When indicated, the psychiatrists and psychotherapists of the psychosocial walk-in clinic recommended treatment with the audio-based stabilizing and guided imagery techniques to patients and made a face-to-face appointment for an introductory session. During the introductory session (T1), the patients first were asked to complete a baseline measurement via a tablet. Afterward, the psychologist discussed psychoeducational issues with the patients and informed them of the content of the audio files and the aim of stabilizing and guided imagery techniques. Then, the psychologist and the patients practiced the stabilizing and guided imagery techniques together once using the audio files and subsequently discussed the effects and difficulties of each technique. Finally, the audio file was transferred to the patient's phone in the appropriate language. The introductory session lasted approximately two hours. The booster session (T2) took place nine days later and particularly focused on counseling and feedback regarding difficulties in practicing the techniques using audio files. Since the focus of this study was put on the self-practice of stabilizing and guided imagery techniques using the provided audio files, no further guided practice sessions were undertaken. An interview was conducted with each patient at the end of the session. Two months after T2, the patients were contacted again by phone and interviewed for a second time (follow-up). Three attempts were made to reach the patients for follow-up interviews. If necessary, a telephone or face-to-face translator was involved during the introductory session and the booster session. Figure 1 provides an overview of the study procedure.

Figure 1. An overview of the study procedure.



The study was approved by the ethics committee of the University of Heidelberg (S-640/2016), and all participants gave their written informed consent in accordance with the Declaration of Helsinki.

Audio Files

The audio file consisted of three parts, namely (1) mindful breathing, (2) the body scan, and (3) the guided imagery technique "The Inner Safe Place," which are described in more detail elsewhere [13,14]. The audio files were available in English, Arabic, Farsi, French, Turkish, and Serbian (spoken by about 80% of the refugee population in the PHV). Before, the text material of the techniques had been translated by professional interpreters and translators into the respective language and then digitally recorded by native speakers. Except for the Arabic version, all versions were narrated by a female native speaker. The audio files and written instructions are available free of charge [40]. If desired, a printed instruction booklet is also available for a nominal fee.

Psychometric Baseline Assessment (T1)

Prior to the introductory session (T1), a baseline measurement including the Primary Care PTSD Screen for DSM-5 (PC-PTSD-5) [41], the two-item Patient Health Questionnaire (PHQ-2) [42], the short version of the General Anxiety Disorder questionnaire (GAD-2) [43], the Self-Assessment Manikin scale (SAM) [44], and the distress thermometer of the Refugee Health Screener-15 (RHS-15 distress thermometer) [45] were used to assess participants' mental distress.

The PC-PTSD-5 [41] assesses PTSD symptoms via a list of different trauma events and five binary questions (0 = "no" and 1 = "yes") on re-experiencing, avoidance, physical reactions, emotional numbness, and trauma-distorted feelings of guilt and blame [41]. With a total score between 0 and 5, individuals with a score ≥3 are identified as patients with probable PTSD. The PC-PTSD-5 shows good sensitivity (.93), acceptable specificity (.85), and good acceptance by patients [46].

The PHQ-2 [42] assesses major depression via two items on anhedonia and depressed mood. Both items are rated on a scale of 0 (not at all) to 3 (nearly every day) and give a total score of 0 to 6. The PHQ-2 shows good construct validity (r from 0.67-0.87), good internal consistency (α=.83) [47]. An overall cut-off score at ≥3 provides a sensitivity of .61 to.87 and specificity of .86 to .92 for major depression in primary care and medical outpatients [42,47,48].

The GAD-2 [43] assesses anxiety disorders via two items on anxiousness and worrying. The GAD-2 ranges from 0 (not at all) to 3 (nearly every day) with a cut-off score of ≥3. The GAD-2 shows a sensitivity of .89 and specificity of .83 for generalized anxiety disorder [43,49]. Internal consistency reliability is acceptable (α=.83) [50].

The SAM [44] is a nonverbal, cross-cultural rating scale [51] widely used for diverse groups of patients, such as for traumatized patients or refugees [52-55]. The individual can choose between five manikin pictures representing the present affective state on the dimensions valence (sad to happy), arousal (exited to calm), and dominance (weak to strong) [53].

The RHS-15 [45] was developed specifically for refugees and asylum seekers. It comprises 14 symptom items that can be answered on a 5-point Likert-Scale. A "distress thermometer" assesses perceived distress on a visual analog scale ranging from 0 to 10. The RHS-15 authors define a screening as positive with a cut-off score set at ≥12 for the symptom items or ≥5 if the distress thermometer is used. The RHS-15 distress thermometer can be used independently and has a good sensitivity (.81-.95) and specificity (.86-.89) [45].

Semistructured Qualitative Interviews

We used qualitative, semistructured interviews to obtain data illuminating the participants' experiences with the audio files and audio-based stabilizing and guided imagery techniques. The semistructured interviews comprised key questions, which were followed by probing questions; for further details, clarifying questions could be added. Multimedia Appendix 1 shows the interview guidelines and key questions.

Quantitative and Qualitative Data Analysis

Quantitative statistical analysis was carried out by using the Statistical Package for the Social Sciences program version 24 [56]. Demographic variables and baseline characteristics were analyzed using descriptive statistics (frequencies, means, and standard deviations). The qualitative interviews were digitally recorded and transcribed verbatim by independent co-workers using the guidelines for interview transcription presented in Mayring [57]. Statements regarding the refugees' practicing behavior were analyzed descriptively. The other research questions were analyzed with the software MAXQDA [58] following the principles of inductive content analysis described by Mayring [57]. Here, sentences are identified as the most basic units of meaning [59], summarized into relevant categories, and further grouped into main themes. The categories and main themes were subsequently discussed to reach

consensus or to be adjusted if required [57]. The T2 and follow-up statements have been summarized to facilitate presentation; noteworthy differences between T2 and follow-up interviews are explicitly highlighted.

Results

Sample Characteristics

A total of 83 patients attending the psychosocial walk-in clinic in PHV were referred to treatment with audio-based stabilizing and guided imagery techniques. Of these, 42 patients (50%) attended the introductory session (T1). All patients attending the introductory session consented to participate. They were aged between 19 and 51 years (mean 33.67, SD 8.3). All patients were invited to attend a booster session nine days later. The booster session (T2) was attended by 19 of 42 patients (45%). Reasons for non-attendance were reallocation (n=7, 17%), self-initiated departure from PHV (n=4, 10%), conflicting appointments (n=2, 5%), as well as reported as inactive (n=1, 2%), hospital stay (n=1, 2%), and imprisonment (n=1, 2%). We do not know why the remaining 7 (17%) patients did not attend. For the follow-up interviews, we attempted to contact as many of the 42 patients as possible. We were unable to reach 18 patients (43%) owing to inactive or incorrect phone numbers, while 7 (17%) patients were not successfully contacted after three attempts. In total, we conducted follow-up interviews with 19 of the participating patients (45%). Table 1 shows the sample characteristics for the total sample of 42 patients attending at least the first session.

Table 1. Sociodemographic characteristics and measurement at baseline for all participants (N=42).

Characteristic	Value
Gender, n (%)	
Male	25 (60)
Female	17 (40)
Region of origin, n (%)	
Middle East	23 (55)
Balkan Peninsula	3 (7)
North Africa	4 (9)
Sub-Sahara Africa	12 (29)
Medication, n (%)	
None	10 (24)
Antidepressant	27 (64)
Neuroleptics	3 (7)
No information	2 (5)
Religion, n (%)	
Christianity	10 (24)
Islam	30 (72)
Judaism	1 (2)
Other	1 (2)
Questionnaire scores, mean (SD)	
Primary Care PTSD[a] Screen for DSM-5[b]	3.93 (0.84)
PHQ-2[c]	3.87 (1.57)
GAD-2[d]	4.12 (1.56)
SAM[e]	
Valence	4.05 (1.10)
Arousal	2.79 (1.60)
Dominance	2.98 (1.12)
RHS-15[f] thermometer	7.21 (2.22)

[a]PTSD: posttraumatic stress disorder.

[b]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition.

[c]PHQ-2: Two-item Patient Health Questionnaire.

[d]GAD-2: General Anxiety Disorder questionnaire.

[e]SAM: Self-Assessment Manikin scale.

[f]RHS-15: Refugee Health Sreener-15.

Psychometric Baseline Assessment (T1)

The baseline scores for PTSD, depression, anxiety disorders, perceived distress, and affective state are presented in Table 1. On average, the patients reported four trauma experiences (mean 4.02, SD 1.71, range 1-7). The most frequently reported traumatic events were experiencing torture (27/42, 64%), being physically or sexually assaulted or abused (25/42, 60%), being imprisoned (23/42, 55%), seeing someone being killed or seriously injured (23/42, 55%), experiencing a war (20/42, 48%), and losing a loved one through homicide or suicide (16/42, 38%). Except for one patient, all patients confirmed three trauma symptoms and fulfilled the criteria of a possible PTSD according to the PC-PTSD-5. Furthermore, 28 (66%) patients reported suffering from four to five different PTSD symptoms. Additionally, 30 patients (71%) fulfilled the criteria of major depression, and 32 patients (76%) displayed symptoms of a generalized anxiety disorder. Furthermore, 38 patients (90%) scored positive on emotional distress assessed by the RHS-15 distress thermometer.

Descriptive Results of Practice Behavior Revealed by Interviews

Table 2 depicts the results of the practice behavior reported at T2 and follow-up.

Table 2. Statements regarding the self-practice behavior of stabilizing and guided imagery techniques.

Variables	T2[a]	Follow-up[b]
Frequency of practice, n (%)		
2-3/day	4 (21)	4 (21)
1/day	6 (31)	0 (0)
2-4/week	5 (26)	6 (31)
1/week	2 (11)	5 (26)
Stopped	2 (11)	4 (21)
Techniques were experienced as, n (%)		
Helpful	14 (74)	11 (58)
Partly helpful	3 (16)	7 (37)
Not helpful	2 (11)	1 (5)
Most helpful technique, n (%)		
Breathing	4 (21)	8 (42)
Body Scan	1 (5)	0 (0)
Guided imagery	10 (53)	8 (42)
No statement	4 (21)	3 (16)
Place for self-practice[c,d], n (%)		
Room	14 (82)	14 (93)
Outside	5 (29)	4 (27)
No statement	3 (18)	0 (0)
Time of self-practice[c,d], n (%)		
Morning	9 (53)	4 (27)
Daytime	9 (53	3 (20)
Evening	8 (47)	9 (60)
If symptoms were perceived	0 (0)	6 (40)
No statement	4 (24)	2 (13)

[a]N=19 patients, who attended the booster session (T2), interviews held face to face directly after the booster session.

[b]N=19 patients who were available via telephone two months after the booster session (follow-up), interviews conducted via telephone.

[c]These sections only include the answers of patients practicing the techniques: T2 n=17, follow-up n=15.

[d]Multiple answers were possible.

Results of the Inductive Content Analysis of the Interviews (T2 and Follow-Up)

We identified and coded 344 single statements from the T2 interviews and 334 single statements from the follow-up interviews. From these codes, we created twelve categories, which were then summarized into four main themes. Examples of each category are shown in Multimedia Appendix 2.

The Audio File as a Tool for Self-Practice (68 Codes)

In the T2 interviews, the patients gave feedback regarding the audio files at the technical and content levels.

Technical Difficulties With the Audio Files (38 Codes)

Most patients (n=15, 79%) denied any technical difficulties with the audio files, while 4 (21%) patients reported technical difficulties largely in connection with their smartphone. One patient could not listen to the audio file because his smartphone was stolen, and another's smartphone broke. One patient reported difficulties in retrieving the audio file on his smartphone. The fourth patient received the audio file via e-mail because his smartphone was unable to connect with the other device for transferring the audio file in the introductory face-to-face session.

Structure of the Audio Files (30 Codes)

Most of the patients stated that the voice and speed of the speech felt comfortable and natural for them. Some patients shared that they memorized the instructions. One patient remarked that the entire audio file might be too long as he sometimes falls asleep before the audio file has finished. Another appreciated the audio files but said it was not comparable to face-to-face contact with a therapist. Furthermore, some patients said that they were unable to follow the instruction of closing their eyes as it caused them discomfort.

Effects of Audio-Based Stabilizing and Guided Imagery Techniques (187 Codes)

The patients described stabilizing and guided imagery techniques via audio files as valuable for the relief of several clinical symptoms.

Arousal (44 Codes)

Most of the patients stated that stabilizing and guided imagery techniques helped them to feel more relaxed and calm immediately after practicing as well as in the long-term. One patient reported being able to regulate his heartbeat better. Others felt more comfortable with practice, both mentally and physically. Some patients noticed that they felt less stressed in the long run, which motivated them to continue practicing.

Tension and Sleep (38 Codes)

Most of the patients stated that stabilizing and guided imagery techniques helped them to reduce body tension immediately after practice. Some felt less pain in the lower back or neck. One patient shared that the techniques helped her with breathing because the tension and stress around her neck had improved. The patients further stated improved sleep during the night and a reduction in feelings of hardly ever getting a good night's sleep or merely a few hours of sleep. In follow-up interviews, some patients highlighted feeling as if they had more physical energy for their everyday lives.

Thoughts and Concentration (50 Codes)

At T2, most of the patients stated that the techniques helped them to focus on the here and now. Ten patients mentioned improvements in their concentration in the short- and long-term, such as when going to German language classes. Some mentioned feeling less forgetful than before or remembering more things than before. Some patients shared that they were able to escape from the problems of their everyday life while practicing the techniques. At follow-up, they frequently mentioned that they had fewer negative thoughts, fewer worries, and more positive thoughts compared to before practicing the techniques.

Mood and Empowerment (55 Codes)

Most patients stated feeling pleasure and relief during self-practice. Some reported feeling better while practicing the techniques, and others reported being in a good mood after practicing. One patient said that he felt increased inner freedom during the practice of the techniques; others felt renewed hope. Additionally, some patients reported feeling revitalized after practicing at follow-up. Three patients mentioned that they were able to get closer to and communicate better with other people. One participant said that he started going into town and engaging in small talk. One patient also reported feeling greater confidence.

Difficulties With the Audio-Based Stabilizing and Guided Imagery Techniques (110 Codes)

Patients reported internal and external challenges with self-practice of the stabilizing and guided imagery techniques.

Accommodation Situation (40 Codes)

The patients struggled with the living conditions in their current accommodation and reported feeling very uncomfortable. For most patients, it was difficult to cohabit with so many different people. Many patients said there was no space for them to calm down and use the techniques. If they were a family, the entire family shared one room without time and space to oneself; some shared with their children and had to take care of them. Many reported difficulties because of the very noisy environment on top of the lack of privacy.

Lack of Concentration (34 Codes)

The patients said it was difficult for them to concentrate on the instructions and remain focused. Several patients mentioned that recurring and painful thoughts, like worries about family members in their home country, the asylum procedure, or fear of persecution, would distract them. Others remarked that memories of bad events or periods sometimes came to mind during self-practice. Some patients stated they were so depressed and withdrawn that they did not want to practice or listen to anything.

Only Short-Term Relief (36 Codes)

Several patients indicated difficulties in accepting that the relief from the stabilizing and guided imaging techniques was only temporary. Some of the patients stated that, at the outset, they had too-high expectations of the long-lasting effect of the techniques. One patient reported feeling disappointed because he still felt burdened by symptoms despite having practiced the techniques. The patients mentioned that they sometimes felt that the techniques did not affect their well-being because they were unable to combat the problems they faced every day.

Appraisal of the Guided Imagery Technique 'The Inner Safe Place" (71 Codes)

Containing statements about positive and negative effects during practice as well as on its content, the guided imagery technique "The Inner Safe Place" was a central aspect within the participants' statements.

Positive Effects of the Guided Imagery Technique (23 Codes)

The participants said that they appreciated the guided technique "The Inner Safe Place." They described this technique as most helpful, liked, particularly positive, and stabilizing as it occasionally allowed them to be far away from their worries during practice. They felt a sense of freedom and security during this technique. One patient reported that he still felt safe three hours after having practiced "The Inner Safe Place." Another said he felt he had regained some hope through this technique.

Difficulties With the Guided Imagery Technique (32 Codes)

Patients also described various difficulties while practicing "The Inner Safe Place." The patients reported that they felt so comfortable imagining an inner safe place that they immediately felt burdened or even sad when they returned to reality. Some patients mentioned that it was difficult for them to visualize a place in their mind's eye. Another patient said that their inner safe place had shattered. The patients found that "The Inner Safe Place" technique could not always protect them from distractions and recurring thoughts during practice.

Content Statements Regarding the Guided Imagery Technique (16 Codes)

The patients shared some of their inner safe places with us. Some chose to be alone in their inner safe place, while others imagined having their family or friends there with them. If they had imagined themselves there alone, they usually thought of places in nature, like a beach, the ocean, or grassy plains. One patient said that he thought of Germany for his inner safe place, while another patient reported seeing himself cooking in a kitchen. Three patients reported imagining themselves either in their future or back in their childhood.

Discussion

Principal Findings

This study aimed to investigate the self-practice of stabilizing and guided imagery techniques via digital audio files in newly arrived refugees living in a state registration and reception center. The qualitative results show that the self-practice of audio-based stabilizing and guided imagery techniques can help traumatized individuals experience symptom relief in the early stages of arrival in their host country. Although some difficulties in practicing the exercise were reported, "The Inner Safe Place" was perceived as the most helpful technique in delivering positive inner images and feelings. However, self-practice of audio-based stabilizing and guided imagery techniques require a high degree of self-motivation and commitment from affected refugees. Finding such motivation and commitment may sometimes be overtaxing in light of impeding internal factors, such as lack of concentration, or postmigratory stressors, like lacking privacy in the accommodations.

Concerning aspects of the user application, the patients' statements indicate that using audio files is feasible and practicable. Issues only arose because of device problems. Various studies have described that the majority of refugees own a smartphone, yet not everyone in this group has access to the internet [60]. In the current study, the audio file was transferred to the patient's smartphone via USB cable or Bluetooth. The MP3 format appears to be a robust and simple format available offline for everyone. In a study by Zehetmair et al [14] assessing stabilizing and guided imagery techniques in a group setting, participating refugees voiced a desire for instructions to facilitate practicing the techniques between the face-to-face group sessions. By developing an audio-based approach in this study, every participant always had the technique instructions with them. The files are available for download [40].

Examining the participants' practicing behavior, we found that 52% of the patients described the daily (to several times daily) self-practice of the stabilizing and guided imagery techniques at T2. However, the statements indicated a decreasing tendency of self-practice and more flexible use of the audio files at follow-up. On the one hand, this may be explained by the high degree of self-motivation and commitment the self-practice of the techniques requires of the highly burdened patients causing them to abate over time. Furthermore, the booster session could also have been perceived as a monitoring element, so they may have practiced more before T2 than after when they practiced at their own accord without an anchoring/monitoring session. Nevertheless, at follow-up, still 52% of the patients practiced the stabilizing and guided imagery techniques frequently, with statements ranging from daily to several times a week.

The qualitative results of the T2 and follow-up interviews showed that audio-based stabilizing and guided imagery techniques were able to alleviate the symptoms associated with mental stress. The participating refugees described changes in arousal levels, mood and empowerment, thoughts and concentration, as well as sleep and tension. Perceived symptom changes through the practice of stabilizing and guided imagery techniques immediately after practice or over time are consistent with other outcomes [13]. Several studies encourage the use of stabilizing and guided imagery techniques in traumatized patients: for example, increases in conscious action (eg, consciously change the attention focus) and decreases in hyperarousal, emotional numbness, and perceived stress have been shown [61,62]. These changes lead to an experience of improved situation control and self-efficacy [10,61,63]. According to Reddemann [63], the combination of self-calming elements and elements of internal process recognition increases internal stabilization.

Nevertheless, the participants described external and internal impeding factors for the self-practice of stabilizing and guided imagery techniques via audio file. Newly arrived refugees are faced with problems such as frequent reallocation, uncertainties regarding their asylum application, as well as uncertainties about social, cultural, or future-related aspects. Moreover, refugees in refugee camps report inhuman living conditions, forced passivity, and waiting, all of which exacerbate their feelings of a lack of control regarding their current situation [64-66]. Accordingly, the patients in our study reported that the stabilizing and guided imagery techniques provided them with short term symptom relief but were unable to offer any solutions for their overwhelming and stressful everyday life's problems. Furthermore, our participants were not only preoccupied with worries about the present and future but also described experiencing trauma-related symptoms, which often made it difficult for them to stay focused or engage in the techniques. Although the audio files were designed to promote stabilization and distressing memory distancing, we cannot rule out that participants may be triggered by hearing the language of their home country, which is often strongly associated with trauma. Despite the high burden of traumatic symptoms, patients reported they were able to consciously shift the focus of their thoughts to the techniques.

"The Inner Safe Place" was considered the most helpful technique by the majority of patients. It helped patients experience positive feelings of security, hope, well-being, and freedom. These emotions are particularly meaningful and soothing for traumatized refugees who have been subjected to persecution, imprisonment, and torture [67,68]. Several of them were able to transfer the resulting positive feelings to their present situation. This ability can motivate the refugees to continue practicing the respective techniques. Guided imagery techniques aim at eliciting positive feelings that traumatized individuals often are unable to feel by creating an inner image that contrasts the traumatic re-experience imagery and is available in situations of overwhelming emotions and thoughts [10]. This further can facilitate self-distancing from the traumatic experiences and the associated symptoms of emotional overload and intrusions [69]. However, participants also reported encountering some difficulties with the technique as the positive inner images often clashed harshly with their realities in the present. On the one hand, this highlights that the patients were able to engage in the technique and felt comfortable at their imaginary inner place; on the other hand, it also underpins the effect of post migratory stressors by contrasting their current living situation in a state registration and reception center with the place they had imagined for themselves.

Limitations

This study has several limitations. First, the study relies on self-reports regarding both the questionnaire assessment as well as the qualitative interviews. Therefore, a tendency towards compliant or socially desirable behavior or answers cannot be ruled out. Nevertheless, we conducted 38 interviews (19 at T2, 19 at follow-up) with n=27 of the total sample of 42 patients participating in the study, which is sufficient to achieve saturation for main themes in heterogeneous populations [70]. Second, the patients were of different ages, education, and country of origin. This heterogeneity limits the generalizability of our results but also reflects the realities of the sample group. Third, the PHV psychosocial walk-in clinic's psychiatrists and psychotherapists recommended treatment with audio-based stabilizing and guided imagery techniques to *N*=83 refugees and asylum seekers. The study included 42 patients who took part in the study and attended the introductory session (T1), equivalent to a dropout rate of 50%. Unfortunately, we do not have any information about the reasons for non-attendance. However, among other reasons, logistical barriers, like reallocation to different accommodations, scheduling conflicts, missing the appointment, as well as treatment enrollment barriers, such as treatment-related insecurities and fears or symptom- and anticipated outcome-related motivational issues may be possible explanations Additionally, there was a high dropout rate of 55% between T1 and T2. As already experienced in previous studies [13,14,71], the registration and reception center presents a research setting that is inherently afflicted by frequent reallocation and high dropout rates. The high dropout rate limits the generalizability of our results and can further have facilitated the occurrence of the selection bias. Fourth, we cannot neglect that there might have been other circumstances that affected the patients' symptom load. So, even though the patients attributed symptom changes to the self-practice of the techniques, psychopharmaceutic or other stabilizing effects cannot be ruled out. Nevertheless, the positive effect attribution leads to increased commitment and motivation to continue practicing the techniques. Fifth, the impact of the respective narrators' voice characteristics (male, female, etc) was not further assessed in this study. However, none of the patients reported difficulties related to the narrator's voice but rather gave positive feedback regarding the audio file narration during their interviews. However, we cannot rule out that some patients might have encountered problems with the specific voice narrating their audio file.

Conclusions

This study explored refugees' perspectives on the self-practice of stabilizing and guided imagery techniques via digital audio files after they arrived in a state registration and reception center. The stabilizing and guided imagery techniques via audio files proved to be a practical and effective tool for self-help regardless of the patient's country of origin or ethnic background. In sum, the participants reported more positive effects than difficulties with the audio-based stabilizing and guided imagery techniques. They described effects on cognitive, emotional, physical, and empowerment levels; difficulties encountered were associated with internal impeding factors, such as lack of commitment, concentration, and only experiencing short-term relief, or external hindering factors, such as lack of privacy in their accommodations. Particularly the use of "The Inner Safe Place" technique was reported to produce pleasant and self-calming feelings. However, the experience sometimes stood in stark contrast to the patients' daily reality, making it difficult for some patients to cope with the difference. Overall, the qualitative data presented show promising results for the use of audio-based stabilizing and guided imagery techniques in this sample group.

Acknowledgments

This research was funded by the Ministry of Science, Research, and Arts, Baden-Wurttemberg, Germany [project identification No. D 100011720; AZ42-04 HV.MED (16)27/1].

Authors' Contributions

CZ, DK, and CN conceived the study. CZ, EN, AC, DK, LR, and CN participated in the design of the study. CZ, EN, and CL carried out the study. CN supervised the project. CZ and EN carried out the quantitative analysis. CZ performed the qualitative analysis. CZ, with assistance from EN and AC, and CN drafted the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guideline for interviews conducted after the booster session (T2-interview) and two months after the booster session (Follow-up).
[DOCX File , 17 KB - jmir_v22i9e17906_app1.docx]

Multimedia Appendix 2
Main themes, their categories, and example codes derived from qualitative inductive content analyses of 38 interviews with 27 participants.
[DOCX File , 20 KB - jmir_v22i9e17906_app2.docx]

References

1. Turrini G, Purgato M, Ballette F, Nosè M, Ostuzzi G, Barbui C. Common mental disorders in asylum seekers and refugees: umbrella review of prevalence and intervention studies. Int J Ment Health Syst 2017 Aug 25;11(1):51 [FREE Full text] [doi: 10.1186/s13033-017-0156-0] [Medline: 28855963]
2. Nikendei C, Greinacher A, Sack M. Therapy of post-traumatic stress disorders: overall treatment plan, treatment methods,their effectiveness. In: Borcsa M, Nikendei C, editors. Psychotherapy after flight and expulsion. Stuttgart: Georg Thieme Verlag; 2017:86-100.
3. Turrini G, Purgato M, Acarturk C, Anttila M, Au T, Ballette F, et al. Efficacy and acceptability of psychosocial interventions in asylum seekers and refugees: systematic review and meta-analysis. Epidemiol Psychiatr Sci 2019 Aug;28(4):376-388 [FREE Full text] [doi: 10.1017/S2045796019000027] [Medline: 30739625]
4. Nosè M, Ballette F, Bighelli I, Turrini G, Purgato M, Tol W, et al. Psychosocial interventions for post-traumatic stress disorder in refugees and asylum seekers resettled in high-income countries: Systematic review and meta-analysis. PLoS One 2017 Feb 2;12(2):e0171030 [FREE Full text] [doi: 10.1371/journal.pone.0171030] [Medline: 28151992]
5. Bozorgmehr K, Mohsenpour A, Saure D, Stock C, Loerbroks A, Joos S, et al. [Systematic review and evidence mapping of empirical studies on health status and medical care among refugees and asylum seekers in Germany (1990-2014)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2016 May;59(5):599-620. [doi: 10.1007/s00103-016-2336-5] [Medline: 27090243]
6. Reddemann L, Piedfort-Marin O. Stabilization in the treatment of complex post-traumatic stress disorders: Concepts and principles. European Journal of Trauma & Dissociation 2017 Jan;1(1):11-17. [doi: 10.1016/j.ejtd.2017.01.009]
7. Robertson MEA, Blumberg JM, Gratton JL, Walsh EG, Kayal H. A group-based approach to stabilisation and symptom management in a phased treatment model for refugees and asylum seekers. Eur J Psychotraumatol 2013;4:1-8 [FREE Full text] [doi: 10.3402/ejpt.v4i0.21407] [Medline: 24371509]
8. Ter Heide FJJ, Mooren TM, van de Schoot R, de Jongh A, Kleber RJ. Eye movement desensitisation and reprocessing therapy v. stabilisation as usual for refugees: randomised controlled trial. Br J Psychiatry 2016 Oct;209(4):311-318. [doi: 10.1192/bjp.bp.115.167775] [Medline: 26892849]
9. Farhood LF, Richa H, Massalkhi H. Group mental health interventions in civilian populations in war-conflict areas: a Lebanese pilot study. J Transcult Nurs 2014 Apr;25(2):176-182. [doi: 10.1177/1043659613515717] [Medline: 24464029]
10. Reddemann L. Imagination als heilsame Kraft. Ressourcen und Mitgefühl in der Behandlung von Traumafolgen. 20 ed. Stuttgart: Klett-Cotta-Verlag; 2017.
11. Kruse J, Joksimovic L, Cavka M, Wöller W, Schmitz N. Effects of trauma-focused psychotherapy upon war refugees. J Trauma Stress 2009 Dec;22(6):585-592. [doi: 10.1002/jts.20477] [Medline: 19960519]
12. Sarkadi A, Ådahl K, Stenvall E, Ssegonja R, Batti H, Gavra P, et al. Teaching Recovery Techniques: evaluation of a group intervention for unaccompanied refugee minors with symptoms of PTSD in Sweden. Eur Child Adolesc Psychiatry 2018 Apr;27(4):467-479 [FREE Full text] [doi: 10.1007/s00787-017-1093-9] [Medline: 29260422]
13. Zehetmair C, Tegeler I, Kaufmann C, Klippel A, Reddemann L, Junne F, et al. Stabilizing Techniques and Guided Imagery for Traumatized Male Refugees in a German State Registration and Reception Center: A Qualitative Study on a Psychotherapeutic Group Intervention. J Clin Med 2019 Jun 22;8(6):894 [FREE Full text] [doi: 10.3390/jcm8060894] [Medline: 31234487]
14. Zehetmair C, Kaufmann C, Tegeler I, Kindermann D, Junne F, Zipfel S, et al. Psychotherapeutic Group Intervention for Traumatized Male Refugees Using Imaginative Stabilization Techniques-A Pilot Study in a German Reception Center. Front Psychiatry 2018 Oct 29;9:533 [FREE Full text] [doi: 10.3389/fpsyt.2018.00533] [Medline: 30420815]
15. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]

16. Fledderus M, Bohlmeijer ET, Pieterse ME, Schreurs KMG. Acceptance and commitment therapy as guided self-help for psychological distress and positive mental health: a randomized controlled trial. Psychol Med 2012 Mar;42(3):485-495. [doi: 10.1017/S0033291711001206] [Medline: 21740624]
17. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
18. Menon V, Rajan TM, Sarkar S. Psychotherapeutic Applications of Mobile Phone-based Technologies: A Systematic Review of Current Research and Trends. Indian J Psychol Med 2017;39(1):4-11 [FREE Full text] [doi: 10.4103/0253-7176.198956] [Medline: 28250552]
19. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct 01;13(3):288-295. [doi: 10.1002/wps.20151]
20. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
21. Yu JS, Kuhn E, Miller KE, Taylor K. Smartphone apps for insomnia: examining existing apps' usability and adherence to evidence-based principles for insomnia management. Transl Behav Med 2019 Jan 01;9(1):110-119. [doi: 10.1093/tbm/iby014] [Medline: 30590862]
22. Wild J, Warnock-Parkes E, Grey N, Stott R, Wiedemann M, Canvin L, et al. Internet-delivered cognitive therapy for PTSD: a development pilot series. Eur J Psychotraumatol 2016;7:31019 [FREE Full text] [doi: 10.3402/ejpt.v7.31019] [Medline: 27837579]
23. Wild J, Ehlers A. Self-study assisted cognitive therapy for PTSD: a case study. Eur J Psychotraumatol 2010;1:1-11 [FREE Full text] [doi: 10.3402/ejpt.v1i0.5599] [Medline: 21994807]
24. Benight CC, Ruzek JI, Waldrep E. Internet interventions for traumatic stress: a review and theoretically based example. J Trauma Stress 2008 Dec;21(6):513-520. [doi: 10.1002/jts.20371] [Medline: 19107724]
25. Litz BT, Engel CC, Bryant RA, Papa A. A randomized, controlled proof-of-concept trial of an Internet-based, therapist-assisted self-management treatment for posttraumatic stress disorder. Am J Psychiatry 2007 Nov;164(11):1676-1683. [doi: 10.1176/appi.ajp.2007.06122057] [Medline: 17974932]
26. Klein B, Mitchell J, Abbott J, Shandley K, Austin D, Gilson K, et al. A therapist-assisted cognitive behavior therapy internet intervention for posttraumatic stress disorder: pre-, post- and 3-month follow-up results from an open trial. J Anxiety Disord 2010 Aug;24(6):635-644. [doi: 10.1016/j.janxdis.2010.04.005] [Medline: 20447802]
27. Knaevelsrud C, Brand J, Lange A, Ruwaard J, Wagner B. Web-based psychotherapy for posttraumatic stress disorder in war-traumatized Arab patients: randomized controlled trial. J Med Internet Res 2015 Mar 20;17(3):e71 [FREE Full text] [doi: 10.2196/jmir.3582] [Medline: 25799024]
28. Wagner B, Schulz W, Knaevelsrud C. Efficacy of an Internet-based intervention for posttraumatic stress disorder in Iraq: a pilot study. Psychiatry Res 2012 Jan 30;195(1-2):85-88. [doi: 10.1016/j.psychres.2011.07.026] [Medline: 21813187]
29. Brand BL, Schielke HJ, Putnam KT, Putnam FW, Loewenstein RJ, Myrick A, et al. An Online Educational Program for Individuals With Dissociative Disorders and Their Clinicians: 1-Year and 2-Year Follow-Up. J Trauma Stress 2019 Feb;32(1):156-166 [FREE Full text] [doi: 10.1002/jts.22370] [Medline: 30698858]
30. Jaycox LH, Ayer L, Vona P, Hehman C, Stein BD, Mahmud A, et al. Development and preliminary evaluation of a self-guided, internet-based tool for coping with stress and trauma: Life Improvement for Teens (LIFT). Psychol Serv 2019 Mar;16(1):85-94. [doi: 10.1037/ser0000277] [Medline: 30431308]
31. James EL, Bonsall MB, Hoppitt L, Tunbridge EM, Geddes JR, Milton AL, et al. Computer Game Play Reduces Intrusive Memories of Experimental Trauma via Reconsolidation-Update Mechanisms. Psychol Sci 2015 Aug;26(8):1201-1215 [FREE Full text] [doi: 10.1177/0956797615583071] [Medline: 26133572]
32. Horsch A, Vial Y, Favrod C, Harari MM, Blackwell SE, Watson P, et al. Reducing intrusive traumatic memories after emergency caesarean section: A proof-of-principle randomized controlled study. Behav Res Ther 2017 Jul;94:36-47 [FREE Full text] [doi: 10.1016/j.brat.2017.03.018] [Medline: 28453969]
33. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: A meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
34. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, Acceptability, and Potential Efficacy of the PTSD Coach App: A Pilot Randomized Controlled Trial With Community Trauma Survivors. Psychol Trauma 2016 Jan 25:384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
35. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
36. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]

37. Brown FL, Carswell K, Augustinavicius J, Adaku A, Leku MR, White RG, et al. Self Help Plus: study protocol for a cluster-randomised controlled trial of guided self-help with South Sudanese refugee women in Uganda. Glob Ment Health (Camb) 2018;5:e27 [FREE Full text] [doi: 10.1017/gmh.2018.17] [Medline: 30128163]
38. Manok N, Huhn D, Kohl RM, Ludwig M, Schweitzer J, Kaufmann C, et al. Ambulanz für Geflüchtete mit Traumafolgestörungen und psychischen Belastungen in einer Landeserstaufnahmeeinrichtung. Psychotherapeut 2017 Jun 26;62(4):333-340. [doi: 10.1007/s00278-017-0205-9]
39. Nikendei C, Huhn D, Adler G, von Rose PB, Eckstein TM, Fuchs B, et al. [Development and implementation of an outpatient clinic at an initial reception centre for asylum seekers in the German federal state of Baden-Wuerttemberg]. Z Evid Fortbild Qual Gesundhwes 2017 Oct;126:31-42. [doi: 10.1016/j.zefq.2017.07.011] [Medline: 28935457]
40. Kaufmann C, Zehetmair C, Nagy E, Reddemann L, Nikendei C. Heidelberger Standardübungen zur Stabilisierung von traumatisierten Geflüchteten. Handlungsanweisungen für imaginative Stabilisierungsübungen für traumatisierte Geflüchtete und Helfer. In: Heidelberg Clinical Standards. Heidelberg: Medizinische Fakultät Heidelberg; 2020. URL: www.heidelbergerklinischestandards.de
41. Prins A, Bovin M, Kimerling R, Kaloupek D, Marx B, Pless Kaiser A, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5). 2015. URL: https://www.ptsd.va.gov/professional/assessment/screens/pc-ptsd.asp [accessed 2020-04-27]
42. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
43. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
44. Lang P. Behavioral treatment and bio-behavioral assessment: Computer applications. In: Sidowski JB, Johnson JH, Williams TA, editors. Technology in Mental Health Care Delivery Systems. Norwood, NJ: Ablex; 1980:119-137.
45. Hollifield M, Verbillis-Kolp S, Farmer B, Toolson EC, Woldehaimanot T, Yamazaki J, et al. The Refugee Health Screener-15 (RHS-15): development and validation of an instrument for anxiety, depression, and PTSD in refugees. Gen Hosp Psychiatry 2013 Mar;35(2):202-209. [doi: 10.1016/j.genhosppsych.2012.12.002] [Medline: 23347455]
46. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample. J Gen Intern Med 2016 Oct 11;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]
47. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Mar;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
48. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010 Jul;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
49. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016 Mar;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
50. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
51. Morris J. Observations: SAM: The Self-Assessment Manikin. An Efficient Cross-Cultural Measurement of Emotional Response. Journal of Advertising Research 1995;35(6):63-68.
52. Kolassa I, Eckart C, Ruf M, Neuner F, de Quervain DJ, Elbert T. Lack of cortisol response in patients with posttraumatic stress disorder (PTSD) undergoing a diagnostic interview. BMC Psychiatry 2007 Oct 04;7:54 [FREE Full text] [doi: 10.1186/1471-244X-7-54] [Medline: 17916253]
53. Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9]
54. Spahic-Mihajlovic A, Crayton JW, Neafsey EJ. Selective numbing and hyperarousal in male and female Bosnian refugees with PTSD. J Anxiety Disord 2005 Jan;19(4):383-402. [doi: 10.1016/j.janxdis.2004.03.004] [Medline: 15721571]
55. Adenauer H, Pinösch S, Catani C, Gola H, Keil J, Kissler J, et al. Early processing of threat cues in posttraumatic stress disorder-evidence for a cortical vigilance-avoidance reaction. Biol Psychiatry 2010 Oct 01;68(5):451-458. [doi: 10.1016/j.biopsych.2010.05.015] [Medline: 20619396]
56. IBM Corp. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp; 2016.
57. Mayring P. Qualitative Inhaltsanalyse. Grundlagen und Techniken. 11 ed. Weinheim: Belz Verlag; 2010.
58. VERBI Software. MAXQDA. Software for qualitative data analysis, 11th version (released 11.1.4). Berlin: VERBI Software-Consult-Sozialforschung GmbH; 2012.
59. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2 ed. London: Saga Publications; 1998.

60. Burchert S, Alkneme MS, Bird M, Carswell K, Cuijpers P, Hansen P, et al. User-Centered App Adaptation of a Low-Intensity E-Mental Health Intervention for Syrian Refugees. Front Psychiatry 2018 Jan 25;9:663 [FREE Full text] [doi: 10.3389/fpsyt.2018.00663] [Medline: 30740065]
61. Fabach S. Achtsamkeit in der Praxis der Traumatherapie. Psychotherapie Forum 2017 May 29;22(1-2):24-30. [doi: 10.1007/s00729-017-0098-7]
62. Lampe A, Mitmansgruber H, Gast U, Schüssler G, Reddemann L. Therapieevaluation der Psychodynamisch Imaginativen Traumatherapie (PITT) im stationären Setting. NE 2008:189-197 Treatment outcome of psychodynamic trauma therapy in an inpatient setting. [doi: 10.5414/nepband22189]
63. Reddemann L. Achtsamkeit in der Behandlung von persönlichkeitsgestörten und traumatisierten Patienten. In: Reddemann L, editor. Kontexte von Achtsamkeit in der Psychotherapie. 1 ed. Stuttgart: Kohlhammer; 2011:101-117.
64. Bjertrup PJ, Bouhenia M, Mayaud P, Perrin C, Ben Farhat J, Blanchet K. A life in waiting: Refugees' mental health and narratives of social suffering after European Union border closures in March 2016. Soc Sci Med 2018 Oct;215:53-60 [FREE Full text] [doi: 10.1016/j.socscimed.2018.08.040] [Medline: 30205279]
65. Davies T, Isakjee A. Ruins of Empire: Refugees, race and the postcolonial geographies of European migrant camps. Geoforum 2019 Jun;102:214-217. [doi: 10.1016/j.geoforum.2018.09.031]
66. Porter M, Haslam N. Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: a meta-analysis. JAMA 2005 Aug 03;294(5):602-612. [doi: 10.1001/jama.294.5.602] [Medline: 16077055]
67. Kearney DJ, McManus C, Malte CA, Martinez ME, Felleman B, Simpson TL. Loving-Kindness Meditation and the Broaden-and-Build Theory of Positive Emotions Among Veterans With Posttraumatic Stress Disorder. Medical Care 2014;52:S32-S38. [doi: 10.1097/mlr.0000000000000221]
68. Passardi S, Peyk P, Rufer M, Plichta MM, Mueller-Pfeiffer C, Wingenbach TSH, et al. Impaired Recognition of Positive Emotions in Individuals with Posttraumatic Stress Disorder, Cumulative Traumatic Exposure, and Dissociation. Psychother Psychosom 2018;87(2):118-120. [doi: 10.1159/000486342] [Medline: 29495010]
69. Wöller W. Traumaspezifische Behandlungstechniken. PiD - Psychotherapie im Dialog 2006 Dec;7(4):405-407. [doi: 10.1055/s-2006-951833]
70. Hagaman AK, Wutich A. How Many Interviews Are Enough to Identify Metathemes in Multisited and Cross-cultural Research? Another Perspective on Guest, Bunce, and Johnson's (2006) Landmark Study. Field Methods 2016 Jul 24;29(1):23-41. [doi: 10.1177/1525822x16640447]
71. Nikendei C, Kindermann D, Brandenburg-Ceynowa H, Derreza-Greeven C, Zeyher V, Junne F, et al. Asylum seekers' mental health and treatment utilization in a three months follow-up study after transfer from a state registration-and reception-center in Germany. Health Policy 2019 Sep;123(9):864-872. [doi: 10.1016/j.healthpol.2019.07.008] [Medline: 31345581]

Abbreviations

CBT: cognitive behavioral therapy
GAD-2: General Anxiety Disorder questionnaire
PC-PTSD-5: Primary Care PTSD Screen for DSM-5
PHQ-2: two-item Patient Health Questionnaire
PHV: Patrick Henry Village
PTSD: posttraumatic stress disorder
RHS-15: Refugee Health Screener-15
SAM: Self-Assessment Manikin scale

Edited by G Eysenbach; submitted 21.01.20; peer-reviewed by T Loda, S Brigitte; comments to author 03.04.20; revised version received 20.04.20; accepted 30.04.20; published 23.09.20.

Please cite as:
Zehetmair C, Nagy E, Leetz C, Cranz A, Kindermann D, Reddemann L, Nikendei C
Self-Practice of Stabilizing and Guided Imagery Techniques for Traumatized Refugees via Digital Audio Files: Qualitative Study
J Med Internet Res 2020;22(9):e17906
URL: http://www.jmir.org/2020/9/e17906/
doi:10.2196/17906
PMID:32965229

Original Paper

An Internet-Based Intervention Augmented With a Diet and Physical Activity Consultation to Decrease the Risk of Dementia in At-Risk Adults in a Primary Care Setting: Pragmatic Randomized Controlled Trial

Kaarin J Anstey[1,2,3], PhD; Nicolas Cherbuin[3], PhD; Sarang Kim[4], PhD; Mitchell McMaster[3], BSc; Catherine D'Este[5,6], PhD; Nicola Lautenschlager[7], PhD, MD; George Rebok[8], PhD; Ian McRae[3], PhD; Susan J Torres[9], PhD; Kay L Cox[10], PhD; Constance Dimity Pond[11], PhD, MD

[1]School of Psychology, University of New South Wales, Sydney, Australia
[2]Neuroscience Research Australia, Sydney, Australia
[3]Centre for Research on Ageing, Health and Wellbeing, Australian National University, Canberra, Australia
[4]Wicking Dementia Resaerch & Education Centre, University of Tasmania, Hobart, Australia
[5]National Centre for Epidemiology and Public Health, Australian National University, Canberra, Australia
[6]School of Medicine and Public Health, University of Newcastle, Newcastle, Australia
[7]Academic Unit for Psychiatry of Old Age, University of Melbourne, Melbourne, Australia
[8]Johns Hopkins Centre on Aging and Health, Johns Hopkins University, Baltimore, MD, United States
[9]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne, Australia
[10]Medical School, University of Western Australia, Perth, Australia
[11]Department of General Practice, University of Newcastle, Newcastle, Australia

Corresponding Author:
Kaarin J Anstey, PhD
School of Psychology
University of New South Wales
Matthews Building
Kensington
Sydney, 2031
Australia
Phone: 61 412935746
Email: k.anstey@unsw.edu.au

Abstract

Background: There is a need to develop interventions to reduce the risk of dementia in the community by addressing lifestyle factors and chronic diseases over the adult life course.

Objective: This study aims to evaluate a multidomain dementia risk reduction intervention, Body Brain Life in General Practice (BBL-GP), targeting at-risk adults in primary care.

Methods: A pragmatic, parallel, three-arm randomized trial involving 125 adults aged 18 years or older (86/125, 68.8% female) with a BMI of ≥25 kg/m^2 or a chronic health condition recruited from general practices was conducted. The arms included (1) BBL-GP, a web-based intervention augmented with an in-person diet and physical activity consultation; (2) a single clinician–led group, Lifestyle Modification Program (LMP); and (3) a web-based control. The primary outcome was the Australian National University Alzheimer Disease Risk Index Short Form (ANU-ADRI-SF).

Results: Baseline assessments were conducted on 128 participants. A total of 125 participants were randomized to 3 groups (BBL-GP=42, LMP=41, and control=42). At immediate, week 18, week 36, and week 62 follow-ups, the completion rates were 43% (18/42), 57% (24/42), 48% (20/42), and 48% (20/42), respectively, for the BBL-GP group; 71% (29/41), 68% (28/41), 68% (28/41), and 51% (21/41), respectively, for the LMP group; and 62% (26/42), 69% (29/42), 60% (25/42), and 60% (25/42), respectively, for the control group. The primary outcome of the ANU-ADRI-SF score was lower for the BBL-GP group than the control group at all follow-ups. These comparisons were all significant at the 5% level for estimates adjusted for baseline differences

(immediate: difference in means −3.86, 95% CI −6.81 to −0.90, *P*=.01; week 18: difference in means −4.05, 95% CI −6.81 to −1.28, *P*<.001; week 36: difference in means −4.99, 95% CI −8.04 to −1.94, *P*<.001; and week 62: difference in means −4.62, 95% CI −7.62 to −1.62, *P*<.001).

Conclusions: A web-based multidomain dementia risk reduction program augmented with allied health consultations administered within the general practice context can reduce dementia risk exposure for at least 15 months. This study was limited by a small sample size, and replication on a larger sample with longer follow-up will strengthen the results.

Trial Registration: Australian clinical trials registration number (ACTRN): 12616000868482; https://anzctr.org.au/ACTRN12616000868482.aspx.

(J Med Internet Res 2020;22(9):e19431) doi:10.2196/19431

KEYWORDS

risk reduction behavior; dementia prevention & control; primary prevention; pragmatic clinical trial; prevention; primary care

Introduction

Background

In 2015, it was estimated that approximately 47 million people were living with dementia, and this number is expected to reach 131 million by 2050 [1]. It is estimated that a 10%-25% reduction in 7 key risk factors could prevent 1.1 to 3.0 million cases of dementia worldwide [2,3]. Furthermore, the risk factors for dementia are shared with risk factors for other chronic diseases [4]. Interventions that reduce the risk of dementia are therefore likely to also reduce the risk of other chronic diseases and promote healthy aging. In relation to cognition, prevention and delay of cognitive impairment will also benefit individuals and reduce health service usage [5-7].

To afford a greater chance of producing detectable changes during study time frames, the dementia research community has increasingly focused on multidomain interventions that address multiple risk factors simultaneously [8]. Alzheimer disease (AD) is the most common cause of dementia. Reviews of cohort studies have found that a similar set of risk factors is associated with AD and all-cause dementia [3]. Among individuals with high risk factor burden, cognitive decline can be reduced (and possibly reversed) by cardiovascular risk reduction, by increasing activities that stimulate and protect the brain, including cognitive [9], social [10], and physical activity [11], and by an appropriate diet [12,13]. Dementia and cardiovascular diseases share cardiometabolic and lifestyle risk factors [4]. Both cardiovascular disease and dementia risk reduction can be achieved by smoking cessation; increasing physical activity; adopting a healthy diet; reducing abnormally high blood pressure and cholesterol in midlife; and managing major depression, overweight or obesity in midlife, and diabetes if present [14]. Altogether, the literature supports the view that multidomain interventions aimed at reducing cardiometabolic risk and promoting behaviors protective against dementia will contribute to preventing cognitive decline, reducing the overall risk of dementia, and lowering depressive symptoms [15,16].

To bring about risk reduction and implement current guidelines on dementia risk reduction [17], there needs to be long-lasting behavioral changes in multiple areas. Moreover, strategies are required to identify adults with risk factors for dementia and to encourage them to make appropriate lifestyle changes. Achieving this requires developing pragmatic interventions that could be implemented in existing health or community settings [18,19] and using techniques such as goal setting, decreasing barriers to change, improving self-monitoring, increasing access to information, and maintaining motivation [20,21]. Therefore, this randomized controlled trial (RCT) investigated whether lifestyle management programs that offer not only health-promoting information but also practical behavior change techniques that can be implemented in daily life can reduce dementia risk.

Objectives

Assessment of cardiovascular risk factors is common in primary care, as is lifestyle advice. General practice is therefore a setting where the need for dementia risk reduction interventions may be identified and where interventions may be prescribed [22]. Studies have been conducted in a primary care setting with older adults to address the management of cardiovascular risk factors [23,24] to reduce the risk of dementia. However, the current program is the first of its kind to provide interventions to adults (aged ≥18 years) in the primary care setting, addressing both cardiovascular and lifestyle risk factors of dementia. Specifically, the objectives of this study are to determine, in community-dwelling adults who are overweight or have chronic health conditions, the effectiveness of (1) a web-based multidomain dementia risk reduction intervention (Body Brain Life in General Practice [BBL-GP]) developed by the authors, in comparison to (2) a single clinician–led group, Lifestyle Modification Program (LMP), developed by a general practice cooperative, and (3) a web-based active control condition developed by the authors.

Methods

Trial Design

The trial protocol for the BBL-GP study is published elsewhere [25]. Briefly, it is a three-arm, pragmatic, single-blind RCT to reduce the risk of cognitive decline in at-risk individuals attending a general practice. The study was conducted within the National Health Co-op (NHC), a bulk billing general practice (ie, patients' general practitioner [GP] fees are fully funded by a universal health cover Medicare scheme) in Canberra, with 8 clinics drawn from low- and middle-income areas. Canberra has a cold climate and a slightly higher than average level of education compared with the national average in Australia [26]. It also has a lower rate of bulk billing primary care services than

any other state or territory in Australia [27]. The NHC had an existing program called the LMP, to which practice physicians referred patients diagnosed with or at-risk of chronic health conditions. To integrate the trial into existing referral pathways at the clinics, the referral criteria for LMP were used as the trial eligibility criteria, and this was managed by the clinic's LMP coordinator.

Participants

Recruitment occurred via email, posters displayed in the NHC clinics, and the NHC website. Participants who expressed interest in the study were contacted by the LMP coordinator for a screening assessment of inclusion and exclusion criteria (outlined below). After eligible volunteers provided written informed consent, they were provided log-in details for the BBL-GP study website for web-based data collection purposes; a clinic appointment was also organized. Participant recruitment was undertaken over a 12-month period from mid-July 2016 until the end of July 2017, and the final follow-ups were completed in December 2018. Appointments were booked by the NHC staff. The research team was not involved in recruitment or assessment. The Australian National University Human Research Ethics Committee approved the study.

Eligibility Criteria

The eligibility criteria included being aged 18 years or older, being a resident in the Australian Capital Territory or surroundings, being a registered member of the NHC (which involves paying a joining fee of AUD $30 [US $21.90] and an annual fee of AUD $100 [US $73]), having home access to a computer and internet, having English fluency, having Australian permanent residency or citizenship (for universal health cover eligibility), being the sole member of a household taking part in the study, and having a chronic health condition that would make the participant eligible for the NHC lifestyle management program (eg, hypertension, heart disease, type 2 diabetes, osteoporosis, polycystic ovary syndrome, kidney or liver disease) or being overweight or obese (BMI ≥25 kg/m^2). All participants aged older than 60 years were required to score within the nonimpaired range on the Mini-Mental State Examination (MMSE; ≥25) [28] to be enrolled in the study.

Exclusion criteria included significant and unstable medical and psychiatric conditions precluding participation, sensory deficits, or mobility limitations that would prevent completion of the interventions, cognitive impairment (including AD or dementia), pregnancy, and previous participation in the LMP.

Interventions

The BBL-GP intervention is a 12-week program that includes (1) 8 web-based electronic learning modules on dementia literacy, risk factors, physical activity, nutrition, health condition management, cognitive activity, social activity, and mood and (2) tailored, face-to-face physical activity and nutrition sessions. BBL-GP participants were required to complete all 8 modules. The dementia literacy and risk factor modules were standardized across participants and were released at the rate of one per week after completion of the prior module, with the same timing for all participants. For the remaining modules, the content was tailored to participants' individual risk profiles as indicated by the baseline Australian National University Alzheimer Disease Risk Index Short Form (ANU-ADRI-SF; described below). The tailoring algorithms are included in Multimedia Appendix 1. These modules were also released at the rate of one per week, which was the same for all participants. The same exercise physiologist and 3 dietitians conducted all face-to-face sessions for the BBL-GP participants. The dietitians and exercise physiologist were all staff members within the NHC, but their services for this project were funded by a research grant. Participants at baseline who had unintentional weight loss or weight gain (±5 kg) or scored low on the Australian Recommended Food Score were seen by one of the dietitians (1-hour face-to-face) [25] and reviewed via phone at weeks 4, 12, and 20. Each participant received a single exercise physiology session that involved an evaluation of their current exercise level, fitness, and any preexisting health conditions and the design of a personal exercise program. Although follow-up via phone was planned, this was not conducted. The content of the dietitian's session was dietary education and advice to assist the participant in adapting their diet to a healthy diet in areas that were identified as unhealthy in the dietary questionnaire. Diet and exercise physiologist sessions varied depending on the participants' clinical needs and baseline measures. All sessions related to the BBL-GP were provided free of charge to the participants.

The LMP intervention was a face-to-face, practice-based, 6-week program provided by the NHC to its members for free. It included group sessions providing generic information on basic nutrition, meal planning, physical activity, health conditions, motivation and goals, medications, and sleep. All LMP sessions were conducted by an NHC clinical staff member using existing clinic resources. Each LMP group session involved up to 20 participants, and relevant sessions were conducted by one of the 3 dietitians and an exercise physiologist as for the BBL-GP group.

The 12-week active control arm involved weekly emails sent to participants that included links to information regarding lifestyle risk factors and disease management. At the end of the intervention, participants in the active control group received a 1-hour, face-to-face, group-based risk reduction workshop that provided the information contained in the BBL-GP intervention. This was held at the Australian National University and involved 1 to 6 participants and was delivered by KA with assistance from SK and MM.

Participants from all groups were sent a standard email when they were due for their follow-up appointment (18, 32, and 62 weeks). They were asked to complete their web-based assessments that were located on the BBL website, plus a diet quiz for which links were provided. If the web-based follow-up was not completed after 1 week, participants received a follow-up phone call from the research team with additional phone calls to request completion of missing sections of the assessments. The NHC receptionist arranged follow-up appointments for clinical assessments. The assessment was treated as missing after 3 follow-up phone calls. At the completion of the trial, participants were invited to complete a feedback questionnaire on the web, which included structured questions and one open-ended comment.

Randomization and Blinding

Following completion of the baseline assessment, consenting participants were randomized to one of the 3 intervention groups in a 1:1:1 ratio, stratified by sex and age group (18-49 vs ≥50 years) using permuted blocks of 6. Randomization was performed using the Sealed Envelope software [29] by a researcher not involved in the study, and the allocation sequence was provided to the project manager, who was also not involved in conducting assessments. The project manager then assigned the participant to their intervention group according to the schedule and notified the participant of their group allocation via email. Participants were not blinded to the group allocation.

Study conditions were presented as 3 alternative lifestyle interventions in recruitment materials. Participants were not informed of the intervention of interest to the researchers.

Web-based research data were stored on a server at the Australian National University, which was compliant with Australian data protection laws.

Outcomes

Web-based surveys and face-to-face visits to the NHC were conducted for the baseline evaluation and for the 18-, 36-, and 62-week follow-ups. Immediate follow-up (at the end of the formal program) was conducted on the web at week 7 for LMP and week 13 for the BBL-GP and control groups. These times were chosen to evaluate the immediate effects and the long-term effects of the intervention.

The web-based questionnaire included the primary outcome measure, the ANU-ADRI-SF [30]. The ANU-ADRI-SF is a shortened version of the ANU-ADRI [31]. Intraclass correlation coefficients assessing the agreement of individual components on the original and short-form ANU-ADRI were high (0.77-0.99). The ANU-ADRI has been externally validated in 3 cohort studies to predict dementia [32] and on a fourth cohort to predict mild cognitive impairment (MCI) [33]. The ANU-ADRI-SF questionnaire included secondary outcome measures of self-reported physical activity using the short version of the International Physical Activity Questionnaire (IPAQ) [34]. Additional questionnaires were the Pittsburgh Sleep Quality Index (PSQI) [35], the 12-item Short-Form Health Survey (SF-12) [36], and the Australian Recommended Food Score (ARFS) [37]. The web-based baseline assessment also included the Multidimensional Health Questionnaire [38] and the Adult Pre-exercise Screening System (APSS) [39] to identify individuals with acute or high-risk conditions or those who may be at higher risk of adverse events during exercise. The results of the APSS were provided to the exercise physiologist before the one-to-one exercise session with participants in the BBL-GP group.

Following completion of the web-based surveys, participants underwent a series of assessments at one of the 5 NHC clinics. Participants were assessed on sociodemographic characteristics, anthropometric measures, smoking status, alcohol consumption, blood pressure, cholesterol, high-density lipoprotein, blood glucose, social history (marital status and living arrangements), recreational activities, medical history, medications, and family history. From the information collected, the Framingham cardiovascular disease (CVD) risk score and the Australian Type 2 Diabetes Risk Assessment Tool risk scores were calculated. Body fat composition was also measured at the clinic using a bioelectrical impedance analysis, accounting for age, sex, height, and weight. Cognitive measures included the Trail Making Test A and B [40] and the Symbol Digit Modalities Test [41], which were completed via the BBL website on a desktop computer while participants were at the clinic in the presence of a clinic nurse. The assessments were conducted entirely by practice nurses. The MMSE was administered to individuals aged 60 years or older at baseline (see the Eligibility Criteria section). Objective physical activity was measured using the duration of moderate-to-vigorous physical activity (MVPA) via an ActiGraph monitor (GT9X Link) worn on the wrist for 7 days following the participants' clinic visit. MVPA was calculated as a continuous measure of activity registering 3 or more metabolic equivalents for 10 min or longer on the activity monitor. Self-reported physical activity was measured as part of the ANU-ADRI-SF using IPAQ categories for high, moderate, and low levels of physical activity, and relevant risk scores were assigned as part of ANU-ADRI.

Compliance

Compliance was measured by completion of web-based modules, following recommendations provided by the dietitian and exercise physiologist, or attendance at LMP group sessions, where relevant [25].

Harms and Other Adverse Events

Study team members monitored and managed any risks throughout the trial, including data handling and website access.

Participant Experience and Open-Text Feedback

At the conclusion of study participation, each participant completed a feedback questionnaire requiring them to rate the following aspects on a 5-point scale: overall study experience, perceived relevance of the intervention materials, level of interest in materials, and perceived repetition of intervention materials. Participants also rated whether participation was worthwhile and their perceived effectiveness of the group to which they were randomized.

Sample Size

We planned to recruit 240 participants to detect a difference in continuous outcomes between groups of SD of 0.5 based on ANU-ADRI (medium effect, based on a previous BBL project [42]), assuming 80% power and a two-tailed 5% significance level and allowing for 33% attrition over time (based on the previous LMP programs in the NHC).

Statistical Methods

For each outcome, analysis involved the generation of regression models (linear regression for continuous outcomes: primary outcome of ANU-ADRI and secondary outcomes of standardized cognition score, MVPA per week, ARFS, and PSQI; negative binomial regression for highly right-skewed continuous or count data: Center for Epidemiological Studies Depression Scale [CES-D]; and logistic regression for binary outcomes: proportion with sufficient physical activity), including age, sex, outcome at each time point, intervention group, time

point, and the interaction between the intervention group and time point. Mixed models were used to adjust for the correlation of outcome within individuals over time. The interaction term allowed comparison of outcomes between intervention groups at each time point, unadjusted (primary estimates) and adjusted (secondary estimates) for baseline values, using the *lincom* command in Stata (StataCorp LLC). For continuous outcomes, the differences in means between groups (unadjusted and adjusted for baseline differences in outcome) were obtained with 95% CIs. For binary outcomes, odds ratios with 95% CIs were obtained unadjusted for baseline differences, whereas the adjusted estimates or *intervention effect* was estimated as the ratio of odds ratios for the relevant time point and baseline: that is, the odds ratio at the time point of interest divided by the odds ratio at baseline, with 95% CIs. Outcomes with 95% CIs (means for continuous outcomes and proportions for binary outcomes) were graphed for the intervention group at each time point. A sensitivity analysis adjusted for (in addition to age and sex) sufficient physical activity, standardized cognition score, ARFS (diet), SF-12 Physical Component Score, SF-12 Mental Component Score, and Framingham CVD risk score at baseline was undertaken because these variables were considered to be potentially associated with the outcomes or attrition.

The primary analysis was a complete case analysis [43] for all outcome measures. The secondary analysis involved multiple imputations to account for missing data in a full intention-to-treat analysis. Variables included in the imputation model were sociodemographic characteristics (age, sex, and years of education at baseline), intervention group, values of the outcome at each time point, and baseline values of BMI and SF-12. Imputation was undertaken for all 3 intervention groups combined, rather than separately for each group, because of the small number of observations within each group and the number of observations with missing data. A total of 50 imputation data sets were generated, and an overall estimate was obtained according to the Rules by Rubin [44].

Results

Baseline Characteristics

In total, 125 patients were recruited (42 in the BBL-GP group, 41 in the LMP group, and 42 in the control group). Figure 1 shows the flow of participants through the trial and loss to follow-up for each of the 3 intervention groups. Overall, 31% (13/42) individuals in the BBL-GP group, 41% (17/41) in the LMP group, and 45% (19/42) in the control group provided data at all 5 time points. Although a smaller number of participants were recruited than planned, recruitment was stopped because of funding limitations. The planned intervention and follow-up durations did not change. A total of 66 individuals provided data at week 62 (20 in the BBL-GP group, 21 in the LMP group, and 25 in the control group). Baseline characteristics and outcomes at baseline are shown for the 3 intervention groups in Table 1. The 3 groups were generally well-matched, although the control group appeared to have a slightly lower BMI and ANU-ADRI-SF score and slightly higher scores on CES-D and the BBL-GP group had slightly more minutes of MVPA per week.

Figure 1. Participant flow through the trial, follow-up, and analysis. BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program; NHC: National Health Co-op.

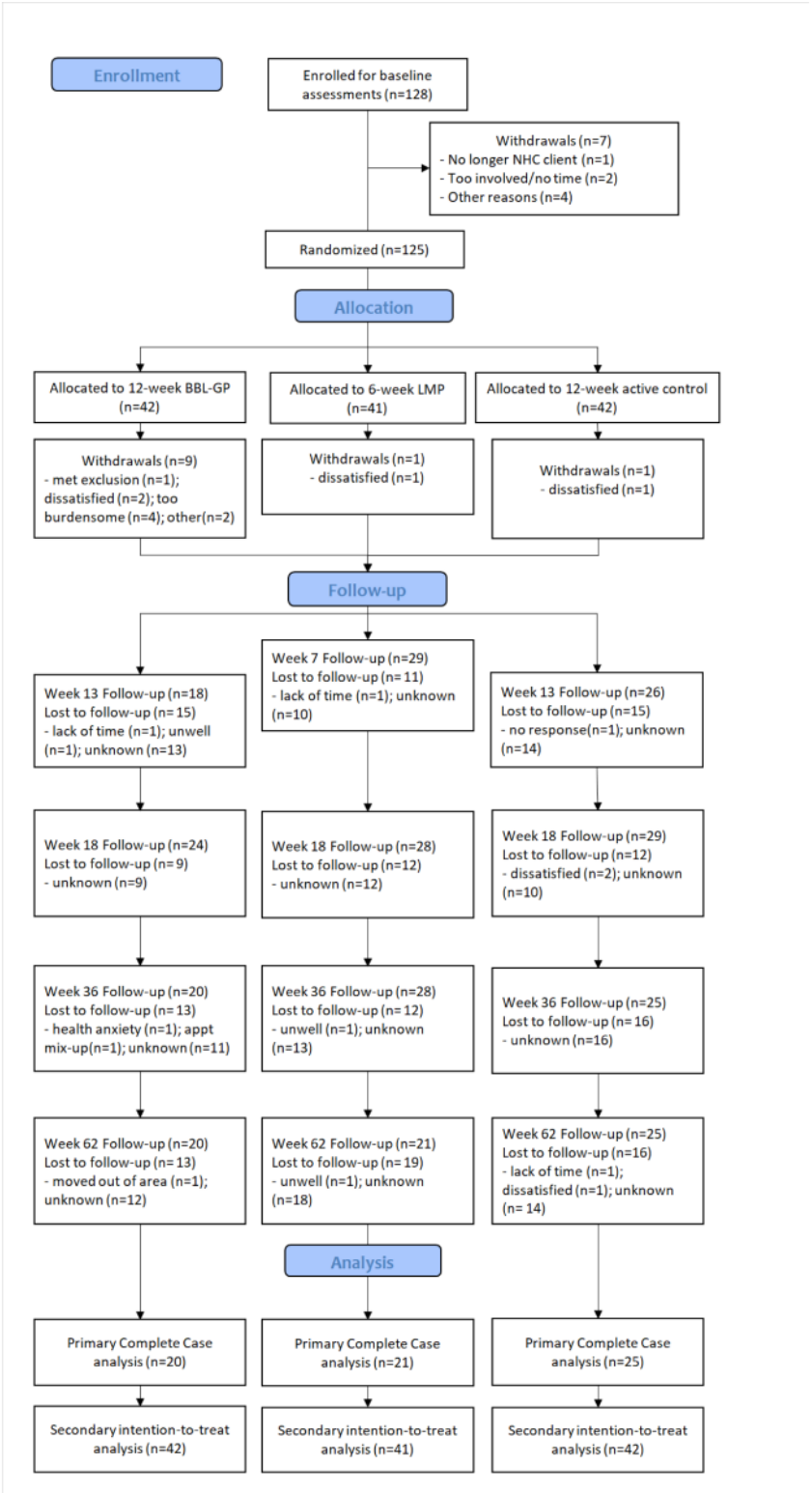


Table 1. Baseline characteristics by intervention group.

Variable	Body Brain Life in General Practice		Lifestyle Modification Program		Control	
	N[a]	Values, mean[b] (SD) or n (%)	N[a]	Values, mean[b] (SD) or n (%)	N[a]	Values, mean[b] (SD) or n (%)
Female	42	28 (67%)	41	27 (66%)	42	31 (74%)
Age, years	42	51.14 (14.24)	41	51.41 (11.69)	42	49.95 (14.03)
Education, years	42	15.51 (4.49)	41	16.41 (4.25)	42	16.11 (4.16)
BMI, kg/m^2	42	34.52 (7.11)	41	34.55 (6.65)	42	32.91 (7.36)
Australian National University Alzheimer Disease Risk Index Short Form	37	3.22 (7.13)	37	3.11 (7.19)	38	1.82 (6.36)
Cognition *z* score	41	−0.1 (1.01)	37	0.03 (1.06)	41	0.07 (0.94)
Total MVPA[c] per week[d]	40	1041.78 (528.97)	36	868.31 (409.62)	37	892.51 (420.01)
Sufficient physical activity	37	19 (51%)	37	20 (54%)	38	20 (53%)
Center for Epidemiological Studies Depression Scale score	42	7 (4,15)	41	7 (3,15)	42	10 (4,14)
Diet (Australian Recommended Food Score)	42	35.86 (8.84)	38	35.92 (10.42)	40	36 (8.36)
Sleep (Pittsburgh Sleep Quality Index)	41	6.63 (3.67)	41	7.46 (4.27)	41	8.02 (4.03)
12-item Short-Form Health Survey Physical Component Score	42	45.21 (8.62)	39	47.12 (10.49)	42	44.62 (7.93)
12-item Short-Form Health Survey Mental Component Score	42	44.95 (11.26)	39	45.42 (11.74)	42	42.81 (11.10)
Diabetes risk (Australian Type 2 Diabetes Risk Assessment Tool)	40	15.88 (6.53)	38	16.92 (5.66)	40	16.3 (5.56)
Framingham cardiovascular disease risk score	38	3.5 (0,7)	35	3.0 (1,7)	34	3.0 (0,6)

[a]Numbers may not add to the total sample size because of missing values.

[b]Median (Q1 and Q3) presented for the Center for Epidemiological Studies Depression Scale and Framingham cardiovascular disease risk scores.

[c]MVPA: moderate-to-vigorous physical activity.

[d]Total minutes of MVPA per week (activity registering 3 or more metabolic equivalents for at least 10 min).

Completion Rates and Missing Data

Multimedia Appendix 2 shows the characteristics of individuals who did (completers) and did not (noncompleters) provide data at all 5 time points. Although there were no statistically significant differences between completers and noncompleters, the sample size was small, and thus, statistical power was low for these comparisons. Some individuals did not provide information on the number of hours of each type of physical activity undertaken, resulting in missing data for both the IPAQ measure and the overall ANU-ADRI score.

Intervention Effects

Figures 2 to 8 and Multimedia Appendix 3 show the results of the comparison of outcomes between groups at each follow-up time from the models adjusted for age and sex. The interaction terms were statistically significant for the ANU-ADRI-SF and the PSQI, indicating significant group differences over time. However, there was only a clear and consistent pattern of better outcomes over time associated with the interventions for the primary outcome, ANU-ADRI-SF. This shows that the BBL-GP intervention reduced the risk of dementia. The difference between BBL-GP and the control group was only statistically significant at the 10% level for weeks 36 and 62 for ANU-ADRI-SF, although the trend for reduced risk scores remained (immediate: difference in means −2.30, 95% CI −5.93 to 1.34, *P*=.22; week 18: difference in means −2.49, 95% CI −5.99 to 1.02, *P*=.16; week 36: difference in means −3.43, 95% CI −7.16 to 0.29, *P*=.07; week 62: difference in means −3.06, 95% CI −6.71 to 0.60, *P*=.10). After adjusting for baseline differences in outcomes, in addition to age and sex, the differences in mean ANU-ADRI-SF between the BBL-GP and control groups increased (Multimedia Appendix 4) and were statistically significant at the 5% level at each follow-up time (immediate: difference in means −3.86, 95% CI −6.81 to −0.90, *P*=.01; week 18: difference in means −4.05, 95% CI −6.81 to −1.28, *P*<.001; week 36: difference in means −4.99, 95% CI −8.04 to −1.94, *P*<.001; week 62: difference in means −4.62, 95% CI −7.62 to −1.62, *P*<.001). Although the graph of mean diet score demonstrated some indication of higher scores for BBL-GP relative to the other group, this was not statistically significant, possibly because of low power.

Figure 2. Mean Australian National University Alzheimer Disease Risk Index Short Form with 95% CI, by time point and intervention group. ANU-ADRI-SF: Australian National University Alzheimer Disease Risk Index Short Form; BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program.

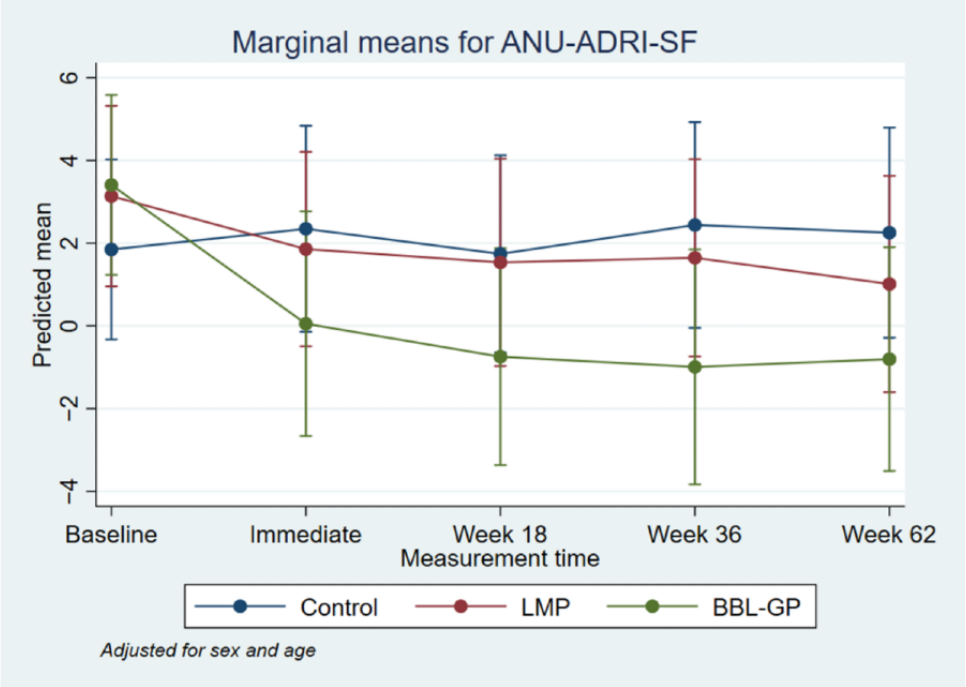


Figure 3. Mean standardized cognition score with 95% CI, by time point and intervention group. BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program.

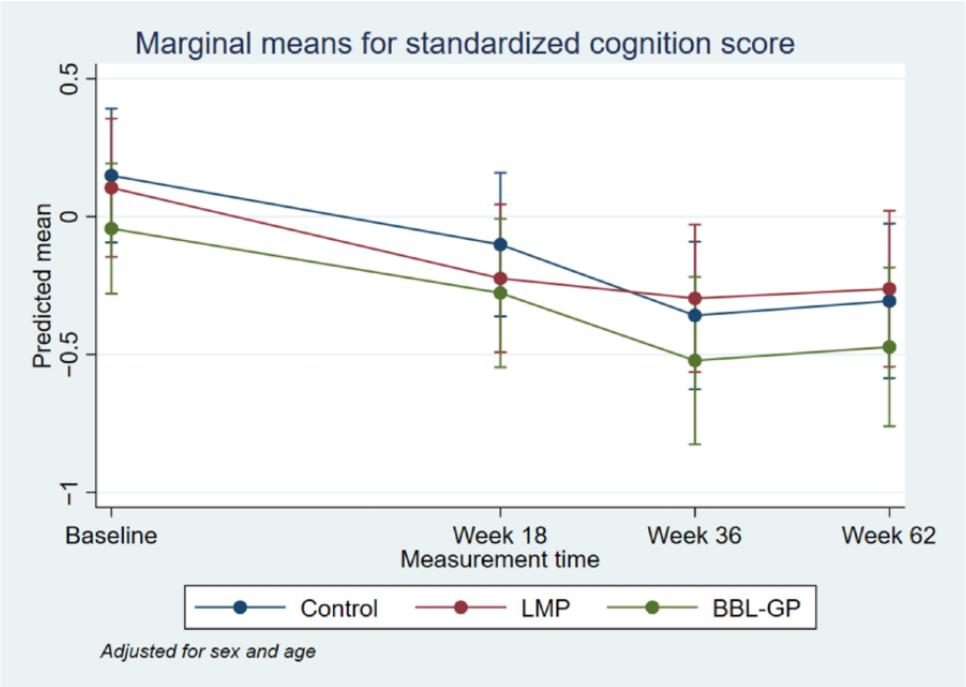


Figure 4. Mean moderate-to-vigorous physical activity with 95% CI, by time point and intervention group. BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program; MVPA: moderate-to-vigorous physical activity.

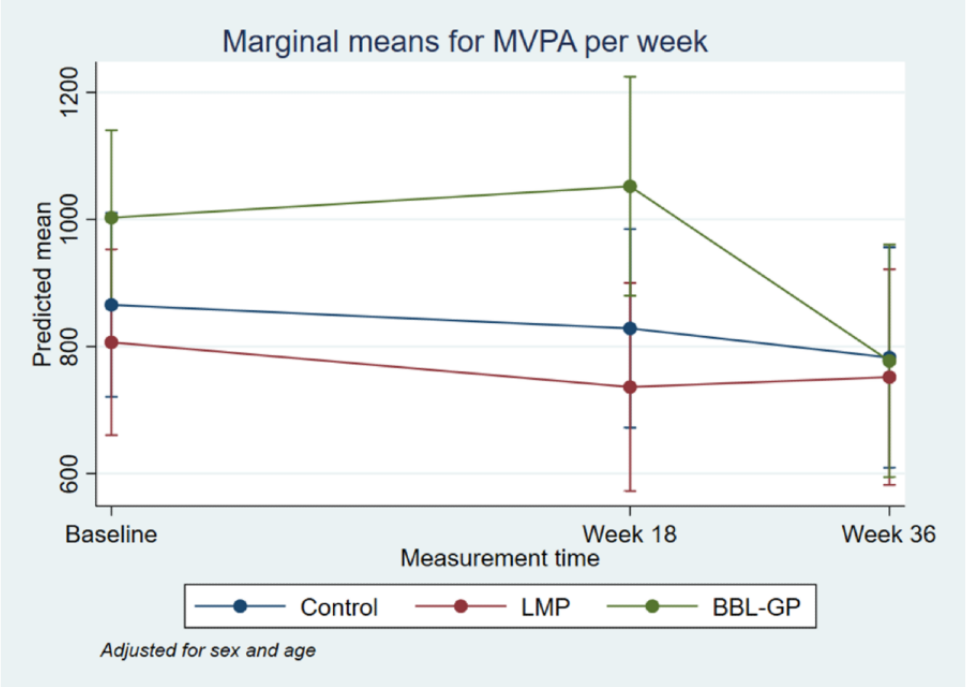


Figure 5. Proportion with sufficient physical activity with 95% CI, by time point and intervention group. Although the upper limit of some 95% CIs has been estimated as >1, it is not possible for a proportion to exceed 1. BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program; PA: physical activity.

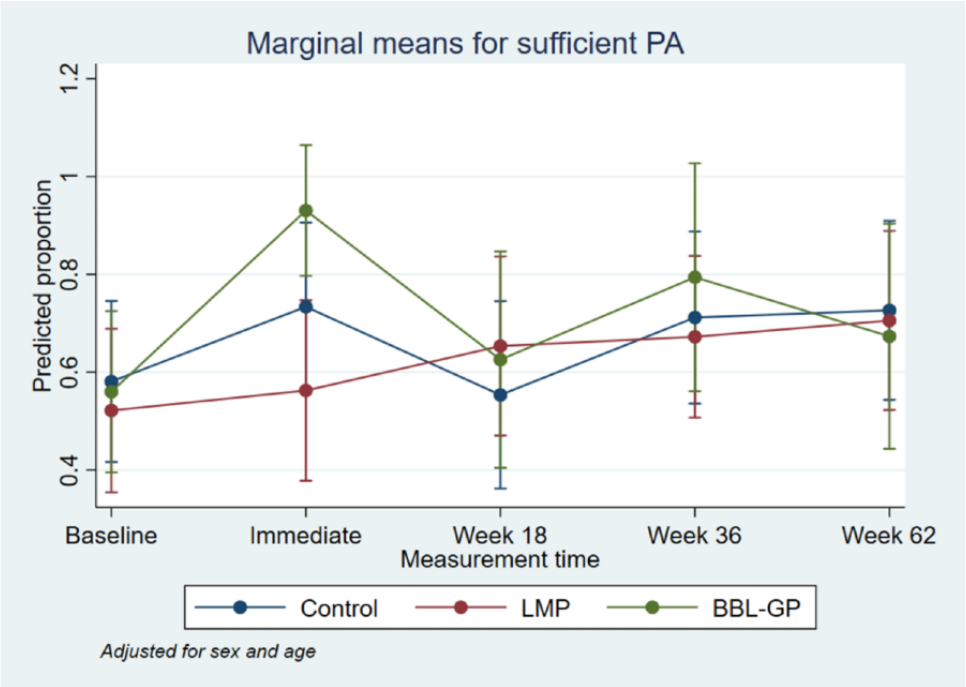


Figure 6. Mean Center for Epidemiological Studies Depression Scale with 95% CI, by time point and intervention group. BBL-GP: Body Brain Life in General Practice; CES-D: Center for Epidemiological Studies Depression Scale; LMP: Lifestyle Modification Program.

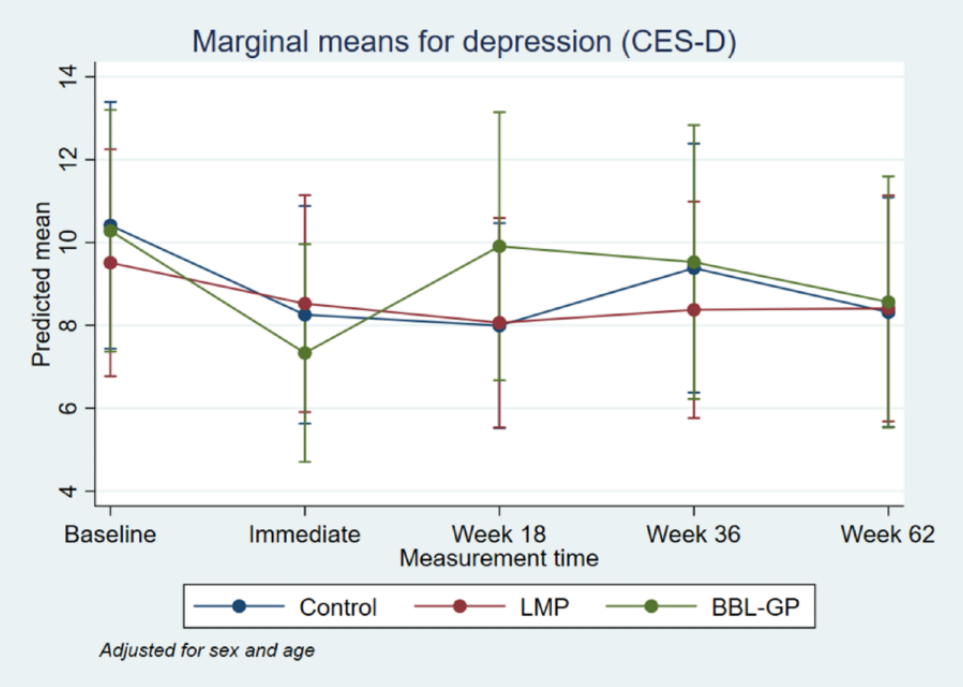


Figure 7. Mean Diet Score (ARFS) with 95% CI, by time point and intervention group. ARFS: Australian Recommended Food Score; BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program; PSQI: Pittsburgh Sleep Quality Index.

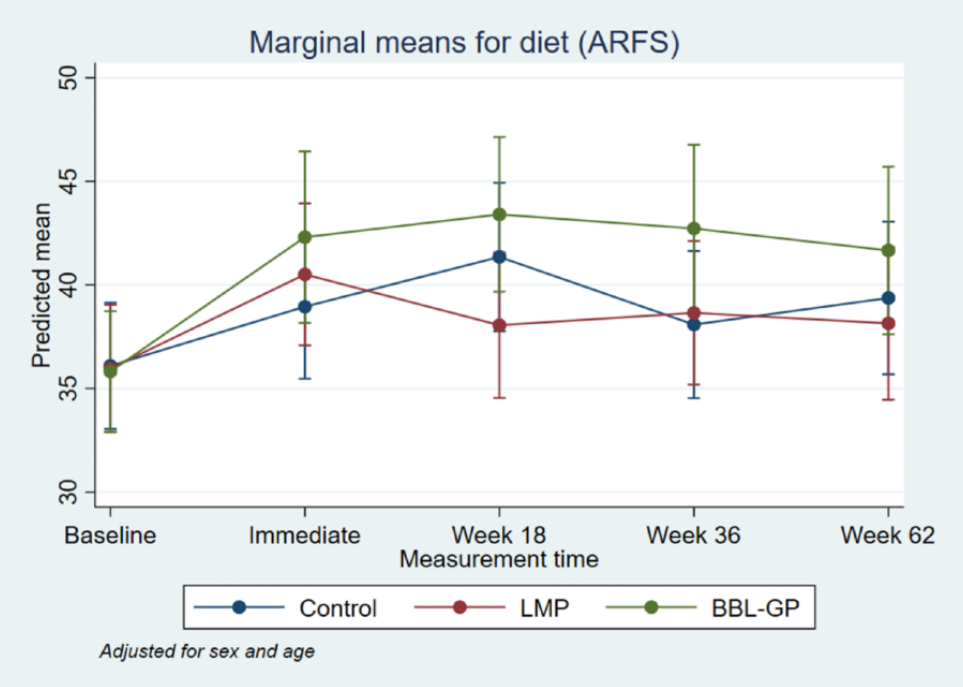
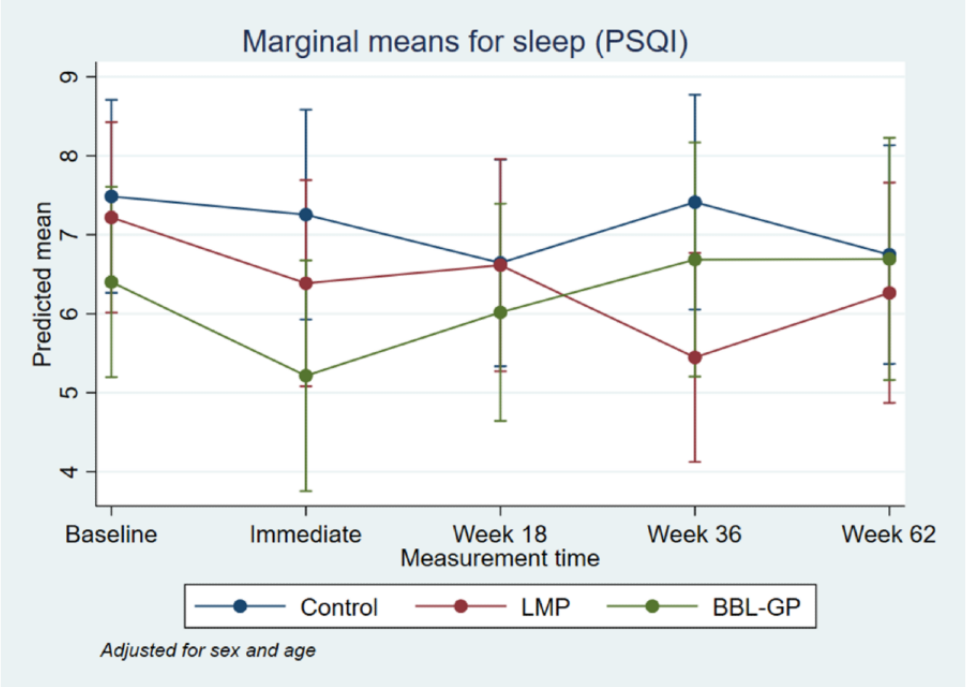


Figure 8. Mean Sleep Score (PSQI) with 95% CI, by time point and intervention group. BBL-GP: Body Brain Life in General Practice; LMP: Lifestyle Modification Program.



Sensitivity or Secondary Analyses

Sensitivity analyses adjusted for baseline physical activity, cognition, diet, mental and physical health, and CVD risk and secondary analyses using multiple imputation produced results that were broadly consistent with the primary analyses (although it was not possible to undertake multiple imputations for all secondary outcomes because of the lack of convergence of imputation and/or analysis models).

Compliance

Although adjustment for adherence (compliance) was originally planned [25], compliance was not considered in the analyses because of the smaller-than-anticipated number of individuals recruited and the amount of loss to follow-up.

Harms and Other Adverse Events

No adverse events were reported. There were no privacy breaches and no major technical problems. Occasional issues with accessing the website were reported and dealt with by the study team.

Participant Experience and Open-Text Feedback

A summary of the participant feedback on their experience of the intervention is found in Multimedia Appendix 5. A total of 21/29 (72.4%) participants who provided feedback rated their study experience as good or very good, and 20/26 (77%) felt their participation was worthwhile.

Discussion

Principal Findings

Our study aimed to evaluate the effectiveness of a multidomain lifestyle intervention in primary care to reduce the risk of dementia compared with a lifestyle management program that was already offered to primary care patients and an internet control condition. The significant group-by-time interaction in our main analysis showed that the multidomain BBL-GP lifestyle intervention was effective in reducing the risk of dementia for a period of at least 15 months. This builds on a growing body of research supporting the benefits of multidomain interventions that focus specifically on factors impacting brain health [8,45]. In contrast, a more generic healthy lifestyle management program that was not personalized did not result in dementia risk reduction. Population-based research has shown that participants aged 60 to 64 years with a one-point lower ANU-ADRI at baseline have an 8% lower chance of developing MCI or dementia [33,46] over a 12-year period. Our intervention achieved a lower ANU-ADRI-SF score of 4.62 (0.62 SD or equivalent to approximately 2 risk factors) at 62 weeks for the BBL-GP group relative to the control group in our analyses, which adjusted for baseline differences in the ANU-ADRI-SF. We therefore infer that maintenance of these benefits would translate into a significant shift in the risk of dementia at the population level, even if the effect size reduced over time.

Limitations and Strengths

Our study had a number of limitations. It was underpowered because of the smaller-than-anticipated sample and the pragmatic design whereby all the assessments and follow-ups were arranged by the NHC staff who assisted with the trial in addition to their usual workloads. This limited our capacity to evaluate the secondary outcomes. The sample attrition was relatively high with regard to attendance at follow-up appointments. We hypothesize that this was partly because of the limited role of the research team in arranging appointments for participants. The study was not resourced for vigilant follow-up of participants, which differs from our previous trial where a research assistant undertook this role [47]. This study utilized the desired model of the NHC and provided a seamless

patient experience, including the use of NHC staff in their usual locations conducting assessments. Future research is needed to develop mechanisms to increase participant follow-up compliance, co-designed with the general practice. In addition, more evidence is required to inform the design of interventions with regard to the number of allied health sessions and the efficacy of using web-based video to administer these sessions. A further study limitation was the relatively short length of follow-up.

The strengths of our study are that the trial was designed in partnership with a large primary care provider using the practice staff, space, and booking systems so that patients from the practice could complete the intervention within their usual settings. It was also implemented by the clinical staff without research training, which is more realistic than most research studies. It is therefore generalizable to the wider primary care setting in Australia. Our study also included 2 comparison conditions, including a lifestyle intervention that is used in clinical practice. Although it was an RCT, the study was implemented in such a way that the practice could continue the program if the BBL-GP site and allied health sessions could be supported and funded. Hence, we were able to demonstrate the benefits of a dementia-specific, multidomain lifestyle intervention. Our outcome measure, being a composite assessment, was sensitive to multidomain risk reduction.

Implications and Conclusions

This trial has several practical implications. The approach to patient selection was straightforward and compatible with usual care. GPs only need to identify patients who have a chronic condition and are appropriate for a lifestyle intervention to select those who will also benefit from a dementia risk reduction intervention. Participants were able to participate in the web-based component of the intervention in their own time and at minimum cost. The use of specialist practitioners to deliver the physical activity and nutrition sessions meant that these were tailored to the individual's clinical profile. This enabled broad inclusion criteria because the clinical risk was managed appropriately. Often, patients with multiple chronic conditions are excluded from trials, which reduces the ecological validity and generalizability of findings to the wider population. We conclude that dementia risk reduction in an adult population is feasible in the primary care setting and worthy of further research to establish scalable and sustainable models.

Acknowledgments

This work was supported by the National Health and Medical Research Council Centre of Research Excellence in Cognitive Health Grant #1100579; KA was funded by the NHMRC Fellowship #1102694, and the development of BBL-GP was funded by the NHMRC Dementia Collaborative Research Centres. This latter funding source had no role in the design, execution, analyses, interpretation of the data, or decision to submit results. The authors would like to thank the NHC and the study participants.

Conflicts of Interest

KA is an adviser for the StaySharp Platform, which is supported by the American Association of Retired Persons. DP has been on the advisory Board for Nutricia, which manufactures a nutritional drink to support early Alzheimers Disease. DP also received payment from a range of primary care providers including Primary Health Networks and the Mental Health Professionals Network, for providing education about dementia. DP also served on the "Silver Book" and "Red Book" advisory boards for the RACGP which provide advice about dementia prevention and management. There are no other declarations of interest.

Multimedia Appendix 1
Algorithm for tailoring risk factor intervention modules.
[XLSX File (Microsoft Excel File), 48 KB - jmir_v22i9e19431_app1.xlsx]

Multimedia Appendix 2
Baseline characteristics by completion.
[DOCX File , 26 KB - jmir_v22i9e19431_app2.docx]

Multimedia Appendix 3
Difference in outcomes between groups at each follow-up.
[DOCX File , 24 KB - jmir_v22i9e19431_app3.docx]

Multimedia Appendix 4
Difference in outcomes between groups at each follow-up, adjusted for baseline differences.
[DOCX File , 22 KB - jmir_v22i9e19431_app4.docx]

Multimedia Appendix 5
Summary of participant feedback.
[DOCX File , 13 KB - jmir_v22i9e19431_app5.docx]

References

1. Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer's Report. London, UK: Alzheimer's Disease International; 2015.
2. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]
3. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol 2011 Sep;10(9):819-828 [FREE Full text] [doi: 10.1016/S1474-4422(11)70072-2] [Medline: 21775213]
4. Peters R, Ee N, Peters J, Beckett N, Booth A, Rockwood K, et al. Common risk factors for major noncommunicable disease, a systematic overview of reviews and commentary: the implied potential for targeted risk reduction. Ther Adv Chronic Dis 2019;10:2040622319880392 [FREE Full text] [doi: 10.1177/2040622319880392] [Medline: 31662837]
5. O'Donoughue Jenkins L, Butterworth P, Anstey KJ. A longitudinal analysis of general practitioner service use by patients with mild cognitive disorders in Australia. Dement Geriatr Cogn Disord 2016;41(5-6):324-333. [doi: 10.1159/000447123] [Medline: 27414026]
6. Griffith LE, Gruneir A, Fisher K, Panjwani D, Gandhi S, Sheng L, et al. Patterns of health service use in community living older adults with dementia and comorbid conditions: a population-based retrospective cohort study in Ontario, Canada. BMC Geriatr 2016 Oct 26;16(1):177 [FREE Full text] [doi: 10.1186/s12877-016-0351-x] [Medline: 27784289]
7. Jacobs JM, Rottenberg Y, Cohen A, Stessman J. Physical activity and health service utilization among older people. J Am Med Dir Assoc 2013 Feb;14(2):125-129. [doi: 10.1016/j.jamda.2012.10.023] [Medline: 23218744]
8. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015 Jun 6;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
9. Wilson RS, Scherr PA, Schneider JA, Tang Y, Bennett DA. Relation of cognitive activity to risk of developing Alzheimer disease. Neurology 2007 Nov 13;69(20):1911-1920. [doi: 10.1212/01.wnl.0000271087.67782.cb] [Medline: 17596582]
10. Seeman TE, Miller-Martinez DM, Stein Merkin S, Lachman ME, Tun PA, Karlamangla AS. Histories of social engagement and adult cognition: midlife in the US study. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66(Suppl 1):i141-i152 [FREE Full text] [doi: 10.1093/geronb/gbq091] [Medline: 21196438]
11. Chang M, Jonsson PV, Snaedal J, Bjornsson S, Saczynski JS, Aspelund T, et al. The effect of midlife physical activity on cognitive function among older adults: AGES--Reykjavik study. J Gerontol A Biol Sci Med Sci 2010 Dec;65(12):1369-1374 [FREE Full text] [doi: 10.1093/gerona/glq152] [Medline: 20805238]
12. Hosking DE, Eramudugolla R, Cherbuin N, Anstey KJ. MIND not Mediterranean diet related to 12-year incidence of cognitive impairment in an Australian longitudinal cohort study. Alzheimers Dement 2019 Apr;15(4):581-589. [doi: 10.1016/j.jalz.2018.12.011] [Medline: 30826160]
13. Prince M, Albanese E, Guerchet M, Prina M. World Alzheimer Report 2014: Dementia and Risk Reduction. Alzheimer's Disease International. 2014. URL: https://www.alz.co.uk/research/world-report-2014 [accessed 2020-09-09]
14. Santos CY, Snyder PJ, Wu W, Zhang M, Echeverria A, Alber J. Pathophysiologic relationship between Alzheimer's disease, cerebrovascular disease, and cardiovascular risk: a review and synthesis. Alzheimers Dement (Amst) 2017;7:69-87 [FREE Full text] [doi: 10.1016/j.dadm.2017.01.005] [Medline: 28275702]
15. Hemingway H, Marmot M. Clinical evidence: psychosocial factors in the etiology and prognosis of coronary heart disease: systematic review of prospective cohort studies. West J Med 1999 Nov;171(5-6):342-350 [FREE Full text] [Medline: 18751201]
16. Wulsin LR, Singal BM. Do depressive symptoms increase the risk for the onset of coronary disease? A systematic quantitative review. Psychosom Med 2003;65(2):201-210. [doi: 10.1097/01.psy.0000058371.50240.e3] [Medline: 12651987]
17. World Health Organization. Risk Reduction of Cognitive Decline and Dementia: WHO Guidelines. Geneva, Switzerland: World Health Organization; 2019.
18. Goldstein MG, Whitlock EP, DePue J, Planning Committee of the Addressing Multiple Behavioral Risk Factors in Primary Care Project. Multiple behavioral risk factor interventions in primary care. Summary of research evidence. Am J Prev Med 2004 Aug;27(2 Suppl):61-79. [doi: 10.1016/j.amepre.2004.04.023] [Medline: 15275675]
19. Grunfeld E, Manca D, Moineddin R, Thorpe KE, Hoch JS, Campbell-Scherer D, BETTER Trial Investigators. Improving chronic disease prevention and screening in primary care: results of the BETTER pragmatic cluster randomized controlled trial. BMC Fam Pract 2013 Nov 20;14:175 [FREE Full text] [doi: 10.1186/1471-2296-14-175] [Medline: 24252125]
20. Middleton KR, Anton SD, Perri MG. Long-term adherence to health behavior change. Am J Lifestyle Med 2013;7(6):395-404 [FREE Full text] [doi: 10.1177/1559827613488867] [Medline: 27547170]
21. Locke EA. Motivation through conscious goal setting. Appl Prev Psychol 1996 Mar;5(2):117-124. [doi: 10.1016/s0962-1849(96)80005-9]
22. The Department of Health. National Primary Health Care Strategic Framework. Sydney, Australia: The Department of Health; 2013.

23. Shelley DR, Gepts T, Siman N, Nguyen AM, Cleland C, Cuthel AM, et al. Cardiovascular disease guideline adherence: an RCT using practice facilitation. Am J Prev Med 2020 May;58(5):683-690. [doi: 10.1016/j.amepre.2019.12.013] [Medline: 32067871]
24. van Charante EP, Richard E, Eurelings LS, van Dalen J, Ligthart SA, van Bussel EF, et al. Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial. Lancet 2016 Aug 20;388(10046):797-805. [doi: 10.1016/S0140-6736(16)30950-3] [Medline: 27474376]
25. Kim S, McMaster M, Torres S, Cox KL, Lautenschlager N, Rebok GW, et al. Protocol for a pragmatic randomised controlled trial of body brain life-general practice and a lifestyle modification programme to decrease dementia risk exposure in a primary care setting. BMJ Open 2018 Mar 17;8(3):e019329 [FREE Full text] [doi: 10.1136/bmjopen-2017-019329] [Medline: 29550779]
26. Fairweather-Schmidt AK, Anstey KJ. Prevalence of suicidal behaviours in two Australian general population surveys: methodological considerations when comparing across studies. Soc Psychiatry Psychiatr Epidemiol 2012 Apr;47(4):515-522. [doi: 10.1007/s00127-011-0369-5] [Medline: 21445626]
27. Department of Health. GP bulk-billing rates by electorate. In: 2019-2020 Supplementary Budget Estimates. Canberra: Australian Government; 2019.
28. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
29. Sealed Envelope. URL: https://www.sealedenvelope.com/ [accessed 2020-11-09]
30. Kim S, Cherbuin N, Anstey KJ. Assessing reliability of short and tick box forms of the ANU-ADRI: convenient alternatives of a self-report Alzheimer's disease risk assessment. Alzheimers Dement (N Y) 2016 Jun;2(2):93-98 [FREE Full text] [doi: 10.1016/j.trci.2016.03.001] [Medline: 29067296]
31. Anstey KJ, Cherbuin N, Herath PM. Development of a new method for assessing global risk of Alzheimer's disease for use in population health approaches to prevention. Prev Sci 2013 Aug;14(4):411-421 [FREE Full text] [doi: 10.1007/s11121-012-0313-2] [Medline: 23319292]
32. Anstey KJ, Cherbuin N, Herath PM, Qiu C, Kuller LH, Lopez OL, et al. A self-report risk index to predict occurrence of dementia in three independent cohorts of older adults: the ANU-ADRI. PLoS One 2014;9(1):e86141 [FREE Full text] [doi: 10.1371/journal.pone.0086141] [Medline: 24465922]
33. Andrews SJ, Eramudugolla R, Velez JI, Cherbuin N, Easteal S, Anstey KJ. Validating the role of the Australian national university Alzheimer's disease risk index (ANU-ADRI) and a genetic risk score in progression to cognitive impairment in a population-based cohort of older adults followed for 12 years. Alzheimers Res Ther 2017 Mar 4;9(1):16 [FREE Full text] [doi: 10.1186/s13195-017-0240-3] [Medline: 28259165]
34. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
35. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
36. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
37. Collins CE, Burrows TL, Rollo ME, Boggess MM, Watson JF, Guest M, et al. The comparative validity and reproducibility of a diet quality index for adults: the Australian recommended food score. Nutrients 2015 Jan 23;7(2):785-798 [FREE Full text] [doi: 10.3390/nu7020785] [Medline: 25625814]
38. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional health assessment questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum 1999 Oct;42(10):2220-2230 [FREE Full text] [doi: 10.1002/1529-0131(199910)42:10<2220::AID-ANR26>3.0.CO;2-5] [Medline: 10524697]
39. Norton K. New Australian Standard for Adult Pre-Exercise Screening. Informit. 2012. URL: https://search.informit.com.au/documentSummary;dn=808362433833839;res=IELHEA;type=pdf [accessed 2020-09-09]
40. Reitan RM. Trail making test results for normal and brain-damaged children. Percept Mot Skills 1971 Oct;33(2):575-581. [doi: 10.2466/pms.1971.33.2.575] [Medline: 5124116]
41. Smith A. (SDMT) Symbol Digit Modalities Test. Western Psychological Services. 1973. URL: https://www.wpspublish.com/sdmt-symbol-digit-modalities-test [accessed 2020-09-09]
42. Anstey KJ, Bahar-Fuchs A, Herath P, Rebok GW, Cherbuin N. A 12-week multidomain intervention versus active control to reduce risk of Alzheimer's disease: study protocol for a randomized controlled trial. Trials 2013 Feb 27;14:60 [FREE Full text] [doi: 10.1186/1745-6215-14-60] [Medline: 23442574]
43. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. Br Med J 2011 Feb 7;342:d40 [FREE Full text] [doi: 10.1136/bmj.d40] [Medline: 21300711]
44. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York, USA: John Wiley & Sons; 2004.

45. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. Nat Rev Neurol 2018 Nov;14(11):653-666. [doi: 10.1038/s41582-018-0070-3] [Medline: 30291317]
46. Cherbuin N, Shaw ME, Walsh E, Sachdev P, Anstey KJ. Validated Alzheimer's disease risk index (ANU-ADRI) is associated with smaller volumes in the default mode network in the early 60s. Brain Imaging Behav 2019 Feb;13(1):65-74 [FREE Full text] [doi: 10.1007/s11682-017-9789-5] [Medline: 29243120]
47. Anstey KJ, Bahar-Fuchs A, Herath P, Kim S, Burns R, Rebok GW, et al. Body brain life: a randomized controlled trial of an online dementia risk reduction intervention in middle-aged adults at risk of Alzheimer's disease. Alzheimers Dement (N Y) 2015 Jun;1(1):72-80 [FREE Full text] [doi: 10.1016/j.trci.2015.04.003] [Medline: 29854927]

Abbreviations

AD: Alzheimer disease
ANU-ADRI-SF: Australian National University Alzheimer Disease Risk Index Short Form
APSS: Adult Pre-exercise Screening System
ARFS: Australian Recommended Food Score
BBL-GP: Body Brain Life in General Practice
CES-D: Centre for Epidemiological Studies Depression Scale
CVD: cardiovascular disease
IPAQ: International Physical Activity Questionnaire
LMP: Lifestyle Modification Program
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MVPA: moderate-to-vigorous physical activity
NHC: National Health Co-op
PSQI: Pittsburgh Sleep Quality Index
RCT: randomized controlled trial
SF-12: 12-item Short-Form Health Survey

Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by K Rockwood, H van Marwijk, C Santos, J Parsons; comments to author 25.06.20; revised version received 28.07.20; accepted 03.08.20; published 24.09.20.

Please cite as:
Anstey KJ, Cherbuin N, Kim S, McMaster M, D'Este C, Lautenschlager N, Rebok G, McRae I, Torres SJ, Cox KL, Pond CD
An Internet-Based Intervention Augmented With a Diet and Physical Activity Consultation to Decrease the Risk of Dementia in At-Risk Adults in a Primary Care Setting: Pragmatic Randomized Controlled Trial
J Med Internet Res 2020;22(9):e19431
URL: http://www.jmir.org/2020/9/e19431/
doi:10.2196/19431
PMID:32969833

Original Paper

Effectiveness of Human Versus Computer-Based Instructions for Exercise on Physical Activity–Related Health Competence in Patients with Hip Osteoarthritis: Randomized Noninferiority Crossover Trial

Jennifer Durst[1,2], MA; Inka Roesel[1,3], MSc; Gorden Sudeck[2,4], PhD; Kai Sassenberg[5,6], Dr rer nat; Inga Krauss[1,2], Dr rer soc

[1]Department of Sports Medicine, University Hospital, University of Tuebingen, Tuebingen, Germany

[2]Interfaculty Research Institute for Sports and Physical Activity, University of Tuebingen, Tuebingen, Germany

[3]Institute for Clinical Epidemiology and Applied Biostatistics, University of Tuebingen, Tuebingen, Germany

[4]Institute of Sports Science, University of Tuebingen, Tuebingen, Germany

[5]Social Processes Lab, Leibniz-Institut für Wissensmedien, Tuebingen, Germany

[6]Faculty of Science, University of Tuebingen, Tuebingen, Germany

Corresponding Author:
Inga Krauss, Dr rer soc
Department of Sports Medicine
University Hospital
University of Tuebingen
Hoppe-Seyler-Str. 6
Tuebingen, 72076
Germany
Phone: 49 7071 2986486
Email: inga.krauss@med.uni-tuebingen.de

Abstract

Background: Hip and knee osteoarthritis is ranked as the 11th highest contributor to global disability. Exercise is a core treatment in osteoarthritis. The model for physical activity–related health competence describes possibilities to empower patients to perform physical exercises in the best possible health-promoting manner while taking into account their own physical condition. Face-to-face supervision is the gold standard for exercise guidance.

Objective: The aim of this study was to evaluate whether instruction and guidance via a digital app is not inferior to supervision by a physiotherapist with regard to movement quality, control competence for physical training, and exercise-specific self-efficacy.

Methods: Patients with clinically diagnosed hip osteoarthritis were recruited via print advertisements, emails and flyers. The intervention consisted of two identical training sessions with one exercise for mobility, two for strength, and one for balance. One session was guided by a physiotherapist and the other was guided by a fully automated tablet computer-based app. Both interventions took place at a university hospital. Outcomes were assessor-rated movement quality, and self-reported questionnaires on exercise-specific self-efficacy and control competence for physical training. Participants were randomly assigned to one of two treatment sequences. One sequence started with the app in the first session followed by the physiotherapist in the second session after a minimum washout phase of 27 days (AP group) and the other sequence occurred in the reverse order (PA group). Noninferiority was defined as a between-treatment effect (gIG)<0.2 in favor of the physiotherapist-guided training, including the upper confidence interval. Participants, assessors, and the statistician were neither blinded to the treatment nor to the treatment sequence.

Results: A total of 54 participants started the first training session (32 women, 22 men; mean age 62.4, SD 8.2 years). The treatment sequence groups were similar in size (PA: n=26; AP: n=28). Seven subjects did not attend the second training session (PA: n=3; AP: n=4). The app was found to be inferior to the physiotherapist in all outcomes considered, except for movement quality of the mobility exercise (gIG –0.13, 95% CI –0.41-0.16). In contrast to the two strengthening exercises in different positions (supine gIG 0.76, 95% CI 0.39-1.13; table gIG 1.19, 95% CI 0.84-1.55), movement quality of the balance exercise was close to noninferiority (gIG 0.15, 95% CI –0.17-0.48). Exercise-specific self-efficacy showed a strong effect in favor of the physiotherapist

(gIG 0.84, 95% CI 0.46-1.22). In terms of control competence for physical training, the app was only slightly inferior to the physiotherapist (gIG 0.18, 95% CI –0.14-0.50).

Conclusions: Despite its inferiority in almost all measures of interest, exercise-specific self-efficacy and control competence for physical training did improve in patients who used the digital app. Movement quality was acceptable for exercises that are easy to conduct and instruct. The digital app opens up possibilities as a supplementary tool to support patients in independent home training for less complex exercises; however, it cannot replace a physiotherapist.

Trial Registration: German Clinical Trial Register: DRKS00015759; http://www.drks.de/DRKS00015759

(J Med Internet Res 2020;22(9):e18233) doi:10.2196/18233

KEYWORDS

digital app; exercise; movement control; self-efficacy; control competence; mHealth; osteoarthritis; tablet

Introduction

Background

Osteoarthritis (OA) is characterized by pathological changes of the joint structure and is accompanied by pain and functional limitations. The global burden of hip and knee OA was ranked as the 11th highest contributor to worldwide disability [1]. The prevalence of OA increases with age. In Germany, almost 50% of all women and 30% of all men suffer from OA [2].

The promotion of physical activity in general, and exercise in particular, is highly relevant for patients with OA, as it has been shown to decrease pain and improve physical functioning [3]. Furthermore, physical inactivity is a particular characteristic of critical OA patients, with higher all-cause mortality compared to that of the general population, and is even more pronounced in people with severe walking disabilities [4]. Accordingly, international and national guidelines strongly recommend physical exercise as a core treatment in the management of hip and knee OA [5-8]. However, only a small to moderate proportion of patients with knee and hip OA meet physical activity guidelines [9]. An essential barrier for regular physical exercise could be that people with knee or hip OA suffer from pain at the weight-bearing joints and may further believe that physical activity is not beneficial or is even harmful to their joint [10]. Accordingly, OA patients experience fear of worsening their symptoms through exercise or by executing the exercises incorrectly [11,12].

Supervision seems to be an effective means to promote safe and correct exercise techniques, especially in the initial stages of the disease, and to ensure the right exercise dosage for each individual according to their physical ability and dosage principles from a training science perspective [13,14]. It is recommended that initial supervision and knowledge transfer on activity pacing be guided by a health care provider [7]. However, there is a considerable discrepancy between activity-related recommendations and practice in health care: 63% of patients with hip or knee OA aged 60 years or older that were part of a large German statutory health insurance company were treated with medication, whereas only 41% received physical therapy, including physiotherapy and exercise therapy [15]. Therefore, exercise-related advice from health care providers to community-dwelling people is not yet satisfactory. This is especially true for information related to modifying exercise behavior [16,17].

As a consequence, exploring alternate cost-efficient forms of delivery modes for people with limited access to therapeutic services appears to be indicated [13]. In this context, digital apps could be particularly suitable, since geographical proximity and time synchronicity between a therapist and patient are no longer necessary requirements for therapeutic interventions.

Digital interventions offer various types of interfaces and degrees of supervision. First, blended physiotherapy partially replaces face-to-face appointments with a digital app comprising instructions for graded activity, complementary unsupervised exercises, and behavior-change techniques. The physiotherapist plays an active role in patient interaction with regard to exercise content, as well as in supervision of intervention progress [18]. Second, the intervention can be delivered without the local presence of a physiotherapist, but with permission for regular Skype video sessions and individual support of the home training program in terms of exercise selection, instruction, and dosage adjustment. Exercise videos are also provided [19]. Other interventions still build upon personal supervision, yet without face-to-face contact. Web-based intervention programs are used that provide information, exercise instructions, an online physiotherapist (synchronous and asynchronous chats, and telephone consultations), and education regarding factors of relevance to OA, including lifestyle [20,21]. Third, there are fully automated apps that do not involve any interaction with a real person [22,23]. In the complete absence of personal supervision, a digital app should support the patients' tasks required for the correct and safe execution of exercises, as well as adequately control physical load and handling of pain as well as possible.

In recent years, a large number of digital exercise interventions have emerged, but there is a lack of studies comparing apps or digital interventions with human-delivered approaches in the field of chronic diseases. Hsu et al [24] compared a person-to-person and a digital-assisted approach for successful aging in older people. The results indicated that healthy behavior improved for the person-to-person group, although not significantly. However, the ability to search for health information improved in the digital-assisted group. Some other studies compared guided (email or telephone) and unguided mental health interventions. Guided interventions were significantly superior to unguided interventions [25]. However, none of the above-mentioned studies included patients with OA. Furthermore, systematic reviews have shown that the theoretical

foundation of digital interventions is often not very well-developed [26].

In the present study, we evaluated a fully automated digital app that was developed on the basis of a paper-based home exercise program, which emerged as one of the important features of an evidence-based exercise intervention for patients with hip OA [27,28]. The app was designed to be used without personal supervision and comprises a selection of videos with hip-specific exercises, information on dosage principles, as well as individualized feedback mechanisms (for further details see the Methods section).

Theoretical Underpinning of Digital Exercise Interventions

With regard to the effectiveness of exercise interventions in general, and for digital interventions in particular, a theory-based approach has been shown to optimize the targeted effects [29-31]. The digital app examined in this study was developed as part of the framework of the model of physical activity-related health competence (PAHCO) [32]. This model follows the general objective of exercise interventions for patients with chronic diseases to promote individual competence for an increasingly independent realization of regular physical exercise [33]. From this health educational perspective, it is particularly important to empower the patients to adequately master the requirements of specific exercises and to enable them to carry out physical exercise in the best possible health-effective and low-risk manner, taking into account their own physical condition.

The PAHCO model considers three subcompetences, each of which specifically helps in coping with demands that arise during the initiation and maintenance of physical exercise in a health-enhancing way. *Movement competence* relates to motor demands and is observable based on high movement quality while performing physical exercises. Movement competence is mainly based on motor abilities and skills but also requires the confidence to accomplish the task with one's own abilities (corresponding to task self-efficacy [34]). *Control competence* enables people to gear their own physical exercise to optimize health benefits and minimize health risks. It is mainly based on action-related knowledge but also requires the ability to perceive and interpret body signals (eg, to adjust intensities based on muscle soreness). *Self-regulation competence* ensures the required regularity of physical exercise. Thus, in theoretical terms, this subcompetence is closely related to social-cognitive theories of health behavior that address motivational and volitional determinants of exercise behavior (eg, outcome expectancies, behavioral self-efficacy; for an overview see Biddle et al [35]). The subcompetences can be considered as proxies for regular physical exercise, which in turn leads to the desired health benefits [32]. The promotion of these subcompetences should therefore be an important aim of an exercise intervention. Although pain, physical functioning, fear of movement, and self-efficacy have already been the focus of previous studies on the effectiveness of digitally assisted training interventions [23,36,37], there is a lack of studies related to proxies of health outcomes such as movement quality, control competence, and self-regulation.

Aims and Hypotheses

Based on the aforementioned knowledge deficits, the aim of this crossover study was to evaluate whether exercise instruction and guidance of one training session via a fully automated tablet computer-based app results in comparable benefits in subcompetences of PAHCO in comparison to one session supervised by a physiotherapist in subjects with hip OA. Our hypotheses were as follows: (1) movement quality of exercises guided by the app is not inferior to the movement quality under supervision by a physiotherapist; (2) the effect of the app on control competence is not inferior to the effect of the physiotherapist-guided intervention; and (3) the effects of the app on exercise self-efficacy (as a prerequisite of self-regulation competence) are not inferior to the effects of the physiotherapist-guided intervention.

Methods

Study Design

This study was designed as a randomized 2×2 crossover trial. The participants were randomly assigned to one of two exercise treatment sequences with an allocation ratio of 1:1. The AP sequence started with a training session instructed by a fully automated tablet computer-based digital app, followed by an intervention supervised by a physiotherapist, whereas the PA sequence started with the physiotherapist and the second training intervention was conducted with the app. The washout phase between the two interventions was set to range between 3 and 5 weeks. This time period seemed sufficient to washout relevant treatment effects of a single exercise. Ethical approval was obtained from the ethical committee of Tuebingen University Hospital. The study was registered in the German clinical trial register (DRKS00015759).

Participants

Community-dwelling individuals with diagnosed hip OA were recruited via advertisements in regional newspapers, as well as by emails sent to employees of the University of Tuebingen and Tuebingen University Hospital. In addition, flyers were distributed by orthopedic surgeons and physiotherapists. Interested participants were asked to call staff members for further information. According to ethical guidelines, the subjects were fully informed about the positive effects of exercise therapy for hip OA. They were also informed about the structure and details of the app (ie, feedback mechanisms) and the research questions. The screening for eligibility took place in the context of this phone call. Eligible people were then randomly allocated to one of the two treatment sequences and the dates for the two treatment sessions were scheduled. The treatment sequence was blinded until the first treatment session. Both treatments took place at Tuebingen University Hospital. Inclusion and exclusion criteria for participants are described in Textbox 1.

Textbox 1. Inclusion and exclusion criteria for the noninferiority study.

- Inclusion criteria
  - 50 years and older
  - self-reported lifetime prevalence of hip osteoarthritis diagnosed by a medical practitioner
  - informed consent to study participation
- Exclusion criteria (general)
  - comorbidities leading to major impairments in everyday life and representing contraindications for physical activities
  - self-reported acute illness
  - significant established osteoporosis requiring treatment, previous spontaneous or low-impact fracture
  - musculoskeletal surgery at the lower extremity within the last 3 months
  - regular use of gait aids (eg, walker, crutch)
  - insufficient German language skills for self-administered questionnaires
  - previous experience in hip exercise groups
- Exclusion criteria (in cases of an artificial joint replacement at the other hip or the knee joints)
  - artificial joint replacement at the knee or hip joint within the last 6 months, with unstable anchoring or with known radiological signs of implant loosening
  - current pain at rest or with activity due to artificial joint replacement
  - luxation as an adverse event of artificial hip joint replacement
  - acute joint inflammation at the knee or hip joint

Trial Interventions

Commonalities of Both Interventions

The interventions used in this study were extracted from an evidence-based 12-week exercise program that was specifically designed for patients with hip OA [28,38,39]. The home-based exercises (2 per week) as well as additional information related to exercise execution, graded exercise dosage, and exercise-related coping with pain are comprehensively described in a book, which also includes training and pain logs [27]. The contents of the book formed the basis for both interventions (physiotherapist and app). Four exemplary exercises of the home training sessions and their instructions were selected from the entire exercise program. The first exercise was related to mobility and movement learning, the second and third were strengthening exercises for the muscles surrounding the hip, and the fourth was a balance task. The training material included elastic rubber bands of different colors, weight cuffs, and soft aero pads. Details on exercise prescription, number of sets, and repetitions are given in Table 1 and Multimedia Appendix 1-4. Participants had to assess their current state of pain prior to each session.

Participants were asked to comment on perceived exertion and OA-related pain after each set using a 10-point Likert scale. The target value for intensity (last row of Table 1) was the same for participants using the app and those supervised by a physiotherapist. In the former case, the app introduced the user to the right intensity. In the latter case, the physiotherapist knew about the adequate intensity and guided the patients to adapt the training intensity if necessary. The instructions for the physiotherapist on how to deal with patients with increasing pain were the same as the algorithm included in the app (first: movement control, second: downgrading of intensity, third: skipping the exercise). A training session, regardless of whether guided by the app or the physiotherapist, lasted between 45 and 60 minutes. The same training area with a size of about 30 m^2 was used in both cases.

Table 1. Details of the exercises.

Detail	Pelvic tilt	Hip abduction	Hip extension	Balance
Name	mobility_seated	strength_supine	strength_table	balance_stance
Position	Seated	Supine	Table-supported stand	Step position
Kit	Stool	Mat, elastic band, pillow	Table, weight cuff, upper body padding	Balance pad
Aim	Pelvic, hip, and lumbar spine mobility; movement learning	Strength endurance	Strength endurance	Balance improvement, fall prevention
Description	Tilting the pelvis back and forth in the sagittal plane	An elastic band is placed below the knees. The feet are set. The knees are slowly tilted outward in the frontal plane	The upper body rests on the table and is supported by the arms. One leg is angled and slowly led back up in the sagittal plane. After adjusting the intensity, the exercise is performed with either an extended leg or an additional weight cuff	Step position both with open and closed eyes as well as with stable and unstable ground
Repetitions	30	20	25 (if exercise is performed with an additional weight cuff, the number of repetitions is reduced to 15)	15 seconds
Sets	2	3	3	6
Intensity	Low, no physical strain	6-7 after the last repetition, still able to perform the exercise correctly	6-7 after the last repetition, still able to perform the exercise correctly	Intensity is related to the difficulty of the balance task and is upgraded as long as the exercise can be performed correctly

Specification of Physiotherapist-Guided Exercises

The supervisor was a qualified physiotherapist with 5 years of work experience. She was responsible for (1) introducing exercises, (2) correcting deficient or improper execution of exercises, (3) adjusting intensity, and (4) instructing participants in the case of increasing pain. In addition, it was at the physiotherapist's judgement whether a participant should skip a simple intensity level and start directly with a more demanding variation.

Specification of App-Guided Exercises

The exercises were video-supported. Movement speeds were set using an auditory "click" sound and visually by the training partner in the video (see Multimedia Appendix 1-4). After each set, users were asked to comment on exercise-induced pain and its intensity via a digital visual analog scale on the interface monitor. To ensure a high quality of handling, the tablet was mounted to a holder before the start of the intervention. The participant trained independently and was only guided by the instructions of the app.

Characteristics of the App

The app was designed for a 9.7-inch (24.64 cm) Apple iPad and was developed by Ambigate GmbH (Tuebingen, Germany) according to the specifications of the authors. It is not open to the public. In line with the interventional implications of the PAHCO model, the content is basically a combination of practical exercises, cognitive and motor learning, and the processing of personal experience with movement [40]. Based on this, the following 5 different app components were compiled and individual intervention elements were further elaborated with reference to specific theoretical foundations: (1) technical introduction, (2) creation of an individual user profile, (3) pedagogical agent "Emil," (4) exercise introductions, and (5) feedback-based dose adjustments and further instructions. A detailed list of the app's components, specific elements, and the theoretical background of interaction principles can be found in Multimedia Appendix 5. An exemplary exercise flow and the underlying simplified algorithm of the software are described in Multimedia Appendix 6.

Videos and acoustic signals were implemented in the software to guide the different exercises and to support the participant during the exercises. The videos can be divided into different categories: (1) instruction video, (2) exercise video, (3) focus video, and (4) video for intensity adjustment. The characteristics of the videos, such as the perspective or benefits of close-ups, were tested in a pilot study during the app development process. The pilot study was conducted on 13 participants aged 50 years and older. This resulted in the implementation of close-ups in both focus videos and exercise videos. In addition, the camera's perspective was optimized so that starting positions and movements were optimally visible. In addition to the video structure, the pilot study also focused on the choice of adequate actors. The pilot study showed a tendency for gender and age to play only a minor role. It was much more important for the participants that the exercise was performed by the model in a clearly visible and correct manner. Hence, the actors in the videos are middle-aged, a man and a woman, and represent an average of the healthy population.

Outcome Measures

Baseline Data

Sociodemographic, anthropometric, personal, and OA-related variables were used to characterize the sample, including age, gender, educational level, work-related life situation, previous experience with exercise groups, importance of sport throughout life, and technical affinity, which were collected at baseline. In addition, fear of movement, OA-related symptoms, and physical activity in the preceding 4 weeks were collected at the beginning of the first (T0) and second (T1) treatment sessions.

Movement Quality

Movement quality was assessed by two independent raters. Both were research assistants with a bachelor degree in exercise science. They were introduced to the scoring procedure of movement quality using a rating sheet including different categories of movement quality for each exercise. The categories described the starting and ending position of all four exercises, as well as the movement sequence for the exercise. The categories are based on the description of the movement order and do not allow for high variance in the response. However, the aim was to query each movement step of the respective exercise. Raters had to classify whether the execution in a category was fulfilled, partly fulfilled, or not fulfilled, and whether all quality criteria points were met throughout all repetitions and the whole movement execution, temporarily, or not at all. The different categories were weighted according to their relevance. The weighting of the individual categories was based on the therapeutic relevance of the respective category for a correct, effective, and safe execution of the movement.

Most of the categories used a 3-point Likert scale with the grades as defined above. If a category could only be fulfilled or not fulfilled, a dichotomous scale was used. The judgments were based on video and audio recordings of the training sessions. Each session was videotaped from two perspectives, so that each movement could be assessed in the appropriate body axis. The raters were allowed to review the videos using the control unit of the video player, if necessary. The values for each exercise were transformed to a scale of 0-100%. All sets were included and averaged across raters. Different scores were calculated: the average value across all categories was defined as the primary outcome for movement quality, and the four average values for each of the exercises were defined as secondary outcomes for movement quality.

Exercise-Specific Self-Efficacy

Exercise-specific self-efficacy was measured with a self-assessed questionnaire based on the Multidimensional Self-Efficacy for Exercise Scale [41] before and after each session. The scale ranges from not at all safe (0) to absolutely safe (10). The 9 Likert-like items of the scale can be classified into three subcategories: task, coping, and scheduling. The overall score of exercise-specific self-efficacy was defined as the primary outcome measure of this scale. Overall exercise-specific self-efficacy and its subscales were tested for internal consistency using Cronbach α.

Control Competence for Physical Training

The assessment of control competence for physical training was conducted in accordance with the self-rating scale developed by Sudeck and Pfeifer [32]. The participants filled in a questionnaire related to control competence for physical training before and after both training sessions. Six Likert-scale items addressed the application of training-specific knowledge, and using the perception of body signals and perceived exertion to pace and structure exercise and training, targeting either endurance or strength. The Likert scale ranged from “totally disagree” (1) to “totally agree” (4). To further reflect specific demands for patients with hip OA, four additional items were created similar to the existing item set, especially focusing on hip-related exercises. The mean value across all items was used to calculate primary outcome measures. Control competence for physical training was tested for internal consistency using Cronbach α.

Sample Size

The sample size was predefined to a minimum of N=40 without a sample size calculation.

Randomization

A computer-generated randomized order list in blocks of 5 entries for each of the 2 treatment sequences (AP and PA) was created prior to the start of the study by the study personnel. Eligible participants were randomly assigned to the sequences at the end of the initial phone call. The study staff member entered the name of the caller into the list consecutively in the order of the calls. The list was visible to the study personnel with no further concealment. If randomized subjects canceled the first treatment session, the randomization slot was opened again and used for the first new incoming call of an eligible subject. The order in which raters assessed the movement quality of each participant was randomized for each rater and test day separately using the internet tool random.org [42].

Blinding

Participants, assessors (raters), and statisticians were not blinded to the treatment sequence or type of intervention. However, participants did not know which treatment sequence they were assigned to before the start of the first training session. The two raters of movement quality were not blinded to treatment sequence or intervention; however, they were not included in any other part of the study, such as data collection or data analysis.

Statistical Analysis

Baseline Data

Baseline sociodemographic, activity-related, and clinical characteristics are summarized for the overall study population and for each of the two treatment sequences separately (PA, AP). Categorical data are presented as absolute numbers and percentages, and continuous data are presented as mean (SD) or median (IQR), as appropriate. Differences between treatment sequences were compared using Pearson chi-square test for categorical data, unpaired Student *t* test, or Mann-Whitney *U* test. The latter was used if the assumption of normally distributed data was violated.

Treatment Effects

This study used a 2×2 crossover design to test for noninferiority of the app versus physiotherapist with respect to the primary and secondary outcome measures as outlined above. The following effects must be considered in a crossover design: direct treatment effect (τ), period effect (π), carryover effect (λ), and random subject effect (nested within sequence of treatment order). Therefore, a mixed-model analysis of variance (ANOVA) was used with the fixed factors treatment (physiotherapist, app), period (T1, T2), and treatment sequence (PA, AP), with the latter indicating potential carryover effects.

Treatment effects were averaged over the levels of period and treatment sequence. They were used to calculate crossover standardized effect sizes comparable to independent group designs that were adjusted for medium sample sizes (Hedges g_{IG}) [43]. Two-sided 95% CIs were further calculated for the resulting g_{IG}. Positive treatment effects (τ) and effect sizes (g_{IG}) favor the conventional treatment physiotherapy.

Noninferiority of the app to physiotherapist-guided treatment was established if the upper limit of the 95% CI of g_{IG} was entirely below the predefined noninferiority margin, Δ=0.2 (g_{IG} + 95% CI < 0.2). Determining a margin that defines the largest clinically acceptable difference between two treatments is a critical step of a noninferiority trial and should account for both statistical considerations (ie, estimates based on prior studies and clinical judgement). As there is no established recommended Δ value for a noninferiority trial comparing app-based interventions with conventional physiotherapy sessions and no prior studies with similar primary outcome measures are available (to the best of our knowledge), we chose crossover standardized effect sizes equal or greater to an absolute value of 0.2 to be relevant in accordance with the definition of an effect size of Cohen *d*=0.2 to be small with negligible clinical importance. Effect sizes were further categorized with Cohen *d*=0.5 representing a medium effect size and *d*=0.8 representing a large effect size.

Outcome measures of movement quality were assessed only once during each intervention and were not further adjusted for any variable. Outcome measures of exercise-specific self-efficacy and control competence for physical training were assessed directly prior to and after each treatment for both periods (T1, T2). Post-pre differences (ie, change from baseline) were used as input variables for the mixed-model ANOVA for these measures.

Rater Agreement for Movement Quality

Rater agreement was assessed by calculating the percentage of exercise-related movement quality categories in which raters completely agreed. Values were averaged across all sets for each exercise and were considered separately for the app and physiotherapist treatments.

Sensitivity Analysis for Outcomes with Carryover Effects

If the mixed-model ANOVA exhibited a significant carryover effect, an additional analysis of treatment effects was conducted for period T1 only with a simple *t* test for unpaired samples. Effect sizes were calculated using Hedges *g*. Two-sided 95% CIs were further calculated for the resulting *g*.

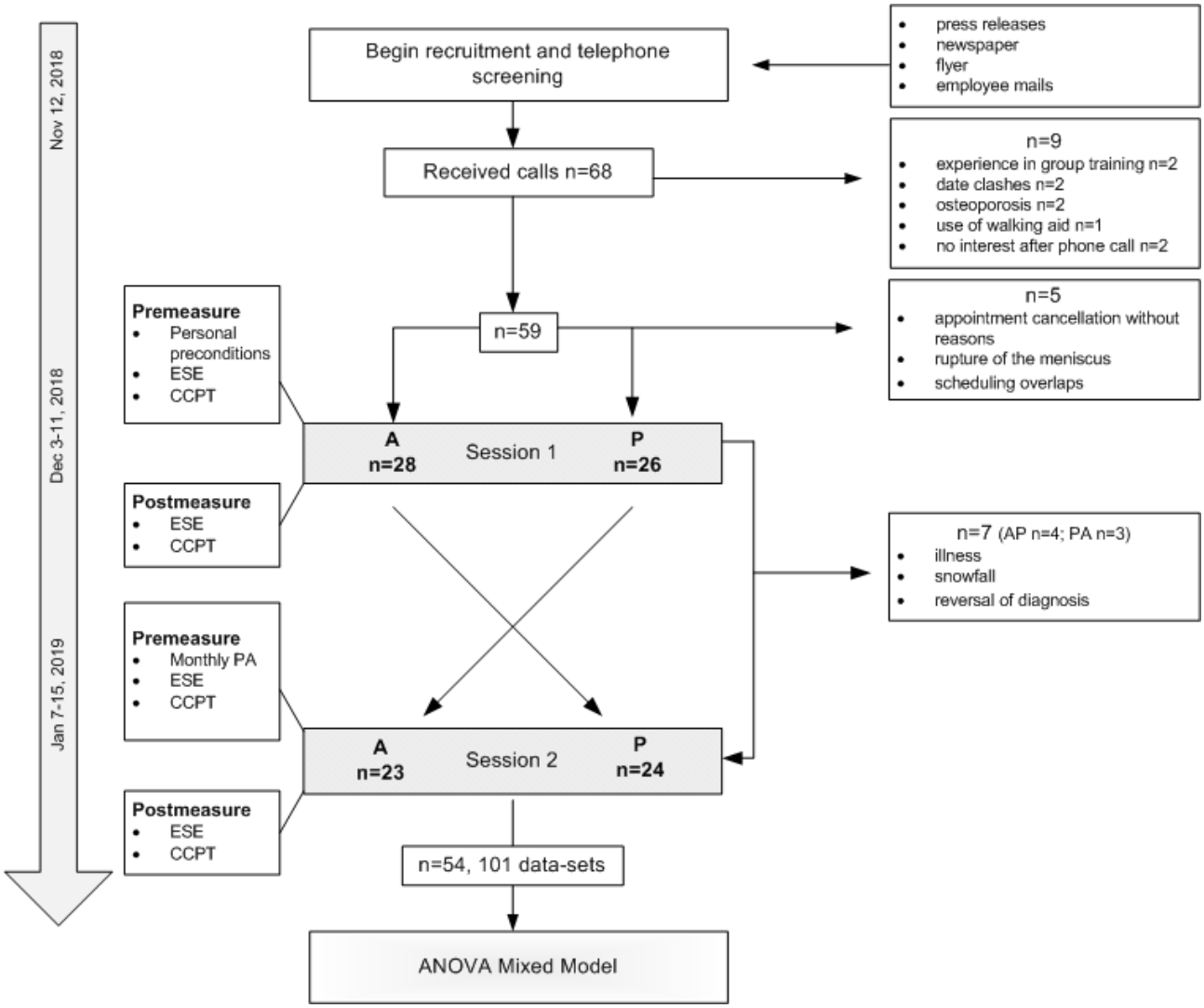
The level of statistical significance was set at the conventional level of α=.05. All data were analyzed using SPSS IBM Version 25 and R version 3.6.1.

Results

Participants

Recruitment started in November 2018. Sixty-eight people contacted the study staff by phone, and 59 were deemed eligible and received a follow-up email, including written study information, details on measuring time points, and travel directions. However, five potential participants cancelled the first training appointment due to physical problems or overlapping appointments. Finally, 54 people completed the first training session. Seven participants could not attend the second session. Further details on participants flow are depicted in Figure 1. Data collection started on December 3, 2018 (first patient in) and was completed on January 16, 2019 (last patient out). The individual time period between T1 and T2 ranged between 27 and 42 days, with an average interval of 34.7 days.

Figure 1. Study flowchart. A: app; P: physiotherapist; AP: app-guided followed by physiotherapist-guided sequence; PA: physiotherapist-guided followed by app-guided sequence; ESE: exercise-specific self-efficacy; CCPT: control competence for physical training; ANOVA: analysis of variance.



Baseline Data

Baseline characteristics are presented in Table 2. Subjects in the PA group had a worse baseline condition for hip-related quality of life compared with that of the AP group. Participants in the PA group also had lower values for overall exercise-specific self-efficacy, as well as the cope and schedule subdimensions, in comparison with those of the AP group at baseline. Other baseline characteristics did not differ between participants allocated to the two treatment sequences.

Table 2. Baseline data for the complete sample differentiated according to treatment sequence.

Characteristic	Total (N=54)	PA[a] (n=26)	AP[b] (n=28)	*P* value
Age (years), mean (SD)	62.4 (8.2)	62.5 (8.0)	62.3 (8.5)	.91
Gender, n (%)				.74
Female	32 (59)	16 (61)	16 (57)	
Male	22 (41)	10 (39)	12 (43)	
Education, n (%)				.19
Academic education	22 (41)	8 (31)	14 (50)	
Vocational education	31 (57)	18 (69)	13 (46)	
No vocational education	1 (2)	0 (0)	1 (4)	
Employment, n (%)				.44
Employed	32 (59)	14 (54)	18 (64)	
Retired	22 (41)	12 (46)	10 (36)	
Technical affinity[c], mean (SD)	2.87 (0.4)	2.90 (0.4)	2.84 (0.5)	.65
Previous experience with similar exercises in group sessions[d], median (IQR)	3.00 (2-3.25)	3.00 (2.0-3.0)	3.00 (2.0-4.0)	.31
Daily activity (minutes of cycling and walking/week), median (IQR)	215 (38-398)	215 (19-349)	225 (60-518)	.49
Sports activity (minutes/week), median (IQR)	209 (59-331)	229 (71-381)	184 (0-308)	.26
Fear of movement[e], median (IQR)	9.0 (8.0-12.0)	9.0 (8.0-13.3)	9.5 (8.3-10.8)	.84
WOMAC[f] pain, mean (SD)	31.4 (16.0)	31.4 (16.2)	31.4 (16.1)	.99
HOOS[g]				
Pain, mean (SD)	62.7 (15.5)	62.7 (16.2)	62.7 (15.1)	.99
Symptoms, mean (SD)	57.8 (17.6)	58.4 (16.4)	57.3 (19.4)	.83
ADL[h], median (IQR)	66.9 (54.4-82.7)	64.7 (54.4-77.9)	72.8 (55.2-86.8)	.26
Sport recreation, mean (SD)	54.2 (22.7)	49.0 (21.7)	58.9 (23.0)	.11
QL[i], median (IQR)	43.8 (29.7-62.5)	31.3 (25.0-50.0)	50.0 (37.5-62.5)	.03
ESE[j]				
Overall, mean (SD)	6.13 (1.4)	5.70 (1.1)	6.52 (1.5)	.03
Task, mean (SD)	6.04 (1.6)	5.85 (1.3)	6.21 (1.8)	.40
Cope, mean (SD)	5.53 (2.0)	4.88 (1.6)	6.13 (2.1)	.02
Schedule, median (IQR)	6.83 (5.7-8.1)	6.0 (5.7-7.3)	7.33 (6.3-8.9)	.02
CCPT[k], mean (SD)	2.67 (0.6)	2.59 (0.5)	2.74 (0.6)	.35

[a]PA: physiotherapist-guided followed by app-guided sequence.

[b]AP: app-guided followed by physiotherapist-guided sequence.

[c]Scored on a 5-point scale from 1 (not true at all) to 5 (fully true); n=3 missing values.

[d]Scored on a 5-point scale from 1 (substantial experience) to 5 (minimal experience).

[e]Scored from 6 (no fear) to 24 (extreme fear).

[f]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; pain subscale transformed values from 0 (no pain) to 100 (extreme pain).

[g]HOOS: Hip Disability and Osteoarthritis Outcome Score; transformed values from 0 (extreme impairment) to 100 (no impairment).

[h]ADL: activities of daily living.

[i]QL: hip-related quality of life.

[j]ESE: exercise-specific self-efficacy; rated on a scale from 0 (not at all safe) to 10 (absolutely safe).

[k]CCPT: control competence for physical training; scored on a scale from 1 (totally disagree) to 4 (totally agree).

Treatment Effects

Movement Quality

Results for movement quality are summarized in Table 3 and in Figures 2 and 3. The app was inferior to the physiotherapist in the primary outcome of overall movement quality as well as in all individual exercises, except for the mobility exercise. In contrast to the large effect sizes for the strengthening exercises (supine and table), the effect size for movement quality of the balance exercise was <0.2. However, the 95% CIs exceeded the noninferiority margin.

Table 3. Effects of treatment on movement quality (MQ).

Outcome[a]	Estimated mean (95% CI)		Analysis of variance mixed model			Effect size, g_{IG}[b] (95% CI)	ni[c]
	Physiotherapist	App	π^d (*P* value)	λ^e (*P* value)	τ_d^f (95% CI)		
Primary: MQ_overall	88.3 (87.2-89.5)	85.9 (84.8-87.0)	.002	.37	2.41 (1.21-3.61)	0.59 (0.29-0.89)	0
Secondary: MQ_mobility_seated	83.5 (80.7-86.3)	84.8 (82.0-87.6)	.01	.86	–1.27 (–4.27-1.72)	–0.13 (–0.41-0.16)	1
Secondary: MQ_strength_supine	91.6 (90.2-93.2)	87.9 (86.5-89.2)	.12	.80	3.76 (2.01-5.50)	0.75 (0.39-1.13)	0
Secondary: MQ_strength_table	87.7 (86.2-89.1)	81.4 (80.0-82.8)	.003	.04	6.25 (4.79-7.72)	1.19 (0.84-1.55)	0
Secondary: MQ_balance_stance	90.5 (89.1-91.9)	89.7 (88.3-91.1)	.58	.28	0.78 (–0.93-2.49)	0.15 (–0.17-0.48)	0

[a] Movement quality (MQ) with 0-100% of quality criteria points fulfilled.

[b]Hedges g_{IG}.

[c]ni: noninferiority for app ("1" if g_{IG} + 95% CI<0.2; else "0").

[d]π: period effect.

[e]λ: carryover effect.

[f]τ_d: treatment effect differences averaged over the levels of period and sequence; positive values indicate a beneficial effect for the physiotherapist.

Figure 2. Violin plots (mirrored estimated kernel density plot on each side of the boxplot, tails are trimmed to the range of the data) to visualize the distribution of movement quality (MQ) depending on the treatment sequence and type of intervention. AP: app-guided followed by physiotherapist-guided sequence; PA: physiotherapist-guided followed by app-guided sequence.

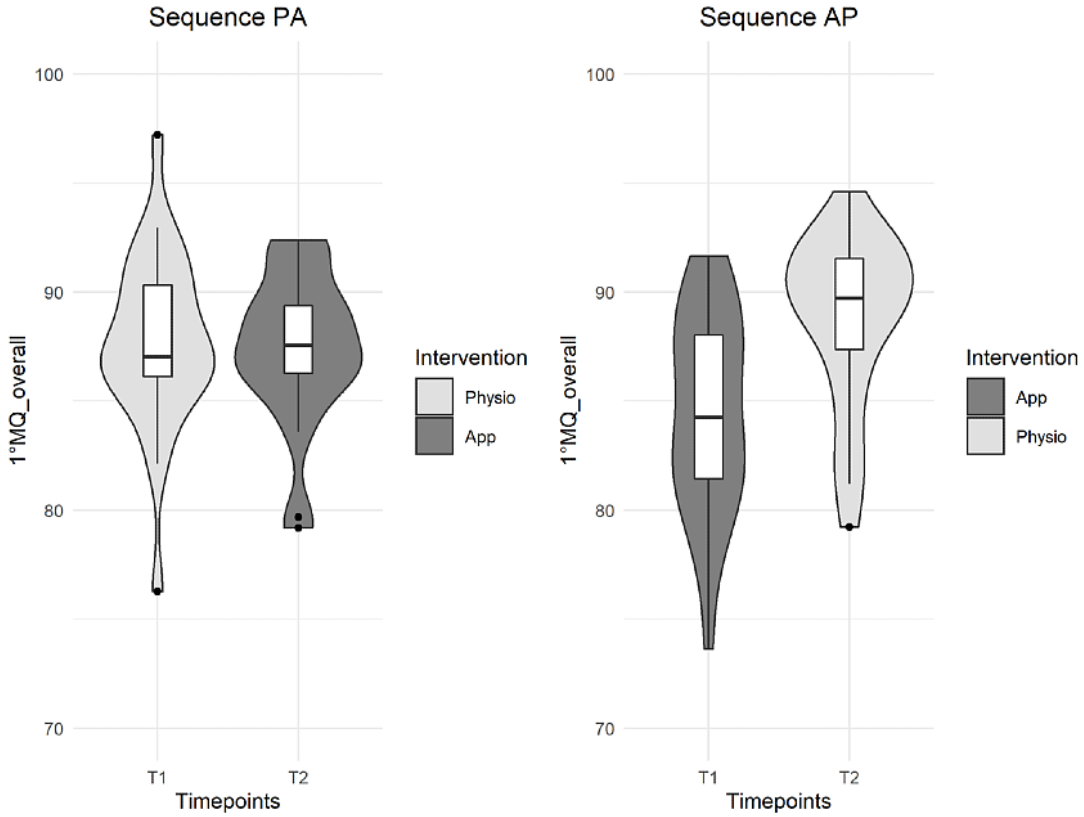
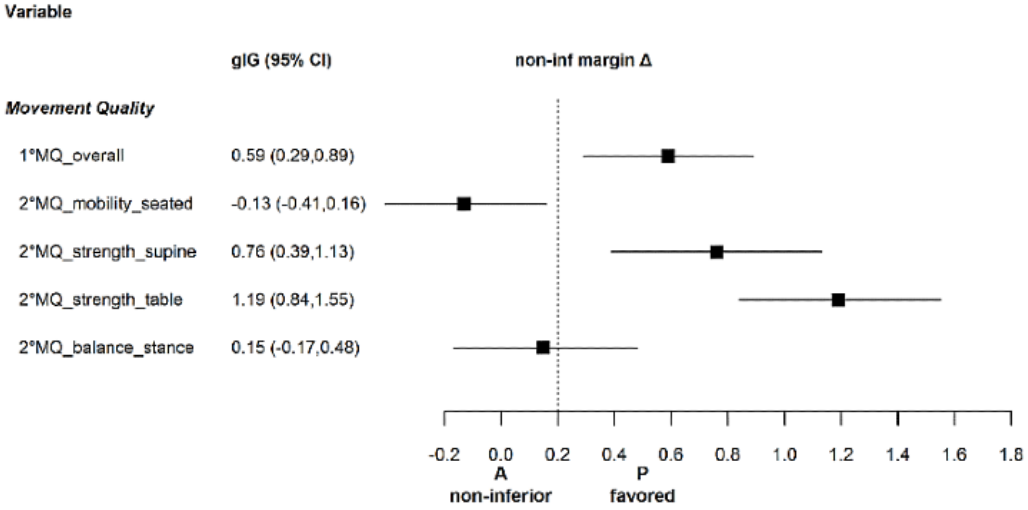


Figure 3. Effect sizes, gIG (95% CI), related to the noninferiority margin for movement quality (MQ).



The overall movement quality score, as well as the movement quality scores for the mobility exercise and strength exercise (table), were significantly better in period 2 (T2) than in period 1 (T1). A statistically significant carryover effect was detected in favor of treatment sequence PA for the movement quality of the table strength exercise.

Rater Agreement for Movement Quality

Total agreement between the two raters varied between 79% and 91%. Apart from movement quality for seated mobility, agreement was 3-5% better with the physiotherapist. The highest rater agreement of the assessors was found for the strength exercise in the supine position. Detailed results for movement quality rater agreement are shown in Multimedia Appendix 7.

The results of sensitivity analysis after excluding the movement quality rating categories with unsatisfactory interrater reliability are shown in Multimedia Appendix 8.

Exercise-Specific Self-Efficacy

The results for exercise-specific self-efficacy are shown in Table 4 and in Figures 4 and 5. The outcomes increased after the intervention in both groups. However, the app was inferior to the physiotherapist.

Overall exercise-specific self-efficacy showed a large effect of the physiotherapist versus the app. Medium effects were found in the exercise-specific self-efficacy subcategories coping and schedule, and a small effect was seen in the subcategory task (Table 4). All measures for exercise-specific self-efficacy showed statistically significant period and carryover effects: intervention effects (post-pre differences) were larger in period 1 compared to those in period 2, and carryover effects were measured in favor of treatment sequence PA. Three of the four scales showed excellent to good scale consistency, with Cronbach α values between .89 and .91. The task subscale also showed an acceptable consistency value with Cronbach α of .79.

Table 4. Effects of treatment on exercise-specific self-efficacy (ESE) and control competence for physical training (CCPT) (N=54).

Variable	Estimated mean (95% CI)		Analysis of variance mixed model			Effect size, g_{IG}[a] (95% CI)	ni[b]	Cronbach α
	Physiotherapist	App	π[c] (*P* value)	λ[d] (*P* value)	τ_d[e] (95% CI)			
Primary: ESE[f]_all_ Δ	1.85 (1.56-2.14)	0.95 (0.67-1.24)	<.001	.002	0.90 (0.52-1.27)	0.84 (0.46-1.22)	0	.89
Secondary: ESE_task_ Δ	1.96 (1.55-2.37)	1.23 (0.83-1.63)	<.001	.03	0.73 (0.13-1.33)	0.49 (0.10-0.88)	0	.79
Secondary: ESE_cope_ Δ	2.16 (1.75-2.57)	1.06 (0.67-1.46)	<.001	.03	1.10 (0.60-1.60)	0.74 (0.38-1.10)	0	.89
Secondary: ESE_schedule_ Δ	1.41 (1.07-1.75)	0.57 (0.23-0.90)	<.001	.001	0.84 (0.36-1.33)	0.68 (0.28-1.09)	0	.91
Primary: CCPT[g]_ Δ	0.31 (0.17-0.45)	0.22 (0.08-0.36)	.25	.36	0.09 (–0.08-0.26)	0.18 (–0.14-0.50)	0	.94

[a]Hedges g_{IG}.

[b]ni: noninferiority for app ("1" if g_{IG} + 95% CI<0.2; else "0").

[c]π: period effect.

[d]λ: carryover effect.

[e]τ_d: treatment effect differences averaged over the levels of period and sequence; positive values indicate a beneficial effect for physiotherapist.

[f]ESE: exercise-specific self-efficacy; rated on a scale of 0 (not at all safe) to 10 (absolutely safe).

[g]CCPT: control competence for physical training; rated on a scale of 1 (totally disagree) to 4 (totally agree).

Figure 4. Violin plots (mirrored estimated kernel density plot on each side of the boxplot, tails are trimmed to the range of the data) for visualization of distribution of exercise-specific self-efficacy (ESE) depending on treatment sequence and type of intervention. ESE values range from 0 (not at all safe) to 10 (absolutely safe). AP: app-guided followed by physiotherapist-guided sequence; PA: physiotherapist-guided followed by app-guided sequence.

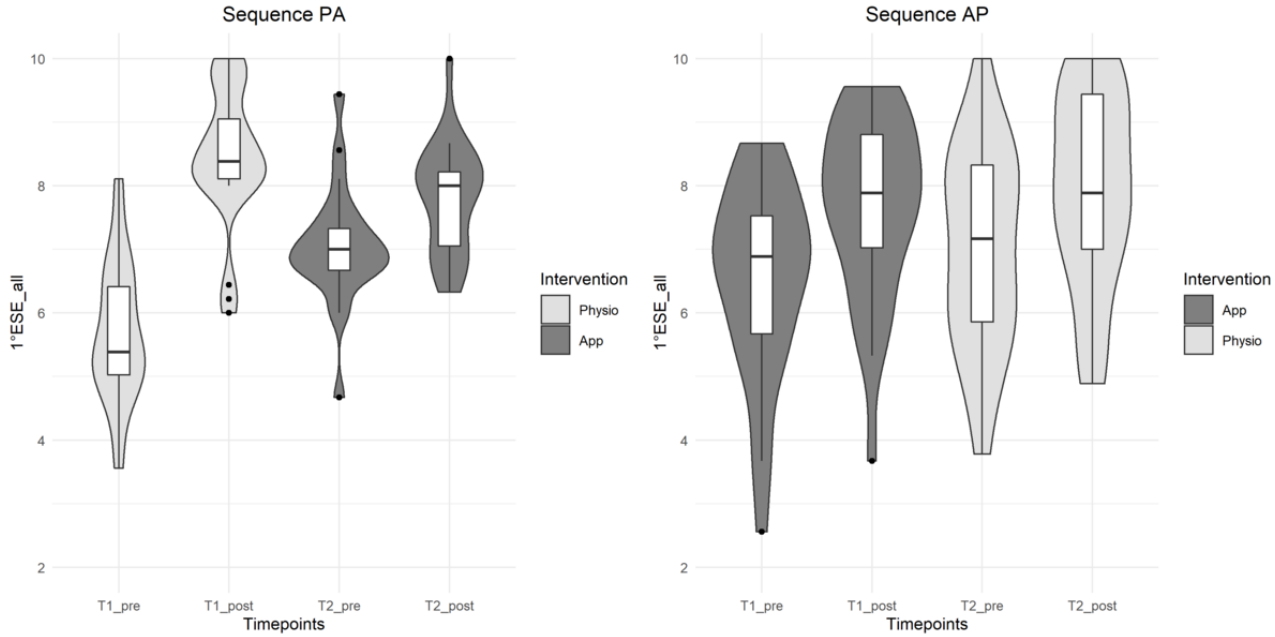
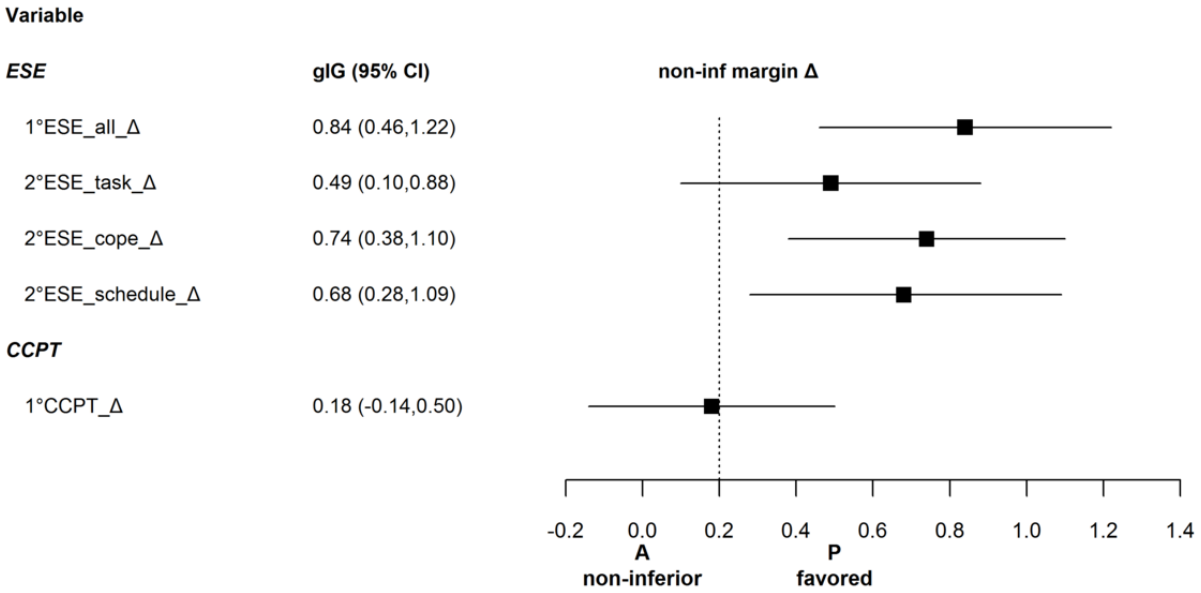


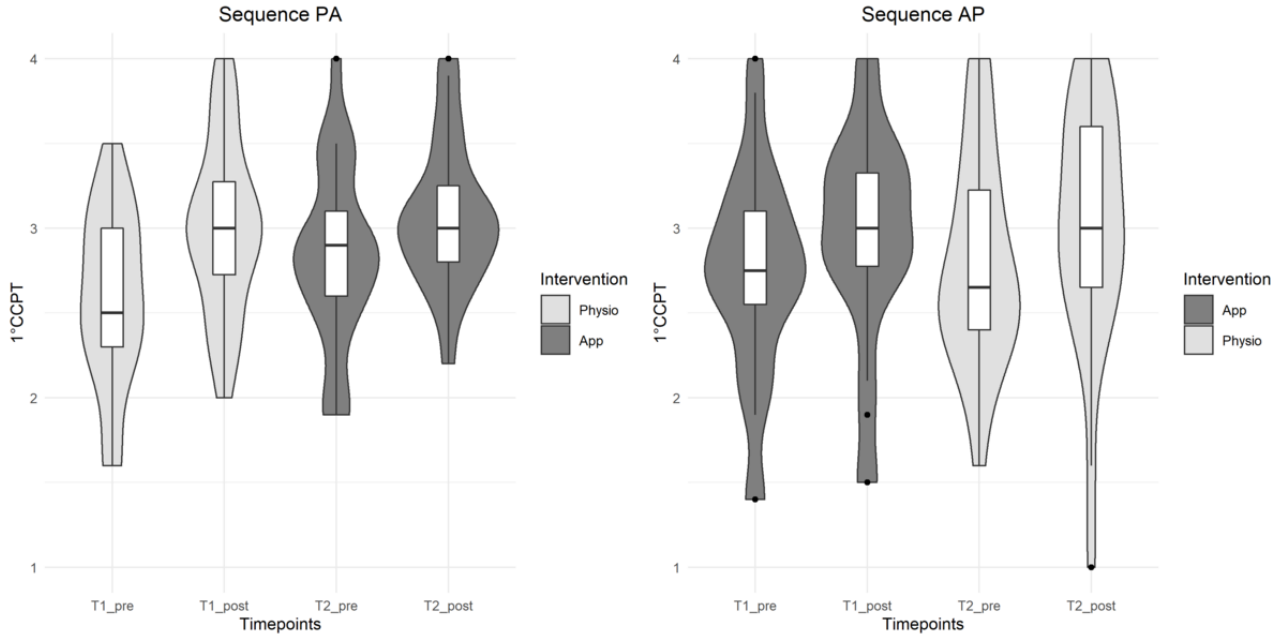
Figure 5. Effect sizes, gIG (95% CI), related to the noninferiority margin for exercise-specific self-efficacy (ESE, overall and with the subdimensions task, cope, and schedule) and control competence for physical training (CCPT). A: app; P: physiotherapist.



Control Competence for Physical Training

As shown in Table 4 and Figures 5 and 6, the outcomes for control competence for physical training increased after the intervention in both groups. However, the app was inferior to the physiotherapist. The effect size was <0.2; however, the 95% CI exceeded the noninferiority margin (Table 4). The scale consistency was excellent with a Cronbach α value of .94.

Figure 6. Violin plots (mirrored estimated kernel density plot on each side of the boxplot, tails are trimmed to the range of the data) for visualization of distribution of control competence for physical training (CCPT) depending on treatment sequence and type of intervention. CCPT values range from 1 (totally disagree) to 4 (totally agree). AP: app followed by physiotherapist sequence; PA: physiotherapist followed by app sequence.



Sensitivity Analysis for Outcomes with Carryover Effects

Outcomes with significant carryover effects were additionally analyzed for period 1 only, and are represented in Table 5. The app was inferior to physiotherapy, with large effect sizes in favor of physiotherapy for all analyzed variables.

Table 5. Carryover effects for period 1 only (Student t test for unpaired samples).

Variable	Mean (SD) change from baseline		*t* test, τ_d[a] (95% CI)	Effect size, *g*[b] (95% CI)	ni[c]
	Physiotherapist (n=26)	App (n=28)			
Primary: ESE[d]_all_Δ	2.77 (1.1)	1.15 (1.4)	1.62 (0.94-2.31)	1.27 (0.68-1.86)	0
Secondary: ESE_task_Δ	2.96 (1.4)	1.62 (2.0)	1.34 (0.39-2.29)	0.76 (0.20-1.32)	0
Secondary: ESE_cope_Δ	3.10 (1.8)	1.18 (1.6)	1.92 (1.01-2.83)	1.14 (0.56-1.72)	0
Secondary: ESE_schedule_Δ	2.26 (1.3)	0.64 (1.6)	1.62 (0.83-2.40)	1.10 (0.52-1.68)	0
Secondary: MQ[e]_strength_table	87. 8 (4.8)	79.0 (7.2)	8.7 (5.38, 12.10)	1.4 (0.79-2.00)	0

[a]τ_d: treatment effect differences averaged over the levels of period and sequence; positive values indicate a beneficial effect for physiotherapist.

[b]Hedges *g*.

[c]ni: noninferiority for app ("1" if g_{IG} + 95% CI<0.2; else "0").

[d]ESE: exercise-specific self-efficacy; values ranging from 0 (not at all safe) to 10 (absolutely safe).

[e]MQ: movement quality.

Harms

No harms or unintended effects occurred during the study.

Discussion

Principal Findings

Digital home training programs that help people support their training routines are urgently needed in the current world of decreasing physical activity. This is particularly true for patients suffering from OA. It is well known that exercises are efficient to decrease pain and increase physical functioning in OA [3,5-8]. To enable patients to train in a health-promoting and low-risk range, the PAHCO model provides a breakdown of subcompetences, each of which helps in the long-term realization of health-effective exercise [32,33]. A main interventional implication of the PAHCO model is to combine exercise practice, transfer of knowledge, and processing of personal experience of movement [40].

To the best of our knowledge, there is no app specifically designed for hip OA patients that combines the transfer of knowledge, exercise instructions, and processing of personal experience with movement. Furthermore, there are no studies comparing the effectiveness of these interventions provided by

humans and apps. Therefore, the aim of this study was to investigate whether digital exercise instruction and guidance leads to benefits in PAHCO's subcompetences in a manner that is comparable to physiotherapist care for patients with hip OA.

Movement quality of exercise execution was one important outcome in our study. Exercising in groups of two, in which the partner is alternately practicing and observing, shows important effects on learning success, especially for practical implementation with patients [44-47]. In our study, the physiotherapist served as an active corrective, but less as a role model, because she did not actively execute the exercises. In contrast, the app provides permanent illustrations of how to conduct the exercises via the virtual training partner and additional acoustic signals. The exercises selected for the study differed in terms of their starting position, used musculature, and complexity. Compared to the other two exercises, the two strength exercises were characterized by a more error-prone starting position and execution of movement, whereas the mobility exercise and balance tasks were simpler. Movement quality was inferior in the training session with the app compared with that guided by the physiotherapist, except for the mobility exercise. The difference between the physiotherapist and app for the balance exercise was smaller than the critical effect size value of 0.2, and the criterion for noninferiority was only missed because the 95% CIs exceeded the noninferiority margin. The effects of exercise for mobility and balance in the intervention with the app were close to those with the physiotherapist. However, both strength exercises showed large effects in favor of the physiotherapist-assisted intervention. It can therefore be concluded that not all exercises with different focuses and difficulties are suitable for being guided by a training app only. To avoid comprehension difficulties and poor movement quality, as well as potentially related impairments of training efficacy and risk of harm, we recommend that patients receive supervised instructions when initiating training, especially for more complex exercises. This is in line with results of other investigations, in which initial face-to-face contact with individuals before a new intervention showed positive changes in physical activity outcomes. This is also recommended in guidelines for the nonpharmacological treatment of OA [7,48].

Self-efficacy is an important key to modifying behavior [37]. People with high self-efficacy set higher goals, invest more effort into the pursuit of their goals, and will be more likely to keep trying when barriers and setbacks arise [49]. Self-efficacy increases when the individual receives positive feedback after successfully completing a task [50]. The results from this study revealed that the app is inferior to a physiotherapist in outcomes related to exercise-specific self-efficacy. In particular, the overall scale for exercise-specific self-efficacy as well as its subscales coping and scheduling showed large effects in favor of the physiotherapist. Coping is related to barriers for carrying out exercises, such as "I am able to exercise when I lack energy" or "…don't feel well," and scheduling is related to the ability to integrate exercises into daily routines. It can be concluded that face-to-face supervision of a single training session by a physiotherapist cannot be equally replaced by a digital app with a pedagogical agent and video-based training instructions. This was confirmed by Danbjørg et al [12], who reported that the participants missed personal contact with the physiotherapist. However, it should not be disregarded that self-efficacy increased even after training with the app. Litman et al [51] also reported an improvement in self-efficacy using digital apps, but also regarded this type of intervention as a complementary and supporting measure.

We found carryover effects for all outcomes related to the self-administered exercise-specific self-efficacy score, but only for one movement quality outcome. The subjective confidence of being able to show a desired behavior seems to be influenced by a personal supervisor much more than the objective ability to perform a movement. This finding underlines the importance of using both subjective and objective outcome measures if PAHCO is to be improved by a special intervention.

The effect size for control competence for physical training in our study was smaller than the critical value of 0.2, and the criterion for noninferiority was only missed because of the 95% CI exceeding the noninferiority margin. Control competence for physical training is quantified using items that are directly related to the competence to control training intensity in the required way (ie, "I know how I can best increase my strength in the leg and hip area with physical training" or "I am able to adjust my training effort well to my physical condition"). These items are comparable to the exercise-specific self-efficacy subscale task to some extent, with items such as "…I am certain that I will be able to carry out exercises with the right technique." The app was inferior to the physiotherapist in this subscale. However, only a small effect size was obtained in comparison to all other exercise-specific self-efficacy scales that showed medium or large effect sizes.

The characteristics of our sample are typical characteristics of OA patients, indicating the good generalizability of the findings. The average age and the gender distribution correspond to the risk profile of the disease [52]. The average Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain score attests a moderate hip OA [53]. However, the fear of movement among our participants was relatively low compared to that of other OA samples [11]. Overall baseline values further indicate that the study sample was physically active according to physical activity guidelines, and had a positive attitude toward technology [54]. These criteria may restrict the external validity of the results to a comparable population with respect to fear of movement, physical activity status, and technical affinity.

No adverse events were reported for any of the interventions. However, safety aspects are extremely relevant for fully automated computer-based interventions as there is no health care professional controlling for nonphysiological or even harmful execution of exercises. The app tested in this study used personalized closed feedback loops to adapt exercise instructions and dosage according to the user's feedback on pain and physical exhaustion. Nonetheless, it has been shown that movement quality was inferior by using the app when it comes to more complex exercises. Future studies with longer intervention periods should therefore evaluate if minor movement competence goes along with higher pain levels, more

adverse events, or poorer health outcomes. Proof of safety and medical benefit or patient-relevant structural and process improvements are mandatory aspects for approval of an "app on prescription" in Germany [55].

Limitations

Results of this study show ceiling effects for the variables self-efficacy and control competence. The study population already had very good values at baseline, thereby reducing the possibility for change. To be able to assess effects in a more differentiated way, samples of future studies should have lower initial values in this outcome dimension. To investigate outcome effects in a more vulnerable population, further research should also focus on a sample that is less physically active, has more severe symptoms, feels greater barriers to technology, and has a greater fear of movement. This limitation and the above-mentioned restrictions with regard to the external validity of the study results may be caused by a potential recruiting bias, which may have affected the outcomes as well. Subjects were fully informed about the rationale and aim of the study in the context of recruitment and inclusion. It therefore cannot be ruled out that this information may have had an effect on user self-selection and expectation, and may have therefore biased the results.

This study also has some methodological issues that should be discussed. Five outcomes of the study showed carryover effects in favor of the physiotherapist, four of which were related to exercise-specific self-efficacy. Although a washout phase with a minimum of 3 weeks was conducted between training sessions, this phase was not long enough to eliminate positive treatment effects. The session with the physiotherapist induced long-lasting effects that sustained during the washout phase. As a consequence, a sensitivity analysis was conducted and results of this sensitivity analysis led to similar but even more pronounced statements related to the inferiority of the app versus the physiotherapist. We are aware that the practice of analyzing data from the first study period as if it had been obtained from a conventional parallel-group design has been shown to be potentially strongly anticonservative [56]. Yet, no conclusive answer is provided by the literature on how to proceed if analyses yield significant carryover differences. Future studies should use a parallel-group design or significantly extend the washout phase to an adequate length to eliminate carryover [57]. However, the results of our study indicate that the latter seems to be difficult to implement in studies with exercise interventions. A complete washout was also not desired in our study.

Exercise-specific self-efficacy and control competence for physical training were assessed prior to and after each intervention, and change from baseline was used as a dependent variable for each intervention. There is little agreement in the current literature as to whether or how to introduce period-specific baseline measurements into the model. The use of change from baseline is discouraged due to its poor type I error rate control and lower power than other methods, yet under the assumption of no carryover [58]. Nevertheless, change from baseline was applied in our study, as it is the most simplistic way to analyze the data, and other methods face similar problems in the presence of carryover effects [59].

Aside from carryover effects, period effects were present in almost all measures. Effects of differences between pre and post values of the training session were larger in period 1, whereas movement quality assessed while exercising was better in period 2. This seems conclusive, as the possibility of improvement within one training session may have a saturation effect partially due to ceiling effects, whereas the better movement quality in period 2 may be related to a learning effect on exercise execution from period 1. As period effects do not affect comparison between groups, this limitation does not seem crucial for the interpretation of the results.

Conclusion

Despite the absence of noninferiority to the physiotherapist in almost all measures of interest, exercise-specific self-efficacy and control competence for physical training also improve using an app, and movement quality is acceptable for exercises that are easy to perform. However, relevant differences in movement quality are present in challenging tasks. The digital app therefore opens up possibilities to take on the role of a supplementary tool to support the patient in independent home training for less complex exercises. Nevertheless, it cannot replace a physiotherapist with an equivalent effect.

Acknowledgments

The authors wish to thank the study participants. They would also like to express their thanks to the scientific assistants Simone Schweda, Hendrik Maier, Lisa-Marie Krehl, Hannah Hofmann, and Phillip Weber, and colleagues Georg Haupt and Dr. Katharina Bernecker for their help in data collection and analysis. The project is funded by the Leibniz-Wissenschafts Campus Tuebingen "Cognitive Interface" with funds from the Ministry of Science, Research and the Arts Baden-Wuerttemberg. Tuebingen University Hospital is the owner of the software; however, the software is not open to the public.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot of the exercise video "Pelvic tilt.".
[PNG File , 627 KB - jmir_v22i9e18233_app1.png]

Multimedia Appendix 2
Screenshot of the exercise video “Strengthening of the hip abductors.".
[PNG File , 718 KB - jmir_v22i9e18233_app2.png]

Multimedia Appendix 3
Screenshot of the exercise video “Strengthening of hip extensors.”.
[PNG File , 684 KB - jmir_v22i9e18233_app3.png]

Multimedia Appendix 4
Screenshot of the exercise video "Balance task.".
[PNG File , 588 KB - jmir_v22i9e18233_app4.png]

Multimedia Appendix 5
App components and theoretical background of the interaction principles.
[DOCX File , 24 KB - jmir_v22i9e18233_app5.docx]

Multimedia Appendix 6
Exemplary decision path of the software algorithm based on the exercise to strengthen the hip abductors.
[PNG File , 164 KB - jmir_v22i9e18233_app6.png]

Multimedia Appendix 7
Rater agreement for movement quality, averaged across all categories and sets for each exercise and intervention in percent. P: physiotherapist; A: app.
[DOCX File , 14 KB - jmir_v22i9e18233_app7.docx]

Multimedia Appendix 8
Sensitivity analysis for movement quality after excluding categories with poor interrater reliability.
[DOCX File , 23 KB - jmir_v22i9e18233_app8.docx]

References

1. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
2. Fuchs J, Kuhnert R, Scheidt-Nave C. 12-Monats-Prävalenz von Arthrose in Deutschland. J Health Monit 2017;2(3):55-60 [FREE Full text] [doi: 10.17886/RKI-GBE-2017-054]
3. Kraus VB, Sprow K, Powell KE, Buchner D, Bloodgood B, Piercy K, 2018 Physical Activity Guidelines Advisory Committee. Effects of Physical Activity in Knee and Hip Osteoarthritis: A Systematic Umbrella Review. Med Sci Sports Exerc 2019 Jun;51(6):1324-1339 [FREE Full text] [doi: 10.1249/MSS.0000000000001944] [Medline: 31095089]
4. Nüesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Jüni P. All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. BMJ 2011 Mar 08;342:d1165 [FREE Full text] [doi: 10.1136/bmj.d1165] [Medline: 21385807]
5. Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie e.V. (DGOOC), Stöve J. S2k-Leitlinie Gonarthrose (Registernummer 033 - 004). Leitlinien. Berlin: AWMF Online; 2018. URL: https://www.awmf.org/leitlinien/detail/ll/033-004.html [accessed 2020-09-11]
6. Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC), Matziolis G. S2k-Leitlinie Koxarthrose (Registernummer 033-001). Leitlinien. Berlin: AWMF Online; 2019. URL: https://www.awmf.org/leitlinien/detail/ll/033-001.html [accessed 2020-09-11]
7. Fernandes L, Hagen KB, Bijlsma JWJ, Andreassen O, Christensen P, Conaghan PG, European League Against Rheumatism (EULAR). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013 Jul;72(7):1125-1135. [doi: 10.1136/annrheumdis-2012-202745] [Medline: 23595142]
8. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014 Mar;22(3):363-388 [FREE Full text] [doi: 10.1016/j.joca.2014.01.003] [Medline: 24462672]
9. Wallis JA, Webster KE, Levinger P, Taylor NF. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. Osteoarthritis Cartilage 2013 Nov;21(11):1648-1659 [FREE Full text] [doi: 10.1016/j.joca.2013.08.003] [Medline: 23948979]

10. Dobson F, Bennell KL, French SD, Nicolson PJA, Klaasman RN, Holden MA, et al. Barriers and Facilitators to Exercise Participation in People with Hip and/or Knee Osteoarthritis: Synthesis of the Literature Using Behavior Change Theory. Am J Phys Med Rehabil 2016 May;95(5):372-389. [doi: 10.1097/PHM.0000000000000448] [Medline: 26945211]
11. Gunn AH, Schwartz TA, Arbeeva LS, Callahan LF, Golightly Y, Goode A, et al. Fear of Movement and Associated Factors Among Adults With Symptomatic Knee Osteoarthritis. Arthritis Care Res (Hoboken) 2017 Dec;69(12):1826-1833. [doi: 10.1002/acr.23226] [Medline: 28371481]
12. Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an Exercise App in the Care for People With Osteoarthritis: User-Driven Exploratory Study. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e11 [FREE Full text] [doi: 10.2196/mhealth.7734] [Medline: 29326092]
13. Bennell KL, Dobson F, Hinman RS. Exercise in osteoarthritis: moving from prescription to adherence. Best Pract Res Clin Rheumatol 2014 Feb;28(1):93-117. [doi: 10.1016/j.berh.2014.01.009] [Medline: 24792947]
14. McCarthy CJ, Mills PM, Pullen R, Roberts C, Silman A, Oldham JA. Supplementing a home exercise programme with a class-based exercise programme is more effective than home exercise alone in the treatment of knee osteoarthritis. Rheumatology (Oxford) 2004 Jul;43(7):880-886. [doi: 10.1093/rheumatology/keh188] [Medline: 15113993]
15. Postler A, Ramos AL, Goronzy J, Günther KP, Lange T, Schmitt J, et al. Prevalence and treatment of hip and knee osteoarthritis in people aged 60 years or older in Germany: an analysis based on health insurance claims data. Clin Interv Aging 2018;13:2339-2349. [doi: 10.2147/CIA.S174741] [Medline: 30532524]
16. Ali SA, Walsh KE, Kloseck M. Patient perspectives on improving osteoarthritis management in urban and rural communities. J Pain Res 2018;11:417-425. [doi: 10.2147/JPR.S150578] [Medline: 29503578]
17. Allen KD, Choong PF, Davis AM, Dowsey MM, Dziedzic KS, Emery C, et al. Osteoarthritis: Models for appropriate care across the disease continuum. Best Pract Res Clin Rheumatol 2016 Jun;30(3):503-535. [doi: 10.1016/j.berh.2016.09.003] [Medline: 27886944]
18. Kloek CJJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health 2018 Aug 31;18(1):1082 [FREE Full text] [doi: 10.1186/s12889-018-5975-7] [Medline: 30170586]
19. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an Internet-Delivered Exercise and Pain-Coping Skills Training Intervention for Persons With Chronic Knee Pain: A Randomized Trial. Ann Intern Med 2017 Apr 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
20. Dahlberg LE, Grahn D, Dahlberg JE, Thorstensson CA. A Web-Based Platform for Patients With Osteoarthritis of the Hip and Knee: A Pilot Study. JMIR Res Protoc 2016 Jun 03;5(2):e115 [FREE Full text] [doi: 10.2196/resprot.5665] [Medline: 27261271]
21. Nero H, Dahlberg J, Dahlberg LE. A 6-Week Web-Based Osteoarthritis Treatment Program: Observational Quasi-Experimental Study. J Med Internet Res 2017 Dec 18;19(12):e422 [FREE Full text] [doi: 10.2196/jmir.9255] [Medline: 29254906]
22. Bossen D, Veenhof C, Dekker J, de Bakker D. The usability and preliminary effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis. BMC Med Inform Decis Mak 2013 May 28;13:61 [FREE Full text] [doi: 10.1186/1472-6947-13-61] [Medline: 23714120]
23. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
24. Hsu H, Kuo T, Lin J, Hsu W, Yu C, Chen Y, et al. A Cross-Disciplinary Successful Aging Intervention and Evaluation: Comparison of Person-to-Person and Digital-Assisted Approaches. Int J Environ Res Public Health 2018 May 04;15(5):913 [FREE Full text] [doi: 10.3390/ijerph15050913] [Medline: 29734674]
25. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
26. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
27. Haupt G, Janßen P, Krauß I, Steinhilber B. Das Tübinger Hüftkonzept. Essen: Verlag hellblau; 2014.
28. Krauß I, Steinhilber B, Haupt G, Miller R, Martus P, Janßen P. Exercise therapy in hip osteoarthritis--a randomized controlled trial. Dtsch Arztebl Int 2014 Sep 01;111(35-36):592-599. [doi: 10.3238/arztebl.2014.0592] [Medline: 25249361]
29. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which Combinations of Techniques and Modes of Delivery in Internet-Based Interventions Effectively Change Health Behavior? A Meta-Analysis. J Med Internet Res 2016 Jun 07;18(6):e155 [FREE Full text] [doi: 10.2196/jmir.4218] [Medline: 27268104]
30. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

31. Gourlan M, Bernard P, Bortolon C, Romain AJ, Lareyre O, Carayol M, et al. Efficacy of theory-based interventions to promote physical activity. A meta-analysis of randomised controlled trials. Health Psychol Rev 2016;10(1):50-66. [doi: 10.1080/17437199.2014.981777] [Medline: 25402606]
32. Sudeck G, Pfeifer K. Physical activity-related health competence as an integrative objective in exercise therapy and health sports – conception and validation of a short questionnaire. Sportwiss 2016 May 17;46(2):74-87. [doi: 10.1007/s12662-016-0405-4]
33. Pfeifer K, Sudeck G, Brüggemann S, Huber G. [DGRW-update: exercise therapy in medical rehabilitation--effects, quality, perspectives]. Rehabilitation (Stuttg) 2010 Aug;49(4):224-236. [doi: 10.1055/s-0030-1261909] [Medline: 20677118]
34. McAuley E, Blissmer B. Self-efficacy determinants and consequences of physical activity. Exerc Sport Sci Rev 2000 Apr;28(2):85-88. [Medline: 10902091]
35. Biddle S, Hagger M, Chatzisarantis N, Lippke S. Theoretical frameworks in exercise psychology. In: Tenenbaum G, Eklund RC, editors. Handbook of Sport Psychology. Hoboken (US): John Wiley & Sons Inc; 2007:537-559.
36. Schäfer AGM, Zalpour C, von Piekartz H, Hall TM, Paelke V. The Efficacy of Electronic Health-Supported Home Exercise Interventions for Patients With Osteoarthritis of the Knee: Systematic Review. J Med Internet Res 2018 Apr 26;20(4):e152 [FREE Full text] [doi: 10.2196/jmir.9465] [Medline: 29699963]
37. Uesugi Y, Koyanagi J, Takagi K, Yamaguchi R, Hayashi S, Nishii T. Exercise Therapy Interventions in Patients With Hip Osteoarthritis: Comparison of the Effects of DVD and Website-Based Interventions. JMIR Rehabil Assist Technol 2018 May 07;5(1):e10 [FREE Full text] [doi: 10.2196/rehab.8251] [Medline: 29735473]
38. Krauss I, Steinhilber B, Haupt G, Miller R, Grau S, Janssen P. Efficacy of conservative treatment regimes for hip osteoarthritis--evaluation of the therapeutic exercise regime "Hip School": a protocol for a randomised, controlled trial. BMC Musculoskelet Disord 2011 Nov 24;12:270 [FREE Full text] [doi: 10.1186/1471-2474-12-270] [Medline: 22114973]
39. Steinhilber B, Haupt G, Miller R, Janssen P, Krauss I. Exercise therapy in patients with hip osteoarthritis: Effect on hip muscle strength and safety aspects of exercise-results of a randomized controlled trial. Mod Rheumatol 2017 May;27(3):493-502. [doi: 10.1080/14397595.2016.1213940] [Medline: 27486681]
40. Pfeifer K, Sudeck G, Geidl W, Tallner A. Bewegungsförderung und Sport in der Neurologie? Kompetenzorientierung und Nachhaltigkeit. Neurol Rehab 2013;19:7-19.
41. Rodgers WM, Wilson PM, Hall CR, Fraser SN, Murray TC. Evidence for a multidimensional self-efficacy for exercise scale. Res Q Exerc Sport 2008 Jun;79(2):222-234. [doi: 10.1080/02701367.2008.10599485] [Medline: 18664046]
42. Random Integer Generator. Random.org. Dublin: Randomness and Integrity Services Ltd URL: https://www.random.org/ [accessed 2020-09-11]
43. Kitchenham B, Madeyski L, Curtin F. Corrections to effect size variances for continuous outcomes of crossover clinical trials. Stat Med 2018 Jan 30;37(2):320-323. [doi: 10.1002/sim.7379] [Medline: 29250841]
44. Wulf G. Motorisches Lernen - Therapierelevante Forschungsergebnisse. ergoscience 2007 May;2(2):47-55. [doi: 10.1055/s-2007-963010]
45. Shea CH, Wright DL, Wulf G, Whitacre C. Physical and observational practice afford unique learning opportunities. J Mot Behav 2000 Mar;32(1):27-36. [doi: 10.1080/00222890009601357] [Medline: 11008269]
46. Shebilskem WL, Regian JW, Arthur W, Jordan JA. A Dyadic Protocol for Training Complex Skills. Hum Factors 2016 Nov 23;34(3):369-374. [doi: 10.1177/001872089203400309]
47. McCullagh P, Meyer KN. Learning versus correct models: influence of model type on the learning of a free-weight squat lift. Res Q Exerc Sport 1997 Mar;68(1):56-61. [doi: 10.1080/02701367.1997.10608866] [Medline: 9094763]
48. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
49. Bandura A. Self-efficacy: The exercise of control. New York: WH Freeman; 1997.
50. Saemi E, Porter J, Ghotbi-Varzaneh A, Zarghami M, Maleki F. Knowledge of results after relatively good trials enhances self-efficacy and motor learning. Psychol Sport Exerc 2012 Jul;13(4):378-382. [doi: 10.1016/j.psychsport.2011.12.008]
51. Litman L, Rosen Z, Spierer D, Weinberger-Litman S, Goldschein A, Robinson J. Mobile Exercise Apps and Increased Leisure Time Exercise Activity: A Moderated Mediation Analysis of the Role of Self-Efficacy and Barriers. J Med Internet Res 2015 Aug 14;17(8):e195 [FREE Full text] [doi: 10.2196/jmir.4142] [Medline: 26276227]
52. Arden N, Nevitt M. Osteoarthritis: epidemiology. Best Pract Res Clin Rheumatol 2006 Feb;20(1):3-25. [doi: 10.1016/j.berh.2005.09.007] [Medline: 16483904]
53. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Differences between men and women in the rate of use of hip and knee arthroplasty. N Engl J Med 2000 Apr 06;342(14):1016-1022. [doi: 10.1056/NEJM200004063421405] [Medline: 10749964]
54. Pfeifer K, Banzer W, Ferrari N, Füzéki E, Geidl W, Graf C, et al. Nationale Empfehlungen für Bewegung und Bewegungsförderung. Erlangen-Nürnberg: Friedrich-Alexander-Universität; 2016. URL: https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/b/bewegungsempfehlungen.html [accessed 2020-09-11]
55. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V. Digitale Gesundheitsanwendungen. Bonn: Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM); 2020 Jul 31. URL: https://www.bfarm.de/DE/Medizinprodukte/DVG/_node.html [accessed 2020-09-11]

56. Freeman PR. The performance of the two-stage analysis of two-treatment, two-period crossover trials. Stat Med 1989 Dec;8(12):1421-1432. [doi: 10.1002/sim.4780081202] [Medline: 2616932]
57. Senn S. In: Senn S, Barnett V, editors. Cross-over Trials in Clinical Research. Chichester, UK: John Wiley & Sons, Ltd; 2002.
58. Mehrotra DV. A recommended analysis for 2 × 2 crossover trials with baseline measurements. Pharm Stat 2014;13(6):376-387. [doi: 10.1002/pst.1638] [Medline: 25230245]
59. Chen X, Meng Z, Zhang J. Handling of baseline measurements in the analysis of crossover trials. Stat Med 2012 Jul 30;31(17):1791-1803. [doi: 10.1002/sim.4466] [Medline: 22715129]

Abbreviations

ANOVA: analysis of variance
AP: app-guided followed by physiotherapist-guided sequence
OA: osteoarthritis
PA: physiotherapist-guided followed by app-guided sequence
PAHCO: physical activity-related health competence

Edited by G Eysenbach; submitted 13.02.20; peer-reviewed by G Signorelli, L Brosseau; comments to author 12.06.20; revised version received 23.06.20; accepted 25.06.20; published 28.09.20.

Please cite as:
Durst J, Roesel I, Sudeck G, Sassenberg K, Krauss I
Effectiveness of Human Versus Computer-Based Instructions for Exercise on Physical Activity–Related Health Competence in Patients with Hip Osteoarthritis: Randomized Noninferiority Crossover Trial
J Med Internet Res 2020;22(9):e18233
URL: http://www.jmir.org/2020/9/e18233/
doi:10.2196/18233
PMID:32985991

Original Paper

Behavior Change Text Messages for Home Exercise Adherence in Knee Osteoarthritis: Randomized Trial

Kim Bennell[1], PhD; Rachel K Nelligan[1], BPTh; Sarah Schwartz[1], MSci; Jessica Kasza[2], PhD; Alexander Kimp[1], DPT; Samuel JC Crofts[3], MPH; Rana S Hinman[1], PhD

[1]Department of Physiotherapy, University of Melbourne, Centre for Health, Exercise and Sports Medicine, Carlton, Australia

[2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia

[3]Melbourne School of Population and Global Health, University of Melbourne, Centre for Epidemiology and Biostatistics, Parkville, Australia

Corresponding Author:
Kim Bennell, PhD
Department of Physiotherapy
University of Melbourne
Centre for Health, Exercise and Sports Medicine
Level 7, Alan Gilbert Building
161 Barry St
Carlton, 3053
Australia
Phone: 61 383444135
Email: k.bennell@unimelb.edu.au

Abstract

Background: Exercise is a core recommended treatment for knee osteoarthritis (OA), yet adherence declines, particularly following cessation of clinician supervision.

Objective: This study aims to evaluate whether a 24-week SMS intervention improves adherence to unsupervised home exercise in people with knee OA and obesity compared with no SMS.

Methods: A two-group superiority randomized controlled trial was performed in a community setting. Participants were people aged 50 years with knee OA and BMI ≥30 kg/m^2 who had undertaken a 12-week physiotherapist-supervised exercise program as part of a preceding clinical trial. Both groups were asked to continue their home exercise program unsupervised three times per week for 24 weeks and were randomly allocated to a behavior change theory–informed, automated, semi-interactive SMS intervention addressing exercise barriers and facilitators or to control (no SMS). Primary outcomes were self-reported home exercise adherence at 24 weeks measured by the Exercise Adherence Rating Scale (EARS) Section B (0-24, higher number indicating greater adherence) and the number of days exercised in the past week (0-3). Secondary outcomes included self-rated adherence (numeric rating scale), knee pain, physical function, quality of life, global change, physical activity, self-efficacy, pain catastrophizing, and kinesiophobia.

Results: A total of 110 participants (56 SMS group and 54 no SMS) were enrolled and 99 (90.0%) completed both primary outcomes (48/56, 86% SMS group and 51/54, 94% no SMS). At 24 weeks, the SMS group reported higher EARS scores (mean 16.5, SD 6.5 vs mean 13.3, SD 7.0; mean difference 3.1, 95% CI 0.8-5.5; *P*=.01) and more days exercised in the past week (mean 1.8, SD 1.2 vs mean 1.3, SD 1.2; mean difference 0.6, 95% CI 0.2-1.0; *P*=.01) than the control group. There was no evidence of between-group differences in secondary outcomes.

Conclusions: An SMS program increased self-reported adherence to unsupervised home exercise in people with knee OA and obesity, although this did not translate into improved clinical outcomes.

Trial Registration: Australian New Zealand Clinical Trials Registry 12617001243303; https://tinyurl.com/y2ud7on5

International Registered Report Identifier (IRRID): RR2-10.1186/s12891-019-2801-z

(J Med Internet Res 2020;22(9):e21749) doi:10.2196/21749

KEYWORDS

knee osteoarthritis; exercise; patient compliance; mobile phone; randomized controlled trial

Introduction

Background

Knee osteoarthritis (OA) is a global public health problem [1]. As OA has no cure, supporting patients to self-manage their condition is vital. Exercise is a core recommended treatment for knee OA [2,3] and is important for common comorbidities such as obesity, diabetes, and heart disease [4]. Exercise programs often involve initial supervision by a clinician, followed by unsupervised home exercise. Ideally, regular participation in exercise should be one of the long-term goals of self-management. Unfortunately, adherence to home exercise is often poor [5], particularly once clinician input ceases [6]. Numerous barriers can impact adherence, such as pain, negative beliefs about OA and exercise, and poor self-efficacy [7,8]. This decline in exercise adherence is typically mirrored by a gradual loss of initial clinical benefits [6,9]. Thus, scalable strategies to improve adherence to structured home exercise are thought to be important for better long-term patient outcomes [10].

There is uncertainty about how best to help people with knee OA adhere to exercise. Interventions that show promise include *booster* or *refresher* sessions with a physiotherapist and behavioral graded exercise, involving gradual increases in physical activity plus *booster* sessions [11]. However, ongoing clinician involvement may be unfeasible or impractical for many patients due to access challenges and/or cost. Instead, the use of digital communications such as SMS, email, or apps may be inexpensive and accessible options to help promote exercise adherence. As patients with knee OA tend to be older, SMS may have advantages over other forms of digital communication due to its widespread use, familiarity, and potential to overcome barriers related to device ownership (eg, not owning a smartphone) and access to and availability of Wi-Fi cellular data. The effectiveness of SMS-based interventions to promote healthy behaviors relevant to OA, such as physical activity, diet, and/or weight loss, has also been demonstrated in various settings and other conditions [12-14]. To date, the use of SMS to improve adherence to home exercise or physical activity in people with knee OA has only been evaluated in three pilot or feasibility studies [15-17].

Objectives

The primary aim of the ADHERE randomized controlled trial (RCT) was to evaluate the effects of a theoretically informed 24-week SMS program [18] on self-reported adherence to a prescribed, unsupervised, structured home exercise program, undertaken after an initial 12-week period of physiotherapist supervision. We hypothesized that the SMS intervention would lead to greater exercise adherence than no SMS contact.

Methods

Trial Design

This parallel, two-arm superiority RCT is reported according to CONSORT (Consolidated Standards of Reporting Trials) [19], CONORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth) [20], Template for Intervention Description and Replication (TIDieR) [21], and Consensus on Exercise Reporting Template (CERT) recommendations [22]. It was prospectively registered (Australian New Zealand Clinical Trials Registry #12617001243303), and the trial protocol is published [23]. Approval was obtained from the Institutional Human Research Ethics Committee (#1544919).

Participants

This trial used participants completing another study, the TARGET trial [24], where participants visited a physiotherapist five times over 12 weeks for prescription of either a weight-bearing functional exercise program or a non–weight-bearing quadriceps strengthening exercise program. TARGET trial participants were recruited from the community in Melbourne, Australia, between September 2017 and May 2019 via advertisements through consumer organizations, social media, community locations, media, and our volunteer database. Inclusion criteria were as follows: (1) aged ≥50 years, (2) knee pain on most days of the past month, (3) knee pain for ≥3 months, (4) average overall pain severity ≥4 on an 11-point numeric rating scale (NRS), (5) tibiofemoral osteophytes on x-ray, (6) obesity (BMI ≥30 kg/m^2), and (7) own a mobile phone with text messaging. The exclusion criteria are found in Multimedia Appendix 1.

The TARGET trial included face-to-face visits with members of the research team at the University of Melbourne. Only those who completed the TARGET trial final 12-week assessment and did not withdraw at this time point were enrolled into the ADHERE trial. Participants provided written informed consent to participate in the subsequent ADHERE trial at the time of TARGET trial enrollment.

Randomization, Allocation Concealment, and Blinding

On completion of the TARGET trial final assessment (which served a dual purpose as the ADHERE trial baseline assessment), participants underwent 1:1 randomization into either SMS intervention or control (no SMS). Computer-generated randomization was prepared by the biostatistician (JK) in permuted blocks of sizes 6 to 12, stratified by type of exercise performed in TARGET and by exercise adherence at the final TARGET time point (0-1 sessions in the past week arbitrarily classified as *lower adherence* and 2-4 sessions *higher adherence*). Allocation was concealed in a password-protected computer program and accessed by a researcher not involved in enrollment or assessment. Participants were blinded to the study groups and to the study hypothesis through limited disclosure. They were informed at the TARGET trial enrollment that participation was for 9 months, with the initial 3 months comparing two exercise programs and the following 6 months investigating undisclosed adherence strategies, such as a logbook or text messages. To avoid influencing exercise adherence behavior, participants were not informed that two separate, but related, trials were being conducted or that they were being re-randomized into this trial. Outcome assessment was therefore blinded as the participants were deemed *assessors* in this RCT, given outcomes were participant reported. The statisticians were blinded to the group allocation.

Interventions

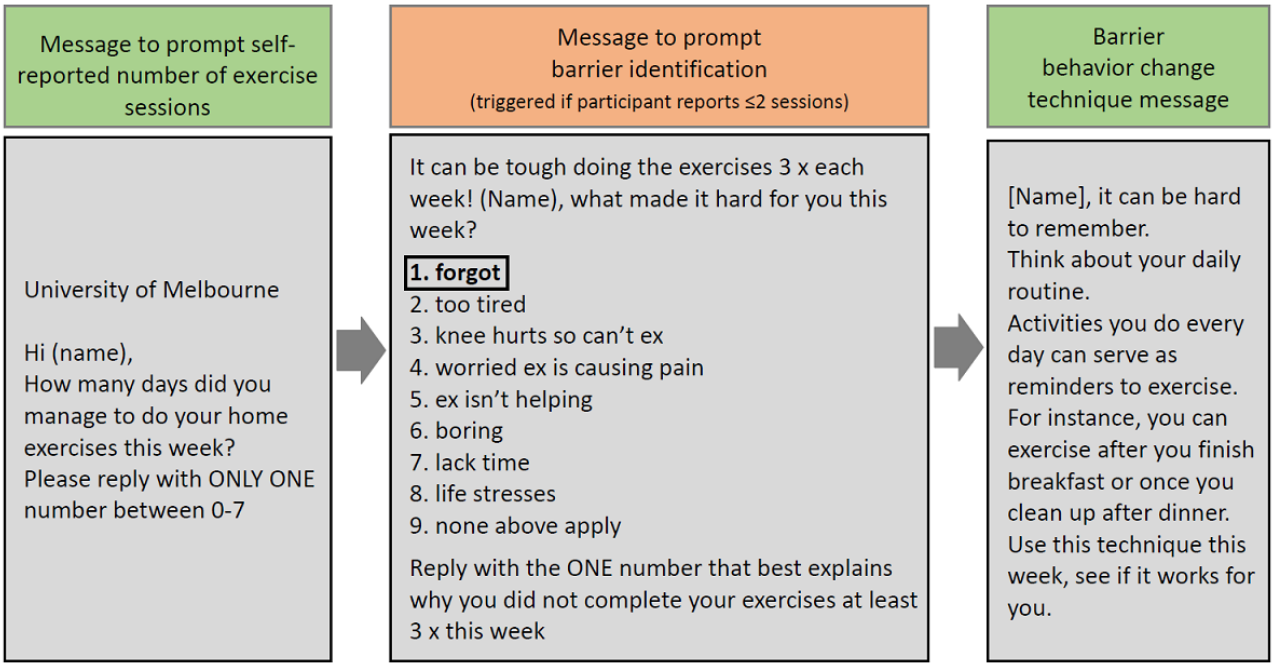
All participants were asked to continue their allocated TARGET prescribed home exercise program unsupervised for 24 weeks [24] but to reduce the frequency from four times per week to three times per week (Multimedia Appendix 2). The frequency was reduced to facilitate adherence over the longer term while still meeting exercise guideline recommendations [25]. In the last appointment in the TARGET trial, the physiotherapists discussed with the participant the importance of progressing the exercises during the subsequent unsupervised phase (eg, by increasing resistance; changing stance surface; and/or varying the number of repetitions, direction, and speed of movement). Participants also received paper-based instructions for each exercise, including how to progress the exercise, and an optional logbook to record their exercise sessions if they wished.

SMS Intervention

Participants received a 24-week automated, semi-interactive SMS intervention delivered via mobile phone to support adherence to the home exercise program. The development of the SMS intervention was based on the Behavior Change Wheel framework [26] and is described elsewhere [18]. In brief, we identified key barriers or facilitators to exercise adherence in knee and hip OA and mapped these to the Theoretical Domains Framework [8]. Behavior change techniques linked to each barrier or facilitator [27] were then used to construct the content of the SMS messages.

Participants received up to five text messages weekly, with message frequency reducing over the 24 weeks. Multimedia Appendices 3 and 4 [18] describe all message types and frequencies, whereas Multimedia Appendix 5 [18] outlines how the automated message sequence functioned. In summary, each week (weeks 1-8) to fortnight (weeks 9-24) participants received a message asking them to self-report the number of home exercise sessions completed in the previous week. Participants who completed ≤2 sessions then received a message prompting them to select their main reason (*barrier*) for not performing exercise sessions as prescribed (3 sessions per week) from a predetermined list (forgot, too tired, knee hurts so cannot exercise, worried exercise is causing pain, exercise is not helping, boring, lack of time, life stress, and none of the above apply to me). Barrier selection then triggered a message providing a suggestion tailored to address the selected barrier (example shown in Figure 1). Those who chose the barrier option of none of the above apply to me received a message encouraging them to continue exercising, but the message was not linked to a specific behavior change technique. Participants who reported being adherent (≥3 exercise session per week) received a positive reinforcement message. Program automation ensured that different messages were received each time. All participants, irrespective of their adherence, also received regular motivational SMS (twice weekly initially then once fortnightly by 24 weeks) containing suggestions linked to exercise facilitators. To enhance engagement, participants received special occasion messages (eg, birthday). Message lengths ranged from 105 to 420 characters, with literacy demands assessed as grade 5.4, well below the maximum eight-grade reading level recommended for consumer health care information [28].

Figure 1. Example automated message sequence for a person with low exercise adherence and reporting their main barrier to exercise as "forgot." Modified from the study by Nelligan et al 2019. Reproduced under the terms of the Creative Commons Attribution 4.0 license.



Control—No SMS

Participants in the control group did not receive any SMS contact.

Outcomes

Outcomes were self-reported and completed electronically (via REDCap) or on paper. The primary outcomes were two measures of adherence, collected at 24 weeks: (1) adherence to

prescribed home exercise using the Exercise Adherence Rating Scale (EARS) Section B and (2) Number of days home exercises completed in the past week. EARS Section B has six items, each scored on a 5-point scale with terminal descriptors of *strongly agree* to *strongly disagree*. The total score ranges between 0 and 24, with higher scores indicating better adherence. This measure has acceptable internal consistency, high test-retest reliability (intraclass correlation coefficients [ICCs] from 0.91 to 0.97), and evidence of construct validity and responsiveness to change [29-31]. Participants were asked "In the past week, how many days did you do your recommended home exercises (maximum of 3 days)?" Response choices ranged from 0 to 3 days. Our test-retest reliability (2-week interval) with such a scale in 54 patients with knee OA was good (ICC [model 2,1]=0.79; 95% CI 0.66-0.87) with fair validity based on agreement with concealed accelerometer-measured session number (Spearman correlations from 0.26 to 0.48 over a 12-week period; method of accelerometer measure reported in the study by Nicolson et al [32]).

Secondary outcomes measured at baseline and 24 weeks, unless otherwise indicated, included the following: (1) adherence to home exercise program three times per week (24-weeks only) based on strength of agreement to the statement "I have been doing my exercise sessions 3 times each week as recommended" using an 11-point NRS with terminal descriptors *strongly disagree*=0 to *strongly agree*=10 [32]; (2) average overall knee pain in the past week using a NRS [33] with terminal descriptors of *no pain* (score=0) and *extreme pain* (score=10) [33]; (3) pain, other symptoms, function in daily living, function in sport and recreation, and knee-related quality of life in the last week using the Knee Injury and Osteoarthritis Outcome Score [34], ranging from 0 to 100, with higher scores indicating better outcomes; (4) health-related quality of life using Assessment of Quality of Life instrument [35] (version AQoL-6D), scores ranging from −0.04 to 1.00 and higher scores indicating better quality of life [35]; (5) Arthritis Self-Efficacy Scale, scores ranging from 0 to 10 and higher scores indicating greater self-efficacy [36]; (6) kinesiophobia using the Brief Fear of Movement Scale for OA, scores ranging from 6 to 24 and higher scores indicating greater kinesiophobia [37]; (7) Pain Catastrophizing Scale, scores ranging from 0 to 52 and higher scores indicating greater catastrophizing [38]; (8) Physical Activity Scale for the Elderly, scores ranging from 0 to >400 and higher scores representing greater physical activity [39]; and (9) participant-reported global overall change using a 7-point scale (terminal descriptors *much worse* to *much better*). Participants who reported *moderately better* and *much better* were classified as improved [40].

Adverse events (any problem participant believed was caused by advice received and required them to seek treatment or take medications and/or interfered with function for ≥2 days) were recorded via a questionnaire at 24 weeks. Medications and other knee OA treatments were recorded at 24 weeks using a customized survey.

Automatically collected SMS data included (1) number who opted to cease receiving messages, (2) mean (SD) number of SMS messages sent per participant, (3) mean (SD) participant reply rate to self-reported exercise sessions, (4) mean (SD) participant reply rate for barrier selection, and (5) group frequency of barriers selected.

Sample Size

We conservatively estimated that 79.6% (102/128 of TARGET participants would be randomized into ADHERE, and of those, 80.3% (82/102) would be retained at week 24. We chose a moderate effect size of 0.6, given that smaller effects are unlikely to be clinically relevant [41]. With 40 participants per group, we would have 83% power to detect an effect size of 0.6 with a two-sided significance level of .05, assuming a correlation between baseline home exercise adherence and adherence outcomes at 24 weeks of 0.4, based on data from our previous trials [42-44], and including baseline adherence as a covariate in regression models.

Statistical Methods

Analyses were performed by biostatisticians (JK and SC) using Stata (StataCorp, version 16) software and intention-to-treat. Baseline characteristics of participants who did and did not provide both primary outcomes were compared using *t* tests or chi-square tests. Missing outcomes were imputed using chained equations with predictive mean matching and five nearest neighbors for continuous outcomes and logistic regression imputation models for binary improvement outcomes. Data were imputed for each group separately. Imputation models for continuous outcomes at 24 weeks included all baseline and outcome variables, where appropriate. Imputation models for binary variables omitted all outcome variables because of the potential for perfect prediction, including only baseline variables. Estimates from 20 imputed data sets were combined using Rubin's rules [45]. Standard diagnostic plots assessed the validity of model assumptions and imputed data sets. For the primary outcome of exercise adherence EARS Section B, the mean between-group difference at week 24 was estimated using a linear regression model adjusted for baseline measures and the stratifying variables of the TARGET exercise group and dichotomized baseline adherence. For the primary outcome of number of days home exercises completed in the past week and the secondary outcome of adherence to home exercise, the mean between-group difference at week 24 was estimated using linear regression models, adjusted only for the stratifying variables. For the continuous secondary outcomes, the mean between-group difference in change (baseline minus follow-up) at week 24 was estimated using linear regression models adjusted for baseline measures and stratifying variables. The proportion of participants with overall self-perceived improvement was compared between groups using a logistic regression model adjusted for stratifying variables, with results presented as odds ratios and risk ratios. Complete case analyses were also conducted.

Results

Participants

We randomized 110 participants (56 SMS group and 54 no SMS), with 99 (90.0%) completing both primary outcome measures at week 24 (48/56, 86% SMS group and 51/54, 94% no SMS; Figure 2). The sample had a mean (SD) BMI of 37.3

(SD 6.4) kg/m^2 and were predominantly female (74/110, 67.2%). Groups were comparable at baseline (Table 1). Participants who provided both primary outcomes were comparable with those who were missing at least one (Multimedia Appendix 6).

Figure 2. Flow diagram of preceding TARGET trial and ADHERE trial procedures. PT: physiotherapy.

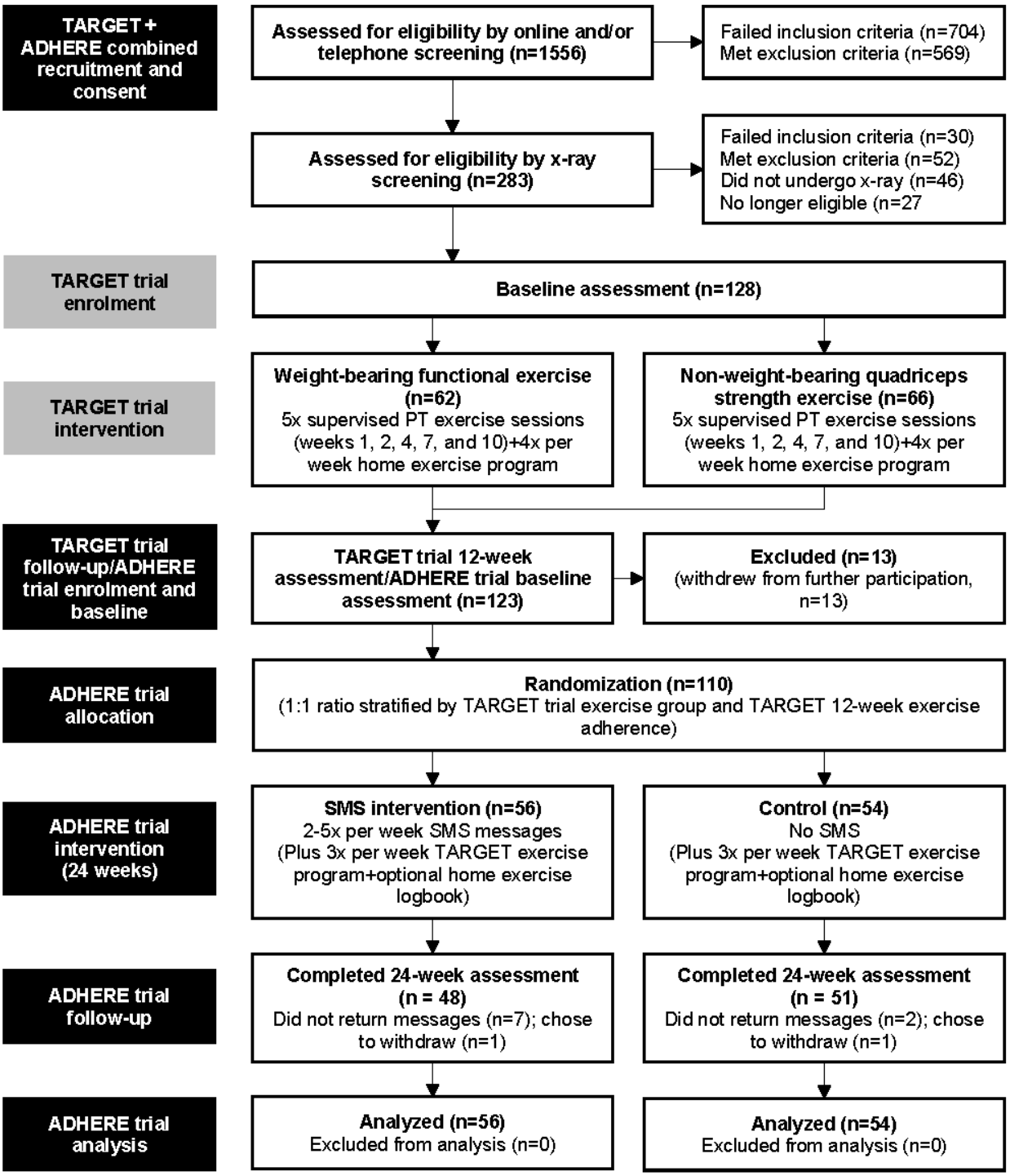


Table 1. Baseline characteristics of participants by group.

Characteristics	SMS (n=56)	Control (n=54)
Age (years), mean (SD)	61.7 (6.7)	62.9 (6.8)
Males, n (%)	21 (38)	15 (28)
Height (m), mean (SD)	166.1 (8.5)	165.4 (9)
Body mass (kg), mean (SD)	102.7 (18.6)	102.3 (18.3)
BMI (kg/m^2), mean (SD)	37.3 (6.8)	37.4 (6)
Radiographic disease severity KL[a] grade, n (%)		
2	9 (16)	11 (20)
3	35 (63)	32 (59)
4	12 (21)	11 (20)
Currently employed, n (%)	35 (63)	32 (59)
Symptom duration (years), mean (SD)	8.1 (6.9)	8.3 (8)
Unilateral symptoms, n (%)	13 (23)	11 (20)
Problems in other joints, n (%)	45 (80)	41 (76)
Hand	22 (39)	20 (37)
Neck	23 (41)	7 (13)
Back	30 (54)	23 (43)
Hip	20 (36)	17 (31)
Foot	17 (30)	13 (24)
Shoulder	17 (30)	10 (19)
Treatments for knee OA[b] in the last 6 months, n (%)[c,d]		
At least one treatment	45 (80)	43 (80)
Hot or cold treatment	15 (27)	17 (31)
Shoe orthotics	4 (7)	9 (17)
Massage	10 (18)	9 (17)
Knee braces	6 (11)	3 (6)
Hydrotherapy	5 (9)	6 (11)
Manual therapy	4 (7)	2 (4)
Walking stick	4 (7)	3 (6)
Acupuncture	4 (7)	0 (0)
Arthroscopic surgery	0 (0)	0 (0)
Transcutaneous electrical nerve stimulation	1 (2)	0 (0)
Ultrasound	0 (0)	0 (0)
Injections	0 (0)	0 (0)
Current pain medication use, n (%) [c,d]	39 (70)	40 (74)
Nonsteroidal anti-inflammatories	20 (36)	23 (43)
Cyclooxygenase-2 inhibitors	2 (4)	4 (7)
Analgesics (paracetamol combinations)	31 (55)	34 (63)
Topical anti-inflammatories	21 (38)	15 (28)
Oral corticosteroids	0 (0)	0 (0)
Oral opioids	3 (5)	0 (0)
TARGET exercise group, n (%)		

Characteristics	SMS (n=56)	Control (n=54)
Nonweight-bearing exercise	30 (54)	28 (52)
Weight-bearing exercise	26 (46)	26 (48)
TARGET adherence to home exercise[e,f], mean (SD)	18.1 (6.8)	17.8 (6.6)
TARGET no of days exercised in past week[f], mean (SD)	2.8 (1.5)	2.4 (1.5)

[a]KL: Kellgren and Lawrence.

[b]OA: osteoarthritis.

[c]Defined as at least once per week in the prior month.

[d]Numbers do not add up to total as participants could choose more than one.

[e]Measured by the Exercise Adherence Rating Scale (EARS) Section B.

[f]Measured at week 12 in the TARGET trial.

A total of 17 participants reported adverse events (none serious), mostly increased knee pain or pain elsewhere (Multimedia Appendix 7). The use of pain medications and co-interventions was similar between groups (Multimedia Appendix 7).

In the SMS group, two participants chose to stop receiving messages. Over the 24 weeks, the mean (SD) number of SMS sent to each participant was 57.9 (SD 9.1) messages. The mean (SD) reply rate per participant to self-reporting home exercise sessions was 66% (SD 34%) and to selecting a barrier (if <3 exercise sessions reported) was 73% (SD 35%). Across the group, the most commonly chosen barrier was lack of time (n=44), followed by none apply (n=43), life stress (n=28), knee hurts so cannot exercise (n=16), worried exercise is causing pain (n=7), forgot (n=6), too tired (n=5), exercise is not helping (n=2), and exercise is boring (n=1).

Primary Outcomes

Both primary outcomes provided evidence of greater home exercise adherence with the SMS intervention compared with control (Tables 2 and 3). At week 24, the SMS group reported higher scores on the EARS (mean 16.5, SD 6.5 vs mean 13.3, SD 7.0; mean difference 3.1, 95% CI 0.8-5.5; *P*=.01) and more days performing home exercise in the past week (mean 1.8, SD 1.2 vs mean 1.3, SD 1.2; mean difference 0.6, 95% CI 0.2-1.0; *P*=.01) than the control group. Specifically, in the SMS group, 23% (11/48) participants did not perform home exercises in the past week, 8% (4/48) performed home exercises on 1 day, 29% (14/48) on 2 days, and 40% (19/48) on 3 days in the past week. In the control group, 35% (18/51) participants did not perform home exercises in the past week, 20% (10/51) performed home exercises on 1 day, 22% (11/51) on 2 days, and 24% (12/51) on 3 days in the past week. Analyses using complete case data yielded similar results (Multimedia Appendix 8).

Table 2. Mean (SD) scores on continuous outcome measures across time, by group.

Outcomes	Baseline		24 weeks	
	SMS (n=56), mean (SD)	Control (n=54), mean (SD)	SMS (n=48)[a], mean (SD)	Control (n=49)[b], mean (SD)
Primary outcomes				
Adherence to prescribed home exercise (EARS[c] Section B)[d]	__[e]	—	16.3 (6.6)	13.4 (7.1)
Number of days on which home exercises were completed in the past week[d]	—	—	1.9 (1.2)	1.3 (1.2)
Secondary outcomes				
Adherence to home exercise thrice weekly (NRS)[d,f,g]	—	—	6.0 (3.8)	5.1 (3.7)
Overall average knee pain (NRS)[h]	3.5 (2.1)	3.8 (2.4)	4.1 (2.2)	4.0 (2.3)
Pain (KOOS)[i]	64.3 (14.9)	63.2 (19.8)	64.9 (17.3)	64.4 (20.1)
Other symptoms (KOOS)	64.3 (17.0)	64.3 (17.4)	66.6 (18.5)	64.9 (18.3)
Function (KOOS)	72.2 (15.6)	70.6 (20.7)	72.4 (17.6)	70.0 (21.1)
Sport and recreation (KOOS)	33.4 (22.3)	39.5 (23.3)	37.3 (24.8)	41.6 (27.7)
Knee-related quality of life (KOOS)	44.4 (19.9)	47.9 (21.7)	46.1 (22.0)	47.8 (23.0)
Health-related quality of life (AQoL)[j]	0.76 (0.18)	0.81 (0.12)	0.77 (0.15)	0.78 (0.15)
Self-efficacy: pain (ASES)[k]	6.6 (2.0)	6.9 (2.1)	6.6 (2.2)	6.4 (2.1)
Self-efficacy: function (ASES)	8.4 (1.2)	7.8 (2.2)	8.3 (1.5)	8.2 (1.7)
Self-efficacy: other (ASES)	7.2 (1.8)	7.5 (2.1)	7.4 (2.0)	7.6 (1.9)
Kinesiophobia (BFOMS)[l]	12.5 (3.4)	12.2 (4.0)	12.1 (3.6)	12.1 (3.8)
Pain catastrophizing (PCS)[m]	6.0 (7.7)	7.4 (9.9)	6.9 (9.6)	6.2 (7.1)
Physical activity (PASE)[n]	176.7 (86.9)	173.9 (82.8)	190.5 (111.3)	174.0 (71.0)

[a]n=48 for both primary outcomes and n=45 for all secondary outcomes.

[b]n=49 for both primary outcomes and n=45 for sport and recreation (KOOS), AQoL, BFOMS, and PASE. n=46 for all other secondary outcomes.

[c]EARS: Exercise Adherence Rating Scale (0-24; higher scores indicate better adherence).

[d]Adherence data only collected at 24 weeks.

[e]Represents data about adherence that can only be collected at 24 weeks and not baseline.

[f]NRS: numeric rating scale.

[g]Adherence to home exercise thrice weekly—agreement with statement "I have been doing my exercise sessions 3 times each week as recommended" with responses collected using an 11-point NRS and terminal descriptors *strongly disagree*=0 to *strongly agree*=10.

[h]Overall average pain NRS (0-10; higher scores indicate worse pain).

[i]KOOS: Knee Injury and Osteoarthritis Outcome Score (0 to 100; lower scores indicate worse pain/symptoms/function/quality of life).

[j]AQoL: Assessment of Quality of Life instrument (−0.04 to 1.0; higher scores indicate better quality of life).

[k]ASES: Arthritis Self-Efficacy Scale (1-10; higher scores indicate better efficacy).

[l]BFOMS: Brief Fear of Movement Scale (6-24; higher scores indicate greater fear).

[m]PCS: Pain Catastrophizing Scale (0-52; higher scores indicate greater catastrophizing).

[n]PASE: Physical Activity Scale for the Elderly (0-400+; higher scores indicate greater activity).

Table 3. Mean (SD) scores at week 24 or mean (SD) change within groups, from baseline to week 24, and mean (95% CI) difference between groups (adjusted for baseline value of outcome, TARGET exercise group, and dichotomized baseline adherence), for continuous outcomes, using multiply imputed data.

Outcomes	SMS, mean (SD)	Control, mean (SD)	SMS−control[a], mean difference (95% CI)	*P* value
Mean (SD) at week 24 and mean (95% CI) difference between groups				
Primary outcomes				
Adherence to prescribed home exercise (EARS[b] Section B)	16.5 (6.5)	13.3 (7.0)	3.1 (0.8 to 5.5)	.01
Number of days on which home exercises were completed in the past week[c,d]	1.8 (1.2)	1.3 (1.2)	0.6 (0.2 to 1.0)	.01
Secondary outcomes				
Adherence to home exercise three times weekly (NRS)[c,e]	6.0 (3.8)	4.9 (3.7)	1.1 (−0.4 to 2.6)	.14
Mean (SD) change within group (baseline minus week 24) and mean (95% CI) difference in change between groups				
Overall average knee pain (NRS)[d,f]	−0.6 (2.4)	−0.2 (2.2)	−0.2 (−1.1 to 0.6)	.59
Pain (KOOS)[g]	−0.8 (14.9)	−2.6 (14.1)	1.3 (−4.6 to 7.3)	.66
Other symptoms (KOOS)	−2.9 (17.3)	−1.7 (13.5)	−1.2 (−7.5 to 5.0)	.70
Function (KOOS)	−0.0 (18.5)	−0.5 (14.0)	−0.2 (−6.7 to 6.3)	.95
Sport and recreation (KOOS)	−3.5 (22.1)	−3.0 (21.9)	1.2 (−8.4 to 10.8)	.81
Knee-related quality of life (KOOS)	−2.2 (23.0)	−2.3 (16.2)	1.3 (−6.7 to 9.4)	.75
Health-related quality of life (AQoL)[h]	0.00 (0.13)	0.03 (0.15)	−0.01 (−0.06 to 0.04)	.68
Self-efficacy: pain (ASES)[i]	0.0 (2.1)	0.6 (2.6)	−0.4 (−1.2 to 0.4)	.35
Self-efficacy: function (ASES)	0.1 (1.7)	−0.4 (2.1)	0.1 (−0.5 to 0.7)	.78
Self-efficacy: other (ASES)	−0.0 (1.9)	−0.1 (2.3)	0.1 (−0.6 to 0.9)	.69
Kinesiophobia (BFOMS)[j,k]	0.4 (2.6)	−0.2 (3.6)	0.5 (−0.8 to 1.7)	.44
Pain catastrophizing (PCS)[k,l]	−1.9 (7.9)	0.9 (10.1)	−2.0 (−5.2 to 1.2)	.22
Physical activity (PASE)[m]	−15.1 (90.1)	0.9 (82.6)	−17.5 (−53.0 to 18.0)	.33

[a]For mean difference between groups, positive differences favor SMS.

[b]EARS: Exercise Adherence Rating Scale (0-24; higher scores, better adherence).

[c]Not adjusted for baseline value of outcome.

[d]Adherence to home exercise thrice weekly: agreement with statement “I have been doing my exercise sessions 3 times each week as recommended” collected using an 11-point numeric rating scale (NRS) and terminal descriptors *strongly disagree*=0 to *strongly agree*=10.

[e]NRS: numeric rating scale.

[f]Overall average knee pain NRS (0-10; higher scores, worse pain).

[g]KOOS: Knee Injury and Osteoarthritis Outcome Score (0-100; lower scores, worse pain/symptoms/function/quality of life).

[h]AQoL: Assessment of Quality of Life instrument (−0.04 to 1.0; higher scores, better quality of life).

[i]ASES: Arthritis Self-Efficacy Scale (1-10; higher scores, better efficacy).

[j]BFOMS: Brief Fear of Movement Scale (6-24; higher scores indicate greater fear).

[k]For change within groups, positive changes indicate improvement.

[l]PCS: Pain Catastrophizing Scale (0-52; higher scores, greater catastrophizing).

[m]PASE: Physical Activity Scale for the Elderly (0-400+; higher scores, greater activity).

Secondary Outcomes

There was no evidence of a between-group difference in adherence (NRS) at 24 weeks or in change in clinical outcomes (Tables 2 and 3). Within-group changes in both groups were relatively small (Tables 2 and 3). Similar proportions of participants in both groups reported global improvement overall since baseline (SMS 23/45, 51% vs control 19/46, 41%; odds ratios 1.48, 95% CI 0.61-3.57; *P*=.38; relative risk, 1.20, 95% CI 0.70-1.71; *P*=.39). Analyses using complete case data yielded similar results (Multimedia Appendix 9).

Discussion

Principal Findings

We found that an automated behavior change theory–informed, semi-interactive SMS intervention improved self-reported adherence to a prescribed unsupervised home-based exercise program over 24 weeks, evidenced by both primary outcomes, when compared with no SMS contact in people with knee OA and obesity. However, greater adherence to home exercise with SMS support did not translate into improvements in secondary clinical outcomes.

The greater exercise adherence may be linked to the rigorous development of the SMS program based on a widely used framework, the Behavior Change Wheel [26]. As our program incorporated several elements, we could not determine which elements were most effective in eliciting the desired exercise behavior. Regular receipt of a message asking to self-report exercise completion can act as a reminder, and self-monitoring behavior increases physical activity adherence in patients with OA [46] and in adults who are overweight or have obesity [47]. Our bank of 198 different text messages included 20 behavior change techniques targeting 13 modifiable barriers and 9 facilitators to exercise previously identified in people with OA [8]. As the SMS program was semipersonalized, the number and type of behavior change techniques employed differed across participants, depending on their adherence and barrier selection. However, in 28% of barrier replies, participants selected *none apply*, meaning that no targeted response message could be sent, which may have attenuated the benefits of the program.

Despite greater adherence to home exercise in the SMS group, we found no between-group difference in any secondary clinical outcome. Both groups either maintained or had slight diminution of the clinical improvements resulting from the preceding 12-week physiotherapist-supervised exercise phase (TARGET trial) [24]. This lack of further improvement may relate to inadequate exercise progression, including suboptimal intensities, during this unsupervised phase. Although it is presumed that there is a relationship between exercise adherence and outcomes in knee OA, surprisingly, the nature of the relationship is unclear given the limited research and contradictory findings [6,48-50]. For example, some studies using self-report adherence measures have found that greater adherence is linked with better pain and function outcomes [6,48,50]. In contrast, a study using concealed accelerometers to accurately assess adherence to a 12-week home strengthening program showed no evidence of an association [49]. It is possible that the greater adherence seen in our SMS group may have been of insufficient magnitude (between-group difference of 3.1 EARS units and 0.6 days exercised) to impact clinical outcomes, especially as adherence actually decreased in both groups from relatively high levels after the supervised exercise phase. This is supported by a recent study that showed a minimally important change of 5.5 units for the Brazilian Portuguese version of the EARS, albeit in people with chronic low back pain [30]. Another explanation relates to the multi-dimensional nature of adherence, which is not fully captured in our measures. For example, exercise regularity and intensity may be elements of the prescribed exercise program that may be important to clinical outcomes [51].

How our Findings Compare With Those of Other Studies

Although numerous strategies are suggested to improve adherence to exercise in people with OA, many of which are clinician-centric such as *booster sessions*, there is limited RCT evidence available to inform clinical practice [11]. To our knowledge, only three pilot or feasibility studies have specifically investigated an SMS intervention aimed at improving exercise adherence in people with knee OA [15-17] and none measured adherence. In one study, short video messages providing visual prompts to home exercises were sent every second day for 6 weeks to 5 people [16]. Participants found the messages *very useful*, and although there was no difference in functional outcomes compared with a control group (n=9), the direction of improvement favored the intervention. In another study, 27 people received four activity-promoting text messages per week for 6 weeks to reinforce content from an OA educational booklet. Participants meaningfully engaged with the intervention (100% read messages and 89% replied), reporting it to be enjoyable and personally relevant. Improvements were seen in perceptions of exercise and pain [15]. In the third study, 19 people who had completed an education and exercise program (Good Life with Arthritis: Denmark) received text messages providing general physical activity advice, three times per week for 6 weeks, whereas the control group received no text messages (n=19) [17]. The results did not indicate any potentially beneficial effects of the intervention on physical inactivity or clinical outcomes. As we cannot draw conclusions about efficacy from these studies, we provide the first evidence from a fully powered RCT about the effectiveness of an SMS strategy on adherence to a home exercise program in people with knee OA. Our results also concur with other RCTs in healthy older adults [52] and in those with frozen shoulder [53], which found that automated SMS programs can improve self-reported adherence to home-based exercise.

Strengths and Limitations of the Study

The strengths of our study include a systematically designed SMS intervention informed by behavior change theory; participant, and thereby assessor, blinding to group allocation; excellent retention at 24 weeks; and acceptable user engagement with the SMS program. There are some limitations to this study. First, accurate measurement of exercise adherence is challenging [54], and there is no gold standard. As is the case with most

studies [54], we used self-reported exercise adherence measures to ensure feasibility, given our large sample and extended time frame. However, self-report measures are subjected to recall bias and typically overestimate [32] due to social desirability bias. This latter bias may be accentuated in the SMS group, given that it received regular reminders about the importance of exercise. Second, only those who completed the preceding TARGET study were enrolled as participants. This may have introduced selection bias, particularly by increasing the likelihood that a more adherent group was enrolled, which could make it more difficult to detect an effect of the SMS program on exercise adherence. However, this is less likely given that 90% of the original TARGET study participants took part. Third, we do not know whether the improved exercise adherence is sustainable over time once SMS contact ceases or whether our findings can be generalized to patients who may be less motivated than research volunteers or to those without obesity or generalized to a home exercise program that is unsupervised from its outset. Nonetheless, the characteristics of our sample broadly reflect those of the general knee OA patient population, which includes greater proportion of females, and people who are more likely to have overweight or obesity, and be of an older age [55].

Implications for Clinical Practice and Future Research

Our results provide preliminary evidence that the use of SMS may promote patient adherence to a core recommended OA treatment (exercise), which is an important clinical priority [56]. Mobile phone text messaging programs are an increasingly popular method for delivering health behavior change interventions [12]. This is unsurprising given the ubiquitous use of mobile phones across all populations and age groups and the many benefits of using SMS technology such as convenience, instantaneous communication, cost-effectiveness [57], and user acceptability. Our SMS program is a scalable, inexpensive intervention that could be incorporated into existing or future web-based exercise resources and/or used by clinicians to enhance adherence of their patients to their own exercise prescription. Although we chose SMS as the delivery mode, the message content and program could be converted into formats suitable for delivery by other communication forms such as email or a mobile app. It could also be adapted for use in other health conditions where exercise is a core treatment and its effectiveness is tested. Further research into modification of the program and its implementation is warranted to optimize exercise behavior change and impact clinical outcomes. This could include messages that better address each person's unique exercise barriers, for example, use of the program at more distal time points when adherence substantially declines and symptomatic benefits are reduced and testing the program in a pragmatic setting where patients are less motivated to exercise at the outset. Our results also highlight the need for further research to better understand the nature of the relationship between exercise adherence and clinical outcomes.

Conclusions

This study showed that a behavior change theory–informed SMS program increased self-reported adherence to unsupervised home exercise over 6 months, following an initial period of exercise supervised by a physiotherapist, in people with knee OA and obesity. However, greater exercise adherence did not translate into improvements in pain and function.

Acknowledgments

The authors gratefully acknowledge Professor Lou Atkins from the Centre for Behaviour Change, University College London, United Kingdom, for assisting with formulating the text message content and Mr Ben Metcalf for assisting with manuscript preparation. The study was funded by the National Health and Medical Research Council Program Grant (#1091302). The funder played no role in the study design; data collection, analysis, or interpretation; or manuscript preparation. KLB is supported by a National Health and Medical Research Council Investigator Fellowship (#1174431). RKN is supported by an Australian Government Research Training Program Scholarship. RSH is supported by a National Health and Medical Research Council Senior Research Fellowship (#1154217).

Authors' Contributions

KB and RH conceived the study. KB led the trial and obtained funding. KB, RH, RN, JK, AK, and SS designed the protocol. JK formulated and was responsible for the sample size; randomization schedule; and, together with SC, statistical analyses. RN, KB, and RH developed the SMS intervention. AK and SS recruited participants and coordinated the trial. KB drafted the manuscript. All authors provided input into manuscript preparation and approved the final manuscript.

Conflicts of Interest

KB receives royalties from Wolster Kluxers for UpToDate osteoarthritis clinical guidelines.

Multimedia Appendix 1
Exclusion criteria.
[DOCX File , 25 KB - jmir_v22i9e21749_app1.docx]

Multimedia Appendix 2
Description of the two exercise programs.

[DOCX File , 29 KB - jmir_v22i9e21749_app2.docx]

Multimedia Appendix 3
Description and frequency of behavior change messages included in the SMS intervention.
[DOCX File , 86 KB - jmir_v22i9e21749_app3.docx]

Multimedia Appendix 4
Description and frequency of logistic and other messages included in the SMS intervention.
[DOCX File , 54 KB - jmir_v22i9e21749_app4.docx]

Multimedia Appendix 5
SMS intervention message sequence including message interactions and triggers.
[DOCX File , 61 KB - jmir_v22i9e21749_app5.docx]

Multimedia Appendix 6
Baseline characteristics of participants who were missing at least one primary outcome and those who provided both primary outcomes.
[DOCX File , 29 KB - jmir_v22i9e21749_app6.docx]

Multimedia Appendix 7
Number (percentage) of participants with adverse events, co-interventions, or pain medication use during the 24 weeks.
[DOCX File , 31 KB - jmir_v22i9e21749_app7.docx]

Multimedia Appendix 8
Means (SD) at week 24 or mean (SD) change within groups, from baseline to week 24, and mean (95% CI) difference between groups (adjusted for baseline value of outcome, TARGET exercise group and dichotomized baseline adherence), for continuous outcomes, using complete case data.
[DOCX File , 32 KB - jmir_v22i9e21749_app8.docx]

Multimedia Appendix 9
Number (percentage) of participants reporting global improvement (adjusted for TARGET exercise group and dichotomized baseline adherence), using complete case data.
[DOCX File , 28 KB - jmir_v22i9e21749_app9.docx]

Multimedia Appendix 10
CONSORT E-HEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1878 KB - jmir_v22i9e21749_app10.pdf]

References

1. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
2. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020 Feb;72(2):220-233. [doi: 10.1002/art.41142] [Medline: 31908163]
3. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589. [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
4. Swain S, Sarmanova A, Coupland C, Doherty M, Zhang W. Comorbidities in osteoarthritis: a systematic review and meta-analysis of observational studies. Arthritis Care Res (Hoboken) 2020 Jul;72(7):991-1000. [doi: 10.1002/acr.24008] [Medline: 31207113]
5. Nicolson PJ, Hinman RS, Kasza J, Bennell KL. Trajectories of adherence to home-based exercise programs among people with knee osteoarthritis. Osteoarthritis Cartilage 2018 Apr;26(4):513-521 [FREE Full text] [doi: 10.1016/j.joca.2018.01.009] [Medline: 29360592]
6. Pisters MF, Veenhof C, Schellevis FG, Twisk JW, Dekker J, de Bakker DH. Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. Arthritis Care Res (Hoboken) 2010 Aug;62(8):1087-1094 [FREE Full text] [doi: 10.1002/acr.20182] [Medline: 20235201]

7. Kanavaki AM, Rushton A, Efstathiou N, Alrushud A, Klocke R, Abhishek A, et al. Barriers and facilitators of physical activity in knee and hip osteoarthritis: a systematic review of qualitative evidence. BMJ Open 2017 Dec 26;7(12):e017042 [FREE Full text] [doi: 10.1136/bmjopen-2017-017042] [Medline: 29282257]
8. Dobson F, Bennell KL, French SD, Nicolson PJ, Klaasman RN, Holden MA, et al. Barriers and facilitators to exercise participation in people with hip and/or knee osteoarthritis: synthesis of the literature using behavior change theory. Am J Phys Med Rehabil 2016 May;95(5):372-389. [doi: 10.1097/PHM.0000000000000448] [Medline: 26945211]
9. Fransen M, McConnell S, Harmer AR, van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015 Jan 9;1:CD004376. [doi: 10.1002/14651858.CD004376.pub3] [Medline: 25569281]
10. Rausch Osthoff A, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018 Sep;77(9):1251-1260. [doi: 10.1136/annrheumdis-2018-213585] [Medline: 29997112]
11. Nicolson PJ, Bennell KL, Dobson FL, Van Ginckel A, Holden MA, Hinman RS. Interventions to increase adherence to therapeutic exercise in older adults with low back pain and/or hip/knee osteoarthritis: a systematic review and meta-analysis. Br J Sports Med 2017 May;51(10):791-799. [doi: 10.1136/bjsports-2016-096458] [Medline: 28087567]
12. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
13. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
14. Palmer M, Sutherland J, Barnard S, Wynne A, Rezel E, Doel A, et al. The effectiveness of smoking cessation, physical activity/diet and alcohol reduction interventions delivered by mobile phones for the prevention of non-communicable diseases: a systematic review of randomised controlled trials. PLoS One 2018;13(1):e0189801 [FREE Full text] [doi: 10.1371/journal.pone.0189801] [Medline: 29304148]
15. Blake H. Motive8! Feasibility of a text messaging intervention to promote physical activity in knee osteoarthritis. Int J Sports Exerc Med 2015 Dec 31;1(5):-. [doi: 10.23937/2469-5718/1510027]
16. Dar G, Marx Y, Ioffe E, Kodesh E. The effectiveness of a multimedia messaging service reminder system in the management of knee osteoarthritis: a pilot study. Int J Clin Med 2014;05(09):483-489. [doi: 10.4236/ijcm.2014.59067]
17. Bartholdy C, Bliddal H, Henriksen M. Effectiveness of text messages for decreasing inactive behaviour in patients with knee osteoarthritis: a pilot randomised controlled study. Pilot Feasibility Stud 2019;5:112 [FREE Full text] [doi: 10.1186/s40814-019-0494-6] [Medline: 31516729]
18. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A short message service intervention to support adherence to home-based strengthening exercise for people with knee osteoarthritis: intervention design applying the behavior change wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
19. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309 [FREE Full text] [doi: 10.7326/0003-4819-148-4-200802190-00008] [Medline: 18283207]
20. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
21. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
22. Slade SC, Dionne CE, Underwood M, Buchbinder R, Beck B, Bennell K, et al. Consensus on exercise reporting template (CERT): modified delphi study. Phys Ther 2016 Oct;96(10):1514-1524. [doi: 10.2522/ptj.20150668] [Medline: 27149962]
23. Nelligan RK, Hinman RS, Kasza J, Schwartz S, Kimp A, Atkins L, et al. Effect of a short message service (SMS) intervention on adherence to a physiotherapist-prescribed home exercise program for people with knee osteoarthritis and obesity: protocol for the ADHERE randomised controlled trial. BMC Musculoskelet Disord 2019 Sep 14;20(1):428 [FREE Full text] [doi: 10.1186/s12891-019-2801-z] [Medline: 31521136]
24. Bennell KL, Nelligan RK, Kimp AJ, Schwartz S, Kasza J, Wrigley TV, et al. What type of exercise is most effective for people with knee osteoarthritis and co-morbid obesity?: The TARGET randomized controlled trial. Osteoarthritis Cartilage 2020 Jun;28(6):755-765. [doi: 10.1016/j.joca.2020.02.838] [Medline: 32200051]
25. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, American College of Sports Medicine. American college of sports medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
26. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Great Britain, UK: Silverback Publishing; 2014.

27. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
28. Schilling L, Bennett B, Bull S, Kempe A, Wretling MP, Staton E. Text Messaging in Healthcare Research Toolkit: The Center for Research in Implementation Science and Prevention (CRISP). Denver: University of Colorado. 2013. URL: https://www.careinnovations.org/wp-content/uploads/2017/11/Text_Messaging_in_Healthcare_Research_Toolkit_2.pdf [accessed 2020-06-11]
29. Newman-Beinart NA, Norton S, Dowling D, Gavriloff D, Vari C, Weinman JA, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the exercise adherence rating scale (EARS). Physiotherapy 2017 Jun;103(2):180-185. [doi: 10.1016/j.physio.2016.11.001] [Medline: 27913064]
30. de Lira MR, de Oliveira AS, França RA, Pereira AC, Godfrey EL, Chaves TC. The brazilian portuguese version of the exercise adherence rating scale (EARS-Br) showed acceptable reliability, validity and responsiveness in chronic low back pain. BMC Musculoskelet Disord 2020 May 12;21(1):294 [FREE Full text] [doi: 10.1186/s12891-020-03308-z] [Medline: 32398068]
31. Meade LB, Bearne LM, Godfrey EL. Comprehension and face validity of the exercise adherence rating scale in patients with persistent musculoskeletal pain. Musculoskeletal Care 2018 Sep;16(3):409-412. [doi: 10.1002/msc.1240] [Medline: 29575381]
32. Nicolson PJ, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Self-reported home exercise adherence: a validity and reliability study using concealed accelerometers. J Orthop Sports Phys Ther 2018 Dec;48(12):943-950. [doi: 10.2519/jospt.2018.8275] [Medline: 30053792]
33. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov;63(Suppl 11):S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
34. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
35. Osborne RH, Hawthorne G, Lew EA, Gray LC. Quality of life assessment in the community-dwelling elderly: validation of the assessment of quality of life (AQOL) instrument and comparison with the SF-36. J Clin Epidemiol 2003 Feb;56(2):138-147. [doi: 10.1016/s0895-4356(02)00601-7] [Medline: 12654408]
36. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989 Jan;32(1):37-44. [doi: 10.1002/anr.1780320107] [Medline: 2912463]
37. Shelby RA, Somers TJ, Keefe FJ, DeVellis BM, Patterson C, Renner JB, et al. Brief fear of movement scale for osteoarthritis. Arthritis Care Res (Hoboken) 2012 Jun;64(6):862-871 [FREE Full text] [doi: 10.1002/acr.21626] [Medline: 22290689]
38. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The pain catastrophizing scale: further psychometric evaluation with adult samples. J Behav Med 2000 Aug;23(4):351-365. [doi: 10.1023/a:1005548801037] [Medline: 10984864]
39. Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): development and evaluation. J Clin Epidemiol 1993 Feb;46(2):153-162. [doi: 10.1016/0895-4356(93)90053-4] [Medline: 8437031]
40. ten Klooster PM, Drossaers-Bakker KW, Taal E, van de Laar MA. Patient-perceived satisfactory improvement (PPSI): interpreting meaningful change in pain from the patient's perspective. Pain 2006 Mar;121(1-2):151-157. [doi: 10.1016/j.pain.2005.12.021] [Medline: 16472915]
41. McAlindon TE, Driban JB, Henrotin Y, Hunter DJ, Jiang G, Skou ST, et al. OARSI clinical trials recommendations: design, conduct, and reporting of clinical trials for knee osteoarthritis. Osteoarthritis Cartilage 2015 May;23(5):747-760 [FREE Full text] [doi: 10.1016/j.joca.2015.03.005] [Medline: 25952346]
42. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94 [FREE Full text] [doi: 10.1002/acr.22915] [Medline: 27111441]
43. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
44. Bennell KL, Kyriakides M, Metcalf B, Egerton T, Wrigley TV, Hodges PW, et al. Neuromuscular versus quadriceps strengthening exercise in patients with medial knee osteoarthritis and varus malalignment: a randomized controlled trial. Arthritis Rheumatol 2014 Apr;66(4):950-959 [FREE Full text] [doi: 10.1002/art.38317] [Medline: 24757146]
45. Carpenter J, Kenward M. Multiple Imputation and Its Application. West Sussex, UK: John Wiley & Sons Ltd; 2013.
46. Willett M, Duda J, Fenton S, Gautrey C, Greig C, Rushton A. Effectiveness of behaviour change techniques in physiotherapy interventions to promote physical activity adherence in lower limb osteoarthritis patients: a systematic review. PLoS One 2019;14(7):e0219482 [FREE Full text] [doi: 10.1371/journal.pone.0219482] [Medline: 31291326]

47. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
48. van Gool CH, Penninx BW, Kempen GI, Rejeski WJ, Miller GD, van Eijk JT, et al. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. Arthritis Rheum 2005 Feb 15;53(1):24-32 [FREE Full text] [doi: 10.1002/art.20902] [Medline: 15696558]
49. Nicolson PJ, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Effects of covertly measured home exercise adherence on patient outcomes among older adults with chronic knee pain. J Orthop Sports Phys Ther 2019 Jul;49(7):548-556. [doi: 10.2519/jospt.2019.8843] [Medline: 31258045]
50. Moseng T, Dagfinrud H, van Bodegom-Vos L, Dziedzic K, Hagen KB, Natvig B, et al. Low adherence to exercise may have influenced the proportion of OMERACT-OARSI responders in an integrated osteoarthritis care model: secondary analyses from a cluster-randomised stepped-wedge trial. BMC Musculoskelet Disord 2020 Apr 13;21(1):236 [FREE Full text] [doi: 10.1186/s12891-020-03235-z] [Medline: 32284049]
51. Brosseau L, Taki J, Desjardins B, Thevenot O, Fransen M, Wells GA, et al. The Ottawa panel clinical practice guidelines for the management of knee osteoarthritis. Part two: strengthening exercise programs. Clin Rehabil 2017 May;31(5):596-611. [doi: 10.1177/0269215517691084] [Medline: 28183213]
52. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 7;18(1):e5 [FREE Full text] [doi: 10.2196/jmir.5235] [Medline: 26742999]
53. Chen H, Chuang T, Lin P, Lin Y, Chuang Y. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
54. Bollen JC, Dean SG, Siegert RJ, Howe TE, Goodwin VA. A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercise programmes, and their psychometric properties. BMJ Open 2014 Jun 27;4(6):e005044 [FREE Full text] [doi: 10.1136/bmjopen-2014-005044] [Medline: 24972606]
55. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med 2010 Aug;26(3):355-369 [FREE Full text] [doi: 10.1016/j.cger.2010.03.001] [Medline: 20699159]
56. Eyles JP, Hunter DJ, Bennell KL, Dziedzic KS, Hinman RS, van der Esch M, Joint Effort Initiative Members. Priorities for the effective implementation of osteoarthritis management programs: an OARSI international consensus exercise. Osteoarthritis Cartilage 2019 Sep;27(9):1270-1279 [FREE Full text] [doi: 10.1016/j.joca.2019.05.015] [Medline: 31163271]
57. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]

Abbreviations

AQoL: Assessment of Quality of Life instrument
EARS: Exercise Adherence Rating Scale
ICC: intraclass correlation coefficient
NRS: numeric rating scale
OA: osteoarthritis
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 23.06.20; peer-reviewed by L Bearne, L Brosseau; comments to author 04.08.20; revised version received 10.08.20; accepted 17.08.20; published 28.09.20.

Please cite as:
Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJC, Hinman RS
Behavior Change Text Messages for Home Exercise Adherence in Knee Osteoarthritis: Randomized Trial
J Med Internet Res 2020;22(9):e21749
URL: http://www.jmir.org/2020/9/e21749/
doi:10.2196/21749
PMID:32985994

Original Paper

Promoting Safe Injection Practices, Substance Use Reduction, Hepatitis C Testing, and Overdose Prevention Among Syringe Service Program Clients Using a Computer-Tailored Intervention: Pilot Randomized Controlled Trial

Karli R Hochstatter[1], MPH, PhD; Shawnika J Hull[2], PhD; Ajay K Sethi[3], MHS, PhD; Marguerite E Burns[3], PhD; Marlon P Mundt[4], MA, MS, PhD; Ryan P Westergaard[3,5], MPH, PhD, MD

[1]Columbia University School of Social Work, New York, NY, United States

[2]Department of Prevention and Community Health, Milken Institute School of Public Health, The George Washington University, Washington DC, DC, United States

[3]Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

[4]Department of Family Medicine and Community Health, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

[5]Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

Corresponding Author:
Karli R Hochstatter, MPH, PhD
Columbia University School of Social Work
1255 Amsterdam Ave
New York, NY, 10027
United States
Phone: 1 2128512300
Email: krh2151@columbia.edu

Abstract

Background: Syringe service programs (SSPs) are safe, highly effective programs for promoting health among people who inject drugs. However, resource limitations prevent the delivery of a full package of prevention services to many clients in need. Computer-tailored interventions may represent a promising approach for providing prevention information to people who inject drugs in resource-constrained settings.

Objective: The aim of this paper is to assess the effect of a computer-tailored behavioral intervention, called Hep-Net, on safe injection practices, substance use reduction, overdose prevention, and hepatitis C virus (HCV) testing among SSP clients.

Methods: Using a social network–based recruitment strategy, we recruited clients of an established SSP in Wisconsin and peers from their social networks. Participants completed a computerized baseline survey and were then randomly assigned to receive the Hep-Net intervention. Components of the intervention included an overall risk synthesis, participants' selection of a behavioral goal, and an individualized risk reduction exercise. Individuals were followed up 3 months later to assess their behavior change. The effect of Hep-Net on receiving an HCV screening test, undergoing Narcan training, reducing the frequency of drug use, and sharing drug equipment was assessed. The individual's readiness to change each behavior was also examined.

Results: From 2014 to 2015, a total of 235 people who injected drugs enrolled into the Hep-Net study. Of these, 64.3% (151/235) completed the follow-up survey 3-6 months postenrollment. Compared with the control group, individuals who received the Hep-Net intervention were more likely to undergo HCV testing (odds ratio [OR] 2.23, 95% CI 1.05-4.74; *P*=.04) and receive Narcan training (OR 2.25, 95% CI 0.83-6.06; *P*=.11), and they shared drug equipment less frequently (OR 0.06, 95% CI 0.55-0.65; *P*<.001). Similarly, individuals who received the intervention were more likely to advance in their stage of readiness to change these 3 behaviors. However, intervention participants did not appear to reduce the frequency of drug use or increase their readiness to reduce drug use more than control participants, despite the fact that the majority of the intervention participants selected this as the primary goal to focus on after participation in the baseline survey.

Conclusions: Implementing computer-based risk reduction interventions in SSPs may reduce harms associated with the sharing of injection equipment and prevent overdose deaths; however, brief computerized interventions may not be robust enough to overcome the challenges associated with reducing and ceasing drug use when implemented in settings centered on the delivery of prevention services.

Trial Registration: ClinicalTrials.gov NCT02474043; https://clinicaltrials.gov/ct2/show/NCT02474043
International Registered Report Identifier (IRRID): RR2-10.2196/resprot.4830

(J Med Internet Res 2020;22(9):e19703) doi:10.2196/19703

KEYWORDS

hepatitis C virus; intravenous drug abuse; drug overdose; harm reduction; web-based intervention

Introduction

Background

The concurrent and overlapping epidemics of substance abuse, overdose, and hepatitis C virus (HCV) constitute a public health crisis. In the United States, an estimated 1.7 million people are affected by substance use disorders related to prescription opioids and 652,000 people are affected by heroin use disorders [1]. Moreover, approximately 70,237 Americans died from drug overdoses involving prescription opioids or illicit substances in 2017 alone, a four-fold increase since 1999 [2]. Through the sharing of needles and preparation equipment, injection drug use has fueled the spread of infectious diseases and now serves as the primary risk factor for HCV transmission in the United States [3]. From 2004 to 2014, acute HCV increased by 400% among Americans aged 18 to 29 years [4]. During this same period, there was an 817% increase in treatment admissions for injection of prescription opioids and a 603% increase in admissions for heroin injection [4].

These sharp increases in injection drug use and HCV incidence are concentrated in communities with traditionally poor access to prevention services, affecting rural and suburban residents in particular [5,6]. An epidemiologic investigation conducted to understand the factors associated with the tripling of HCV incidence between the periods of 2004 and 2008 and 2009 and 2010 in 6 contiguous rural counties of Wisconsin found that 94% of infected patients had shared hypodermic needles, drug preparation equipment, or drug snorting equipment [7]. Similar outbreaks of HCV among young persons who inject drugs have occurred across many other communities in the United States, including several Appalachian states, Massachusetts, Indiana, and New York State [6,8-11], and several other jurisdictions have been identified as potentially vulnerable communities to similar outbreaks [12].

Syringe service programs (SSPs) are safe, highly effective programs for promoting health among people who inject drugs. By providing access to sterile needles and other drug paraphernalia, these community-based programs have been shown to reduce the transmission of HCV and HIV among drug-using networks [13-18], without increasing the frequency of drug use [19-21]. In addition to distributing comprehensive prevention materials, SSPs often offer a range of services, including referral to substance use disorder treatment programs, education on safe injection practices, overdose prevention, and counseling and testing for HCV, HIV, and other sexually transmitted diseases. Despite the abundant documented benefits of SSPs, resource limitations prevent the delivery of a full package of prevention services to many in need [22-24].

Computer-tailored interventions (CTIs), in which the output of a persuasive technological system is adapted to the individual, are a low-cost strategy for delivering personalized health information that is specific to the unique psychosocial needs of each individual. Persuasive health technology is a growing field that combines medicine, computer science, and psychology to aid and motivate people to adopt behaviors that benefit them while avoiding the harmful ones and is used for both health promotion, prevention and disease management [25-27]. CTIs have become increasingly common for facilitating a wide range of healthy behaviors, including smoking cessation, diet and exercise, and mammography screening [28]. However, the use of CTIs to promote healthy behaviors among people who inject drugs, particularly in prevention-oriented and other community-based settings such as SSPs, remains to be less studied. CTIs may represent a promising approach for providing a personalized risk reduction plan for people who inject drugs in resource-constrained settings.

Objectives

Recognizing the growing demand for targeted, evidence-based interventions and the existing infrastructure of SSPs, a computer-tailored risk reduction intervention, called Hep-Net, was developed and implemented through a pilot randomized controlled trial (RCT) [29]. The objectives of this study are to evaluate whether incorporating a computerized behavioral intervention into existing prevention services at SSPs can lead to the adoption of healthier behaviors, including safer injection practices, substance use reduction, overdose prevention, and HCV testing.

Methods

Study Population and Settings

From 2014 to 2015, a total of 235 people who inject drugs enrolled in the Hep-Net study. Participants were either clients of an established SSP operating in 2 Wisconsin cities, Madison and Milwaukee, or peers recruited from the social networks of these clients. Eligibility criteria included being 18 years of age or older, having injected drugs in the past 30 days, and willingness to provide contact information for the 3-month follow-up.

To conduct outreach among high-risk populations of people who inject drugs and engage even those who do not regularly use prevention services, we used social network–based referrals to recruit the study sample. SSP clients were informed about the study during a routine needle-exchange encounter, invited to participate, screened for eligibility, and, if eligible, provided with the computerized baseline survey. After completing the baseline survey, participants were randomly assigned to receive

the intervention, in which the computer immediately delivered the intervention content, or the control, in which the computer session was terminated after completion of the baseline survey. Upon completion of the baseline visit, study participants in both arms received referral coupons and were encouraged to refer eligible peers. Using the contact information provided at baseline, individuals were contacted 3 months later to complete the follow-up assessment. A protocol that describes in detail recruitment methods and intervention content and provides survey instruments and example content images of the Hep-Net intervention has been previously published [29].

Risk Reduction Intervention

The Hep-Net intervention aims to increase health-promoting behaviors and reduce risky drug use behaviors among people who inject drugs. Hep-Net targets 4 different behavioral domains: (1) undergoing regular HCV screening, (2) taking steps to prevent opioid overdose, (3) using clean works for every injection, and (4) reducing and ultimately ceasing injection drug use. Components of the intervention included an overall risk synthesis, participants' selection of a behavioral goal (from 1 to 4) and an individualized risk reduction exercise [29]. The risk synthesis had 2 components: first, positively framed feedback messages were delivered, highlighting healthy behaviors reported by participants that can reduce the risk of HCV transmission and opioid overdose, and second, tailored feedback was delivered regarding specific behaviors associated with increased risk reported by the participant. After selecting a behavioral goal, the participants were delivered a series of screens with feedback and educational content tailored to the individual's reported risk behaviors and their stage of readiness to change. Finally, during the interactive risk reduction exercise, the participants were asked to select 3 to 5 action steps from a list of 10 to 12 suggested steps that they were willing to do in the following 3 months.

Action steps considered during the individualized risk reduction exercise were tailored to the participant's readiness to change (as assessed at pretest) and are informed by the transtheoretical model (TTM), which conceptualizes behavior change as a continuum and argues that different processes of change are relevant to different stages of readiness [30]. According to the TTM, individuals who are not yet considering behavior change are in the precontemplation stage. In the contemplation stage, individuals may consider change but have not yet taken steps toward behavioral change. Contemplation is followed by preparation, action, and maintenance [31]. The Hep-Net intervention delivered feedback and risk reduction activities that were informed by the individual's stage of readiness, as described in the RCT's published protocol [29].

Data Collection

The baseline survey and the Hep-Net intervention were delivered from August 6, 2014, to April 16, 2015. Baseline information included demographics, drugs of choice, the duration of injection drug use, incarceration history, HCV status, the frequency of sharing needles, syringes and other works, naloxone (Narcan) training status, the frequency of drug use, and readiness to change each of the 4 targeted behaviors. Three-month follow-up surveys were conducted until August 18, 2015, and the same behaviors were assessed as the baseline questionnaire. All surveys were delivered on laptops and designed to last 20 to 30 min.

In addition to the Hep-Net survey data, HCV testing data were collected and stored at SSP sites through the agency's standard protocol as well as in the Wisconsin Electronic Disease Surveillance System (WEDSS) in the Department of Health Services for those who tested HCV positive. These 2 additional data sources were used to track HCV testing for those who did and did not complete the follow-up survey, allowing us to measure the HCV testing uptake with greater validity. To assess HCV follow-up testing within 6 months of enrollment for all individuals enrolled in the study, we received participant-level data stored in the WEDSS through a secure, encrypted, and Health Insurance Portability and Accountability Act (HIPAA)-compliant data transfer protocol. Furthermore, archived testing records stored at the 2 SSP sites were manually searched to identify individuals who tested HCV negative. Study participants were matched to the WEDSS and SSP data by first and last name and date of birth. The variables collected from these sources include the date of HCV tests and test results.

Main Outcomes

We assessed 4 behavioral goals:

- *Receive HCV follow-up testing*: Individuals were considered tested for HCV at follow-up if they were not already aware of being HCV positive at the time of Hep-Net enrollment and if any of the following conditions were true: (1) self-reported at follow-up being tested for HCV since their first study visit, (2) underwent HCV testing at the time of the follow-up survey, (3) an archived record of an HCV test existed 6 months postenrollment at the SSP, or (4) any HCV test was reported to the WEDSS 6 months postenrollment.
- *Get trained to administer Narcan*: Individuals were considered trained to administer Narcan if they self-reported undergoing training for naloxone, or Narcan, in the follow-up survey. Those who were already trained at baseline were excluded from this analysis under the assumption that the knowledge and skills gained from such training last indefinitely.
- *Reduce frequency of sharing drug equipment*: To measure the frequency of sharing needles, syringes or other works, individuals were asked how often they shared (1) needles, (2) cottons or filters, and (3) cookers with another person when they used drugs in the past month. Responses were recorded using a slider-bar feature and ranged from 0% of the time they used drugs to 100%. For this analysis, the frequency of sharing drug equipment corresponded to the highest percentage reported among these 3 pieces of equipment. When response data for one or more of these 3 questions were missing, it was assumed that the data were missing at random. These observations were excluded because we could not make an inference about why these data would be missing or hypothesize what these data would have been.
- *Reduce the frequency of drug use*: To measure the frequency of drug use, individuals were asked how often they used

heroin; oxycodone, or OxyContin; methamphetamine (crystal or meth); and cocaine or crack in the past month. Four response categories were available, ranging from never to every day. For this analysis, the frequency of drug use corresponded to the highest frequency of use reported among these drug types and was collapsed into 3 categories: (1) never or less than once a week, (2) more than once a week but not every day, or (3) every day.

Understanding that behavior change often occurs through a series of stages as opposed to a single discrete event and that the Hep-Net intervention may not be powerful enough for some individuals to achieve full behavior change over the course of 3 months, we also assessed individuals' *readiness* to engage in these 4 health-promoting behaviors. At baseline and follow-up, individuals were presented with each of the 4 behavioral goals and asked to select their readiness to change each behavior. The 5 answer choices ranged from *I am not even thinking about this goal* to *I have reached this goal*, which corresponds with the 5 stages outlined by the transtheoretical model: precontemplation, contemplation, preparation, action, and maintenance. Owing to the low number of individuals observed in the contemplation and precontemplation stages, these stages have been combined into a single stage for this analysis.

Statistical Analysis

Descriptive statistics were calculated for the baseline variables of the entire study population. Characteristics of the intervention group were compared with those of the control group using Pearson chi-square (if expected cell frequencies are >5) or Fisher exact (if expected cell frequencies are ≤5) tests for testing differences in categorical variables and the two-sample unpaired *t* test (if normally distributed) or Wilcoxon rank-sum test (if not normally distributed) for continuous variables. Baseline characteristics were also examined to determine whether there were any characteristics that differed between those who returned for the follow-up survey and those who were lost, using the same statistical tests.

As the Hep-Net risk reduction intervention presented a menu of behavioral goals, providing the opportunity for behavior change beyond the behavior that was selected, analyses compared all intervention participants (as opposed to only those who selected the particular goal) with control participants. Binomial logistic regression was used to assess differences between intervention and control arms for binary outcomes: receiving HCV follow-up testing (yes or no) and undergoing Narcan training (yes or no). As the frequency of sharing drug equipment is a non-normally distributed outcome, we conducted Poisson regression to analyze the difference between the intervention and control arms in sharing equipment. Ordinal logistic regression was used to assess the difference in the frequency of drug use between the intervention and control arms. Ordinal logistic regression was also used to assess differences in intervention and control participants' readiness to change each behavior. All analyses were conducted using SAS version 9.4 (SAS Institute). Statistical significance was determined at $\alpha \leq .05$, two-sided.

Results

Demographics

During the enrollment period, 235 people who injected drugs provided consent and completed the baseline survey. Of these, 109 people were randomly assigned to the intervention group and 126 were assigned to the control arm. The baseline descriptive characteristics are displayed in Table 1. Among the 235 individuals, 180 (76.6%) were male, 138 (58.7%) were White, 211 (89.8%) had at least a high school diploma or GED, 161 (68.5%) were unemployed at the time of enrollment, and 172 (73.2%) had an annual income of less than US $11,500. A total of 44.7% (105/235) individuals used illicit substances every day at baseline. Heroin was the most common substance used in the past 30 days, with 87.2% (205/235) individuals using heroin, followed by cocaine or crack (147/235, 62.6%), oxycodone or OxyContin (103/235, 43.8%), and methamphetamine (19/235, 8.1%). At baseline, 71.1% (167/235) individuals self-reported ever being tested for HCV before Hep-Net study enrollment, of which 19.2% (32/235) reported testing HCV positive.

Of the 235 individuals who completed the baseline survey, 151 (64.3%) completed the follow-up survey 3 to 6 months postenrollment. Individuals who were lost to follow-up were slightly less likely to have health insurance (81.0% vs 90.0%; P=.05) and more likely to be homeless (65.1% vs 47.0%; P=.008). This finding is expected, considering that homeless populations are often transient [32], which may lead to high study attrition and are less likely to have health insurance [33].

Table 1. Baseline demographic characteristics by intervention and control arm (N=235).

Characteristics	Study arm[a]	
	Control (n=126)	Intervention (n=109)
Gender, n (%)		
Male	92 (73.0)	88 (80.7)
Female	34 (27.0)	21 (19.3)
Race, n (%)		
White	79 (62.7)	59 (54.1)
Black or African American	32 (25.4)	34 (31.2)
Other or multiple	15 (11.9)	16 (14.7)
Age (years), median (IQR)	35 (18)	33 (17)
Education level, n (%)		
Less than high school	17 (13.5)	7 (6.4)
GED[b], HSED[c], or high school diploma	61 (48.4)	50 (45.9)
At least some college	48 (38.1)	52 (47.7)
Health insurance, n (%)		
No	20 (16.0)	11 (10.1)
Yes	105 (84.0)	98 (89.9)
Completed follow-up survey, n (%)		
No	49 (38.9)	35 (32.1)
Yes	77 (61.1)	74 (67.9)
Currently employed, n (%)		
No	92 (73.0)	69 (63.9)
Yes	34 (27.0)	39 (36.1)
Children at home, n (%)		
No	100 (79.4)	88 (80.7)
Yes	25 (19.8)	20 (18.4)
Income, US $; n (%)		
None	31 (24.6)	30 (27.5)
Less than 11,500	62 (49.2)	49 (45.0)
More than 11,500	28 (22.2)	30 (27.5)
Homeless in the past year, n (%)		
No	58 (46.0)	51 (46.8)
Yes	68 (54.0)	57 (52.3)
Incarcerated in the past year, n (%)		
No	82 (66.1)	63 (58.9)
Yes	42 (33.9)	44 (41.1)
Substance used in past 30 days, n (%)		
Heroin[d]	114 (92.7)	91 (84.3)
OxyContin or Oxycodone[d]	62 (50.4)	41 (37.6)
Methamphetamine	11 (8.7)	8 (7.3)
Cocaine or crack	79 (62.7)	68 (62.4)
Ever overdosed, n (%)		

Characteristics	Study arm[a]	
	Control (n=126)	Intervention (n=109)
No	75 (60.5)	70 (64.2)
Yes	49 (39.5)	39 (35.8)
Percentage of time they share drug equipment[e], median (IQR)	24.0 (47)	24.5 (48)
Has shared needles, cottons, filters, or cookers in the past 30 days, n (%)	99 (88.4)	87 (87.9)
Has shared needles in the past 30 days, n (%)	78 (70.9)	67 (69.1)
Has shared cottons or filters in the past 30 days, n (%)	76 (71.0)	62 (63.3)
Has shared cookers in the past 30 days, n (%)	86 (81.9)	72 (77.4)
Frequency of drug use in past 30 days, n (%)		
Less than daily	67 (53.2)	63 (57.8)
Every day	59 (46.8)	46 (42.2)
Years of injection drug use, median (IQR)	5 (13)	7 (13)
Tested for HCV[f] before study enrollment, n (%)		
No	36 (28.6)	21 (19.3)
Yes	85 (67.5)	82 (75.2)
Unsure	5 (4.0)	6 (5.5)
Result of most recent HCV test, n (%)		
Negative	67 (80.7)	62 (79.5)
Positive	16 (19.3)	16 (20.5)
Trained to administer Narcan at baseline, n (%)		
No	77 (61.1)	62 (56.9)
Yes	49 (38.9)	47 (43.1)

[a]Column percentages may not add up to 100 because the data are assumed to be missing at random.

[b]GED: general educational development.

[c]HSED: High School Equivalency Diploma.

[d]Statistically significant at $\alpha<.05$ (unadjusted).

[e]Includes needles, cottons or filters, and cookers.

[f]HCV: hepatitis C virus.

Randomization Checks

There were no significant differences in demographic variables (gender, race, age, education level, health insurance, employment, income, homelessness, incarceration history, or overdose history) between the intervention and control arms. Those randomized to the control group were significantly more likely to use heroin (92.7% vs 84.3%; *P*=.04) and OxyContin or Oxycodone (50.4% vs 37.6%; *P*=.05) in 30 days before baseline. As the use of these substances may affect the behavioral outcomes assessed in this study, we adjusted for heroin and OxyContin or Oxycodone use in all regression models. There were no significant differences at baseline between the intervention and control participants in the 4 behaviors that Hep-Net targets: the proportion screened for HCV or trained to administer Narcan, nor the frequency of sharing drug equipment or using drugs.

Main Outcomes

Behavioral Goal 1: Receive HCV Follow-Up Testing

Among the 235 study participants, 81 (34.5%) individuals agreed to undergo a rapid HCV test at the time of study enrollment; none of the 32 individuals who self-reported already knowing to be HCV positive agreed to an HCV test at enrollment. Of the 81 individuals, 14 (17%) tested reactive, and all 14 had a blood specimen collected and sent for confirmatory testing.

Of the 203 individuals who did not report testing positive for HCV before study enrollment, 38 (18.7%) received HCV follow-up testing within 6 months of enrollment (Table 2). Individuals in the intervention arm were significantly more likely to undergo HCV follow-up testing than those in the control arm (23.7% vs 14.6%; OR 2.23, 95% CI 1.05-4.74; *P*=.04). This trend persisted at 12 months postenrollment, where 26.9% (25/93) of individuals in the intervention arm and 17.3% (19/110) in the control arm received HCV follow-up testing.

Table 2. Proportion who received Hepatitis C virus testing and underwent Narcan training, and the frequency of sharing drug equipment and using drugs at follow-up.

Behavior	Control	Intervention	OR (95% CI)[a]
Received HCV[b] follow-up testing within 6 months (n=203), n (%)			2.23 (1.05-4.74)[c]
Yes	16 (14.5)	22 (24)	
No	94 (85.5)	71 (76)	
Trained to administer Narcan (n=91), n (%)			2.25 (0.83-6.06)
Yes	10 (21)	16 (37)	
No	38 (79)	27 (63)	
Frequency of sharing drug equipment (n=116)			0.60 (0.55-0.65)[c]
Median percentage (IQR)	5 (1-47)	2 (1-24)	
Frequency of drug use (n=150), n (%)			0.90 (0.49-1.65)
Once a week or less	25 (33)	22 (30)	
More than once a week but not every day	30 (39)	28 (38)	
Every day	21 (28)	24 (32)	

[a]Adjusted for heroin and OxyContin or Oxycodone use at baseline.

[b]HCV: Hepatitis C virus.

[c]Statistical significance at α<.05.

Behavioral Goal 2: Gets Trained to Administer Narcan

Among the 235 study participants, 96 (40.9%) had already been trained to administer Narcan at baseline. Of the 151 individuals who participated in the follow-up survey, 60 were excluded from this analysis because they had been trained before Hep-Net enrollment (n=59) or chose not to answer (n=1). Of the remaining 91 individuals, 26 (28.6%) received training between baseline and follow-up (Table 2). Individuals in the intervention arm were more likely to receive Narcan training than those in the control arm (37.2% vs 20.8%; OR 2.25, 95% CI 0.83-6.06; *P*=.11).

Behavioral Goal 3: Reduce Frequency of Sharing Drug Equipment

At baseline, 186/235 (79.1%) reported sharing needles, cottons, filters, or cookers with another person in the 30 days before study enrollment: 145/235 (61.7%) individuals reported sharing injection needles or syringes, 138/235 (58.7%) shared cottons or filters, and 158/235 (67.2%) shared cookers. Thus, the majority of participants shared more than one piece of equipment (98 individuals shared all the 3 pieces, 59 individuals shared 2 pieces, and 29 individuals shared 1 piece). A total of 25/235 (10.6%) individuals reported never sharing any of these 3 pieces of equipment. The median percentage of time participants shared equipment at baseline was 24.0% (IQR 2.0-49.0).

At follow-up, the median percentage of time participants shared equipment dropped to 3.0% overall (IQR 1.0-28.0) among the 116 individuals who provided responses for all 3 pieces of equipment (Table 2). Individuals in the intervention group were significantly less likely to share drug equipment than individuals in the control group (control median, IQR 5, 1-47; intervention median, IQR 2, 1-24). Those in the intervention group shared drug equipment 0.60 times less than those in the control group (OR 0.60, 95% CI 0.55-0.65; *P*<.001).

Behavioral Goal 4: Reduce Frequency of Drug Use

Overall, 44.7% (105/235) of individuals reported using drugs every day at baseline, compared with 30.0% (45/150) at follow-up. This reduction in the proportion of individuals using drugs every day was experienced by both intervention and control participants (Table 2), where 32.4% and 27.6% reported using every day, respectively (OR 0.90, 95% CI 0.49-1.65; *P*=.72).

Readiness to Change

Table 3 shows the distribution of stages of readiness for change at baseline and at follow-up. This table also shows the odds ratios demonstrating the degree to which individuals in the intervention group advanced their stage of readiness in the direction of *maintenance* compared with the control group at follow-up.

Table 3. Proportion of individuals in each stage of readiness for risk reduction behaviors at baseline and follow-up and odds ratios comparing the mean values of readiness to change between intervention and control groups at follow-up.

Stage of readiness	HCV[a] testing intentions	Narcan training intentions	Intentions to use clean works	Intentions to reduce drug use
Baseline (n=235) , n (%)				
Precontemplation and contemplation				
Control group[b]	25 (19.8)	34 (27.0)	15 (11.9)	25 (19.8)
Intervention group[c]	19 (17.4)	24 (22.0)	9 (8.3)	19 (17.4)
Total	44 (18.7)	58 (24.7)	24 (10.2)	44 (18.7)
Preparation				
Control group	31 (24.6)	28 (22.2)	13 (10.3)	32 (25.4)
Intervention group	19 (17.4)	23 (21.1)	17 (15.6)	31 (28.4)
Total	50 (21.3)	51 (21.7)	30 (12.8)	63 (26.8)
Action				
Control group	46 (36.5)	23 (18.3)	55 (43.7)	64 (50.8)
Intervention group	45 (41.3)	20 (18.4)	44 (40.4)	54 (49.5)
Total	91 (38.8)	43 (18.3)	99 (42.1)	118 (50.2)
Maintenance				
Control group	24 (19.1)	41 (32.5)	43 (34.1)	5 (4.0)
Intervention group	26 (23.9)	42 (38.5)	39 (35.8)	5 (4.6)
Total	50 (21.3)	83 (35.3)	82 (34.9)	10 (4.3)
Follow-up (n=151) , n (%)				
Precontemplation and contemplation				
Control group[d]	12 (16)	17 (22)	4 (5)	7 (9.)
Intervention group[e]	9 (12)	11 (15)	3 (4)	9 (12)
Total	21 (13.9)	28 (18.5)	7 (4.6)	16 (10.6)
Preparation				
Control group	4 (5)	10 (13)	3 (4)	11 (14)
Intervention group	9 (12)	13 (18)	2 (3)	15 (20)
Total	13 (8.6)	23 (15.2)	5 (3.3)	26 (17.2)
Action				
Control group	33 (43)	16 (21)	34 (44)	45 (58)
Intervention group	24 (32)	13 (18)	23 (31)	41 (55)
Total	57 (37.7)	29 (19.2)	57 (37.7)	86 (57.0)
Maintenance				
Control group	28 (36)	34 (44)	36 (47)	14 (18)
Intervention group	32 (43)	37 (50)	46 (62)	9 (12)
Total	60 (39.7)	71 (47.0)	82 (54.3)	23 (15.2)
OR (95% CI)[f,g]	1.23 (0.67-2.24)	1.28 (0.70-2.34)	1.92 (0.99-3.71)	0.67 (0.36-1.27)

[a]HCV: hepatitis C virus.

[b]n=126.

[c]n=109.

[d]n=77.

[e]n=74

[f]The precontemplation and contemplation stages were combined into a single stage, resulting in a four-category dependent variable (stage of readiness)

for each behavior. The probability of reaching the goal (maintenance) was modeled, with control participants being the reference group.

[g]Adjusted for heroin and OxyContin or Oxycodone use at baseline.

Behavioral Goal 1: Receive HCV Follow-Up Testing

Overall, the majority of individuals were in the preparation 21.3% (50/235), action 38.7% (91/235), or maintenance 21.3% (50/235) stage of readiness to undergo regular HCV screening. Of the 109 individuals in the intervention group, 3 (2.8%) selected this goal to work on specifically over the next 3 months. At follow-up, a higher proportion of individuals reached the maintenance stage in both groups (60/151, 39.7% overall), with a slightly higher proportion in the intervention group (32/74, 43.2%) than in the control group (28/77, 36.4%). The odds of advancing through the stages of readiness toward the *maintenance* stage were 16% higher in the intervention arm than in the control arm (OR 1.23, 95% CI 0.67-2.24; *P*=.51).

Behavioral Goal 2: Get Trained to Administer Narcan

At baseline, 21.7% (51/235) of individuals were in the preparation stage of readiness to undergo Narcan training, 18.3% (43/235) were in the action stage, and 35.3% (83/235) were in the maintenance stage. Of the 109 individuals who received the intervention, 10 (9.2%) chose this goal to work on specifically over the next 3 months. At follow-up, a higher proportion of individuals had reached the maintenance stage overall (71/151, 47.0%), with a slightly higher proportion in the intervention group (37/74, 50.0%) than in the control group (34/77, 44.2%). The odds of advancing through the stages of readiness toward the *maintenance* stage were 28% higher in the intervention arm than in the control arm (OR 1.28, 95% CI 0.70-2.34; *P*=.43).

Behavioral Goal 3: Reduce Frequency of Sharing Drug Equipment

At baseline, the highest proportion of individuals were in the action stage (99/235, 42.1%) and the maintenance stage (83/235, 34.9%) for using clean works every time they injected. Among the 109 intervention participants, 17 (15.6%) selected this goal to work on over the next 3 months. The proportion of individuals reaching the maintenance stage at follow-up increased (54.3% overall) among both groups but was higher among intervention participants (62.2%) than among control participants (59/126, 46.8%). The odds of advancing through the stages of readiness toward the *maintenance* stage were 80% higher in the intervention arm than in the control arm (OR 1.92, 95% CI 0.99-3.71; *P*=.05).

Behavioral Goal 4: Reduce Frequency of Drug Use

At baseline, the highest proportion of people were in the preparation (63/235, 26.8%) and action (118/235, 50.2%) stages of readiness for the goal of reducing or ceasing drug use. In contrast to the other 3 behavioral goals, the least proportion of individuals were in the maintenance stage (10/235, 4.3% overall). The highest proportion of individuals in the intervention group (72.5%) selected this goal to focus on over the next 3 months. Although the proportion of individuals reaching the maintenance stage at follow-up increased overall to 15.2% (23/151), individuals in the intervention group were less likely to advance along the readiness continuum in the direction of *maintenance* than those in the control group (OR 0.67, 95% CI 0.36-1.27; *P*=.22).

Discussion

Principal Findings

The goal of this pilot RCT was to determine whether implementing a computerized risk reduction intervention into existing prevention services at SSPs can improve health-promoting behaviors among people who inject drugs. We found that individuals who received the Hep-Net intervention were significantly more likely to undergo testing for HCV and less likely to share drug equipment. Although not statistically significant, individuals in the intervention arm were also more likely to undergo the Narcan administration training. There was also a trend toward increased readiness to change among intervention participants for these 3 behaviors. However, intervention participants did not appear to reduce the frequency of drug use or increase readiness to reduce drug use more than control participants, despite the fact that the majority of intervention participants selected this as the primary goal to focus on after participation in the baseline survey. These results demonstrate that implementing computer-based risk reduction interventions in SSPs may reduce harms related to sharing of injection equipment and prevent overdose deaths.

No significant difference in reducing the frequency of drug use was observed between the intervention and control participants. Multiple interpretations of this finding are plausible. Hep-Net was a single-session, brief intervention. Reducing substance use is likely to require multipronged strategies, including provision of social, psychological, and physical support. Although Hep-Net in isolation did not affect significant changes in this behavior, it may be useful if integrated into other existing programs to more holistically address the complexity of addiction. Furthermore, almost half of the participants did not return for follow-up. It may be that those who did not return for follow-up disproportionately included individuals who were successful in reducing their substance use and therefore did not return to the needle exchange. Increasing the dosage and comprehensiveness of the intervention may also bolster the effect and allow individuals to reduce drug use and reach their recovery goals. For example, implementing booster sessions that expose clients to additional intervention content may strengthen the effect it has on reducing the frequency of drug use. Understanding how interventions may be enhanced to effectively reduce drug use and improve long-term recovery success among SSP clients is greatly needed.

Although CTIs have become increasingly common for facilitating a wide range of health-promoting behaviors, very few studies have examined the effect of a single CTI on multiple risk behaviors among people who inject drugs in SSP or other community-based settings. With regard to CTIs addressing substance abuse, previous studies have instead primarily focused on HIV/AIDS prevention [34,35], perinatal drug use [36,37], or alcohol use disorders [38,39] or were coupled with

therapist-delivered treatments [40,41]. To our knowledge, only one randomized trial has evaluated the efficacy of a CTI for the adoption of safer injection practices among SSP clients. This study found that IV drug users visiting SSPs who received the CTI were using dirty syringes 0.47 times in less than 1 month after the intervention began compared with those who did not receive the CTI; however, the positive effect was short-lived [42]. This study is the first to implement a single CTI into SSPs to address safe injection practices, substance use reduction, hepatitis C testing, and overdose prevention simultaneously. Additional research on CTIs for SSP clients is needed because of its potential to address the challenges of resource limitations facing many affected communities.

The transtheoretical model has been used in the development of various substance abuse interventions and has demonstrated success because it accounts for individuals who are not ready to make behavior change or who do not see their behavior as problematic [30]. However, there is limited research on applying the transtheoretical model to promote healthy behaviors specifically among people who inject drugs, who are engaged in prevention services. Prior studies using transtheoretical model-designed interventions to promote behavior change among people who inject drugs have been restricted to increasing condom use and using bleach to clean drug paraphernalia [43-45]. In this study, we found that most people were in the preparation, action, or maintenance stage for all behaviors at baseline. This heightened readiness may have partially limited our ability to detect the statistical significance in this study.

Limitations

We demonstrate the effects of a brief, digitally delivered tailored intervention designed to reduce HCV risk behaviors among hard-to-reach individuals who inject drugs. Although our study has much to offer, it is not without limitations. Although delivering the intervention through a well-established SSP that holds strong relationships with many members of the targeted population was a major strength of this study, communities that lack such a developed SSP may fail to reach less-engaged clientele. Generalizability may also be limited because individuals utilizing SSPs may choose to prioritize their health and value safe injection practices more strongly than those not engaged in prevention services, as suggested by the high proportion of individuals in the preparation, action, and maintenance stages of behavior change. For this reason, we implemented a social network–based recruitment strategy to engage difficult-to-reach individuals.

Assessments of drug use behaviors were self-reported measures. However, we anticipate that the computerized approach to data collection maximizes privacy and, in turn, mitigates social desirability bias.

Finally, the primary aim of this pilot study was to assess the feasibility of the social network recruitment approach [29]. As such, the sample size may not be sufficiently powered to detect small effect sizes, particularly for subgroup analyses. Further analysis with improved power is needed to improve our understanding of the effectiveness of computer-tailored risk reduction interventions implemented in SSP settings.

Conclusions

As the opioid epidemic continues to burden rural communities, expanding the delivery of comprehensive prevention services in resource-poor settings is critically important. To be most effective, prevention services should be patient-centered and provide individual, personalized plans for risk reduction that are responsive to their unique preferences, needs, and motivations. Using fast data-driven decision guidelines, CTIs offer the opportunity to deliver individualized care even in settings facing resource limitations. Implementing CTIs in SSP settings is one approach for engaging highly marginalized and underserved populations of people who inject drugs and facilitate the uptake of safe injection practices, decrease opioid overdose deaths, and reduce the transmission of blood-borne infections. Although such interventions may prevent some devastating consequences of injection drug use, additional efforts are needed to help individuals reduce drug use and overcome the power of addiction.

Acknowledgments

The authors would specifically like to thank the SSP clients who participated in this study as well as the SSP staff, whose established relationships with the SSP clients allowed for the successful Hep-Net implementation and follow-up. This project was supported by the Clinical and Translational Science Award program through the National Institutes of Health-National Center for Advancing Translational Sciences, grants UL1TR000427 and UL1TR002373, and the University of Wisconsin-Madison School of Medicine and Public Health's Wisconsin Partnership Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1629 KB - jmir_v22i9e19703_app1.pdf]

References

1. National Survey on Drug Use and Health: Detailed Tables. Substance Abuse and Mental Health Services Administration. 2018. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.pdf [accessed 2020-05-01]
2. Overdose Death Rates. National Institute on Drug Abuse (NIDA). 2018 Jan 24. URL: https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates [accessed 2020-05-01]
3. Hepatitis C Testing Recommendations. Centers for Disease Control and Prevention. 2015. URL: https://www.cdc.gov/hepatitis/hcv/index.htm [accessed 2020-05-01]
4. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, et al. Increases in acute hepatitis C virus infection related to a growing opioid epidemic and associated injection drug use, United States, 2004 to 2014. Am J Public Health 2018 Mar;108(2):175-181. [doi: 10.2105/AJPH.2017.304132] [Medline: 29267061]
5. Suryaprasad A, White JZ, Xu F, Eichler BA, Hamilton J, Patel A, et al. Emerging epidemic of hepatitis C virus infections among young nonurban persons who inject drugs in the United States, 2006-2012. Clin Infect Dis 2014 Dec 15;59(10):1411-1419. [doi: 10.1093/cid/ciu643] [Medline: 25114031]
6. Zibbell JE, Iqbal K, Patel RC, Suryaprasad A, Sanders KJ, Moore-Moravian L, Centers for Disease ControlPrevention (CDC). Increases in hepatitis C virus infection related to injection drug use among persons aged ≤30 years - Kentucky, Tennessee, Virginia, and West Virginia, 2006-2012. MMWR Morb Mortal Wkly Rep 2015 May 8;64(17):453-458 [FREE Full text] [Medline: 25950251]
7. Centers for Disease Control and Prevention (CDC). Notes from the field: hepatitis C virus infections among young adults--rural Wisconsin, 2010. MMWR Morb Mortal Wkly Rep 2012 May 18;61(19):358 [FREE Full text] [Medline: 22592276]
8. Centers for Disease Control and Prevention (CDC). Hepatitis C virus infection among adolescents and young adults:Massachusetts, 2002-2009. MMWR Morb Mortal Wkly Rep 2011 May 6;60(17):537-541 [FREE Full text] [Medline: 21544042]
9. Christian WJ, Hopenhayn C, Christian A, McIntosh D, Koch A. Viral hepatitis and injection drug use in Appalachian Kentucky: a survey of rural health department clients. Public Health Rep 2010;125(1):121-128 [FREE Full text] [doi: 10.1177/003335491012500116] [Medline: 20402204]
10. Use of Enhanced Surveillance for Hepatitis C Virus Infection to Detect a Cluster Among Young Injection-Drug Users --- New York, November 2004--April 2007. Centers for Disease Control and Prevention. 2008. URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5719a3.htm [accessed 2020-05-01]
11. Peters PJ, Pontones P, Hoover KW, Patel MR, Galang RR, Shields J, Indiana HIV Outbreak Investigation Team. HIV infection linked to injection use of oxymorphone in Indiana, 2014-2015. N Engl J Med 2016 Jul 21;375(3):229-239. [doi: 10.1056/NEJMoa1515195] [Medline: 27468059]
12. van Handel MM, Rose CE, Hallisey EJ, Kolling JL, Zibbell JE, Lewis B, et al. County-level vulnerability assessment for rapid dissemination of HIV or HCV infections among persons who inject drugs, United States. J Acquir Immune Defic Syndr 2016 Nov 1;73(3):323-331 [FREE Full text] [doi: 10.1097/QAI.0000000000001098] [Medline: 27763996]
13. Hagan H, Jarlais DC, Friedman SR, Purchase D, Alter MJ. Reduced risk of hepatitis B and hepatitis C among injection drug users in the Tacoma syringe exchange program. Am J Public Health 1995 Dec;85(11):1531-1537. [doi: 10.2105/ajph.85.11.1531] [Medline: 7485666]
14. Iversen J, Wand H, Topp L, Kaldor J, Maher L. Reduction in HCV incidence among injection drug users attending needle and syringe programs in Australia: a linkage study. Am J Public Health 2013 Aug;103(8):1436-1444. [doi: 10.2105/AJPH.2012.301206] [Medline: 23763399]
15. Aspinall E, Nambiar D, Goldberg DJ, Hickman M, Weir A, van Velzen E, et al. Are needle and syringe programmes associated with a reduction in HIV transmission among people who inject drugs: a systematic review and meta-analysis. Int J Epidemiol 2014 Mar;43(1):235-248. [doi: 10.1093/ije/dyt243] [Medline: 24374889]
16. Vlahov D, Junge B. The role of needle exchange programs in HIV prevention. Public Health Rep 1998 Jun;113(Suppl 1):75-80 [FREE Full text] [Medline: 9722812]
17. Kaplan EH. A method for evaluating needle exchange programmes. Stat Med 1994;13(19-20):2179-2187. [doi: 10.1002/sim.4780131923] [Medline: 7846418]
18. des Jarlais DC, Perlis T, Arasteh K, Torian LV, Hagan H, Beatrice S, et al. Reductions in hepatitis C virus and HIV infections among injecting drug users in New York City, 1990-2001. AIDS 2005 Oct;19(Suppl 3):S20-S25. [doi: 10.1097/01.aids.0000192066.86410.8c] [Medline: 16251819]
19. Hagan H, McGough JP, Thiede H, Hopkins S, Duchin J, Alexander E. Reduced injection frequency and increased entry and retention in drug treatment associated with needle-exchange participation in Seattle drug injectors. J Subst Abuse Treat 2000 Oct;19(3):247-252. [doi: 10.1016/s0740-5472(00)00104-5] [Medline: 11027894]
20. Watters JK, Estilo MJ, Clark GL, Lorvick J. Syringe and needle exchange as HIV/AIDS prevention for injection drug users. J Am Med Assoc 1994 Jan 12;271(2):115-120. [Medline: 8264065]
21. Marx MA, Brahmbhatt H, Beilenson P, Brookmeyer RS, Strathdee SA, Alexander C, et al. Impact of needle exchange programs on adolescent perceptions about illicit drug use. AIDS Behav Vol 2001;5(4):379-386 [FREE Full text] [doi: 10.1023/A:1013195108756]

22. Syringe Exchange Programs --- United States, 2008. Centers for Disease Control and Prevention. 2010. URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5945a4.htm/Syringe-Exchange-Programs-United-States-2008 [accessed 2020-05-01]
23. des Jarlais DC. Harm reduction in the USA: the research perspective and an archive to David Purchase. Harm Reduct J 2017 Jul 26;14(1):51 [FREE Full text] [doi: 10.1186/s12954-017-0178-6] [Medline: 28747189]
24. Rudolph A, Standish K, Amesty S, Crawford N, Stern R, Badillo W, et al. A community-based approach to linking injection drug users with needed services through pharmacies: an evaluation of a pilot intervention in New York City. AIDS Educ Prev 2010 Jul;22(3):238-251 [FREE Full text] [doi: 10.1521/aeap.2010.22.3.238] [Medline: 20528131]
25. Rita O, Regan M, Julita V, Kathrin G. Tailoring Persuasive Health Games to Gamer Type. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27, 2013; Paris, France URL: http://eprints.lincoln.ac.uk/id/eprint/13641/1/__ddat02_staffhome_jpartridge_p2467-orji.pdf [doi: 10.1145/2470654.2481341]
26. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91 [FREE Full text] [doi: 10.1177/1460458216650979] [Medline: 27245673]
27. Orji T, Mandryk RL. Developing culturally relevant design guidelines for encouraging healthy eating behavior. Int J Hum Comput Stud 2014 Feb;72(2):207-223 [FREE Full text] [doi: 10.1016/j.ijhcs.2013.08.012]
28. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
29. Westergaard RP, Hull SJ, Merkow A, Stephens LK, Hochstatter KR, Olson-Streed HK, et al. Computerized tailored interventions to enhance prevention and screening for hepatitis C virus among people who inject drugs: protocol for a randomized pilot study. JMIR Res Protoc 2016 Jan 22;5(1):e15 [FREE Full text] [doi: 10.2196/resprot.4830] [Medline: 26800903]
30. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
31. Prochaska JO, Redding CA, Harlow LL, Rossi JS, Velicer WF. The transtheoretical model of change and HIV prevention: a review. Health Educ Q 1994;21(4):471-486. [doi: 10.1177/109019819402100410] [Medline: 7843978]
32. Richards CM, Sharif F, Eischen S, Thomas J, Wang Q, Guo H, et al. Retention of homeless smokers in the power to quit study. Nicotine Tob Res 2015 Oct;17(9):1104-1111 [FREE Full text] [doi: 10.1093/ntr/ntu210] [Medline: 25301816]
33. Baggett TP, O'Connell JJ, Singer DE, Rigotti NA. The unmet health care needs of homeless adults: a national study. Am J Public Health 2010 Jul;100(7):1326-1333. [doi: 10.2105/AJPH.2009.180109] [Medline: 20466953]
34. Marsch LA, Bickel WK. Efficacy of computer-based HIV/AIDS education for injection drug users. Am J Health Behav 2004;28(4):316-327. [doi: 10.5993/ajhb.28.4.3] [Medline: 15228968]
35. Marsch LA, Grabinski MJ, Bickel WK, Desrosiers A, Guarino H, Muehlbach B, et al. Computer-assisted HIV prevention for youth with substance use disorders. Subst Use Misuse 2011;46(1):46-56 [FREE Full text] [doi: 10.3109/10826084.2011.521088] [Medline: 21190405]
36. Ondersma SJ, Chase SK, Svikis DS, Schuster CR. Computer-based brief motivational intervention for perinatal drug use. J Subst Abuse Treat 2005 Jul;28(4):305-312 [FREE Full text] [doi: 10.1016/j.jsat.2005.02.004] [Medline: 15925264]
37. Ondersma SJ, Svikis DS, Schuster CR. Computer-based brief intervention a randomized trial with postpartum women. Am J Prev Med 2007 Mar;32(3):231-238 [FREE Full text] [doi: 10.1016/j.amepre.2006.11.003] [Medline: 17236741]
38. Hester RK, Delaney HD. Behavioral self-control program for windows: results of a controlled clinical trial. J Consult Clin Psychol 1997 Aug;65(4):686-693. [doi: 10.1037//0022-006x.65.4.686] [Medline: 9256570]
39. Hester RK, Squires DD, Delaney HD. The Drinker's check-up: 12-month outcomes of a controlled clinical trial of a stand-alone software program for problem drinkers. J Subst Abuse Treat 2005 Mar;28(2):159-169. [doi: 10.1016/j.jsat.2004.12.002] [Medline: 15780546]
40. Bickel WK, Marsch LA, Buchhalter AR, Badger GJ. Computerized behavior therapy for opioid-dependent outpatients: a randomized controlled trial. Exp Clin Psychopharmacol 2008 May;16(2):132-143 [FREE Full text] [doi: 10.1037/1064-1297.16.2.132] [Medline: 18489017]
41. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, et al. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. Am J Psychiatry 2008 Jul;165(7):881-888 [FREE Full text] [doi: 10.1176/appi.ajp.2008.07111835] [Medline: 18450927]
42. Gagnon H, Godin G, Alary M, Bruneau J, Otis J. A randomized trial to evaluate the efficacy of a computer-tailored intervention to promote safer injection practices among drug users. AIDS Behav 2010 Jul;14(3):538-548. [doi: 10.1007/s10461-009-9651-x] [Medline: 20033276]
43. Jamner MS, Wolitski RJ, Corby NH. Impact of a longitudinal community HIV intervention targeting injecting drug users' stage of change for condom and bleach use. Am J Health Promot 1997;12(1):15-24. [doi: 10.4278/0890-1171-12.1.15] [Medline: 10170430]
44. Bowen AM, Trotter R. HIV risk in intravenous drug users and crack cocaine smokers: predicting stage of change for condom usep. J Consult Clin Psychol 1995 May;63(2):238-248. [doi: 10.1037//0022-006x.63.2.238] [Medline: 7751484]
45. Patten S, Vollman A, Thurston W. The utility of the transtheoretical model of behavior change for HIV risk reduction in injection drug users. J Assoc Nurses AIDS Care 2000;11(1):57-66. [doi: 10.1016/S1055-3290(06)60422-6] [Medline: 10670007]

Abbreviations

CTI: computer-tailored intervention
HCV: hepatitis C virus
RCT: randomized controlled trial
SSP: syringe service program

Edited by G Eysenbach; submitted 29.04.20; peer-reviewed by A McLean, N Onyeakusi; comments to author 12.06.20; revised version received 05.07.20; accepted 07.07.20; published 29.09.20.

Please cite as:
Hochstatter KR, Hull SJ, Sethi AK, Burns ME, Mundt MP, Westergaard RP
Promoting Safe Injection Practices, Substance Use Reduction, Hepatitis C Testing, and Overdose Prevention Among Syringe Service Program Clients Using a Computer-Tailored Intervention: Pilot Randomized Controlled Trial
J Med Internet Res 2020;22(9):e19703
URL: http://www.jmir.org/2020/9/e19703/
doi:10.2196/19703
PMID:32990630

Review

Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis

Gideon Meyerowitz-Katz[1,2], MPH; Sumathy Ravi[1], MPhil; Leonard Arnolda[2,3], PhD; Xiaoqi Feng[2,4,5], PhD; Glen Maberly[1,5], MD; Thomas Astell-Burt[2,5], PhD

[1]Western Sydney Diabetes, Western Sydney Local Health District, Blacktown NSW, Australia

[2]School of Health and Society, University of Wollongong, Wollongong, Australia

[3]Illawarra Health & Medical Research Institute, Wollongong, Australia

[4]School of Public Health and Community Medicine, University of New South Wales, Kingsford NSW, Australia

[5]Menzies Centre for Health Policy, University of Sydney, Camperdown NSW, Australia

Corresponding Author:
Gideon Meyerowitz-Katz, MPH
Western Sydney Diabetes
Western Sydney Local Health District
Blacktown Road
Blacktown NSW,
Australia
Phone: 61 2 9881 8878
Email: gideon.meyerowitzkatz@health.nsw.gov.au

Abstract

Background: Chronic disease represents a large and growing burden to the health care system worldwide. One method of managing this burden is the use of app-based interventions; however attrition, defined as lack of patient use of the intervention, is an issue for these interventions. While many apps have been developed, there is some evidence that they have significant issues with sustained use, with up to 98% of people only using the app for a short time before dropping out and/or dropping use down to the point where the app is no longer effective at helping to manage disease.

Objective: Our objectives are to systematically appraise and perform a meta-analysis on dropout rates in apps for chronic disease and to qualitatively synthesize possible reasons for these dropout rates that could be addressed in future interventions.

Methods: MEDLINE (Medical Literature Analysis and Retrieval System Online), PubMed, Cochrane CENTRAL (Central Register of Controlled Trials), and Embase were searched from 2003 to the present to look at mobile health (mHealth) and attrition or dropout. Studies, either randomized controlled trials (RCTs) or observational trials, looking at chronic disease with measures of dropout were included. Meta-analysis of attrition rates was conducted in Stata, version 15.1 (StataCorp LLC). Included studies were also qualitatively synthesized to examine reasons for dropout and avenues for future research.

Results: Of 833 studies identified in the literature search, 17 were included in the review and meta-analysis. Out of 17 studies, 9 (53%) were RCTs and 8 (47%) were observational trials, with both types covering a range of chronic diseases. The pooled dropout rate was 43% (95% CI 29-57), with observational studies having a higher dropout rate (49%, 95% CI 27-70) than RCTs in more controlled scenarios, which only had a 40% dropout rate (95% CI 16-63). The studies were extremely varied, which is represented statistically in the high degree of heterogeneity (I^2>99%). Qualitative synthesis revealed a range of reasons relating to attrition from app-based interventions, including social, demographic, and behavioral factors that could be addressed.

Conclusions: Dropout rates in mHealth interventions are high, but possible areas to minimize attrition exist. Reducing dropout rates will make these apps more effective for disease management in the long term.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42019128737; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019128737

(J Med Internet Res 2020;22(9):e20283) doi:10.2196/20283

KEYWORDS

chronic disease; mHealth; mobile apps; attrition; dropout

Introduction

Chronic diseases are a large and growing issue worldwide, with rates increasing dramatically in recent years, including infectious diseases that are now managed chronically, such as HIV. One example is diabetes, with global prevalence nearly doubling from less than 5% in the 1990s to more than 8% today [1]. As with other chronic diseases, the economic and social cost of diabetes is enormous, with large direct health care costs often eclipsed by the societal impacts of the disease [1,2]. This has led to a large body of research focusing on how to prevent and manage these diseases, with many recommendations now advising moving from a model of care that is medically focused to patient-centric and community-focused care [3]. However, there are many difficulties in implementing programs for chronic disease prevention and care, in particular, the challenges posed by catering to a large, diverse, and growing population of people requiring these services [1,4]. One such difficulty is the dropout rate, or attrition, whereby patients discontinue use of interventions either entirely or enough that the benefit from the intervention is negligible. This is an area of particular concern for new technological innovations, such as mobile phone apps.

The management of chronic disease is often complex. Patients may be on numerous medications, follow strict dietary regimens, and have lifestyle goals to fulfil to optimally manage their disease [5]. Professional assistance from health workers—doctors, educators, dieticians, and others—is an important component of this management, but increasingly, international evidence has shown that self-management—empowering patients to manage their own care—is another effective way to improve outcomes [6-8].

Self-management interventions range from providing educational materials to highly supportive, multifaceted programs that include a variety of measures [7]. One method of self-management assistance that is increasingly popular is providing web-based eHealth or mobile apps (ie, mobile health [mHealth]) to people in order to assist in their management of their disease [9]. These interventions have demonstrated efficacy in terms of markers for management, with a recent systematic review finding that, although the evidence is preliminary, mHealth interventions are effective in reducing weight and glycated hemoglobin in people with diabetes [10]. Another recent review looking only at the efficacy of mobile apps for diabetes care found that there was limited evidence supporting the effectiveness of diabetes apps to improve blood glucose markers for people with diabetes [11]. Overall, there is a growing body of evidence that mHealth interventions, and apps in particular, may be an effective method of promoting self-management in patients.

However, a major barrier to patient care in the use of mHealth interventions is attrition. Previous research has identified that up to 80% of all participants in mHealth interventions may engage in only *minimal use* of these interventions, defined as logging in to the service less than twice, and only a small fraction of users consistently use the intervention long term [12,13]. While clinical trials often report 70% or higher retention, these are often short in duration, some fewer than two months, and may not represent the situation in real-world use [10]. One observational trial of app usage in a large real-world cohort found that only 2% had sustained continuous use of the kind that would be expected to improve clinical outcomes [14]. If only 2% of people who download an app actually use it, there is clearly minimal benefit for the majority. Demonstrating that mHealth interventions are effective in clinical trials is not enough: retention in real-world settings is a necessary precondition for these interventions to be considered effective.

This paper presents a systematic review and meta-analysis into the rate and causes of dropout in mHealth interventions for diabetes and other chronic health issues. This is divided into clinical trials and observational research studies in order to estimate the rates in both controlled and uncontrolled settings and to estimate the effect both in studies with a large support network to prevent attrition and in the more *real-world* experience that might be expected when these apps are actually rolled out into clinical practice. These were also qualitatively synthesized.

Methods

A reproducible strategy was used to identify studies examining mHealth interventions for self-management of chronic disease, either mobile app based or internet based. Studies were identified by electronically searching MEDLINE (Medical Literature Analysis and Retrieval System Online), PubMed, Cochrane CENTRAL (Central Register of Controlled Trials), and Embase from 2003 to the present day. Search terms are fully outlined below and are loosely based on previous systematic reviews looking at similar topics [15]. Searches were performed in June 2019 by GMK, and duplicates were excluded using Microsoft Excel 2013 and EndNote, version 8.0 (Clarivate).

Electronic downloads of searched titles were then performed using the data collection process for each individual database, with titles being screened by GMK and SR against inclusion criteria to determine eligibility. Abstracts were then reviewed by these two reviewers independently. Any disagreements were adjudicated between the two authors. References from included studies were also assessed to identify further trials for inclusion. Both experimental and quasi-experimental study designs were included in this review. As the analysis is based on a secondary endpoint (ie, attrition), no formal risk-of-bias tool was used to assess the quality of included studies.

For the meta-analysis, the total rate of dropout was extracted from each study, as well as the number of participants in the control and intervention groups. The primary summary measure was the rate of dropout in these trials.

Eligibility criteria for inclusion of studies are as follows—studies must meet all criteria:

1. Published in English.
2. Addressed to an adult population (≥18 years of age).
3. Either randomized controlled trials (RCTs) or observational interventions (ie, case control or cohort).
4. Look at app usage in chronic disease.
5. Include a measure of dropout and attrition.

A systematic narrative synthesis was produced to describe the included studies and their findings relating to dropout. This narrative synthesis reviewed the findings from all included studies and provided an overall summation of the subject matter.

Stata, version 15.1 (StataCorp LLC), was used to perform the meta-analysis of the included studies, using the *metaprop* command, with results pooled from RCTs looking at the rate of dropout in clinical trials. There was also a second meta-analysis, by trial type: observational versus RCT. The primary outcome was the rate of dropout. Heterogeneity was assessed using the I^2 statistic and visual inspection of funnel plots; the Egger weighted meta-regression test was used to determine the influence of publication bias. If studies are identified that attempted to prevent dropout, these will form the basis of a subgroup analysis. A sensitivity analysis was conducted looking at attrition comparing short (≤2 months) studies with long (>2 months) studies.

This study was registered at the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42019128737).

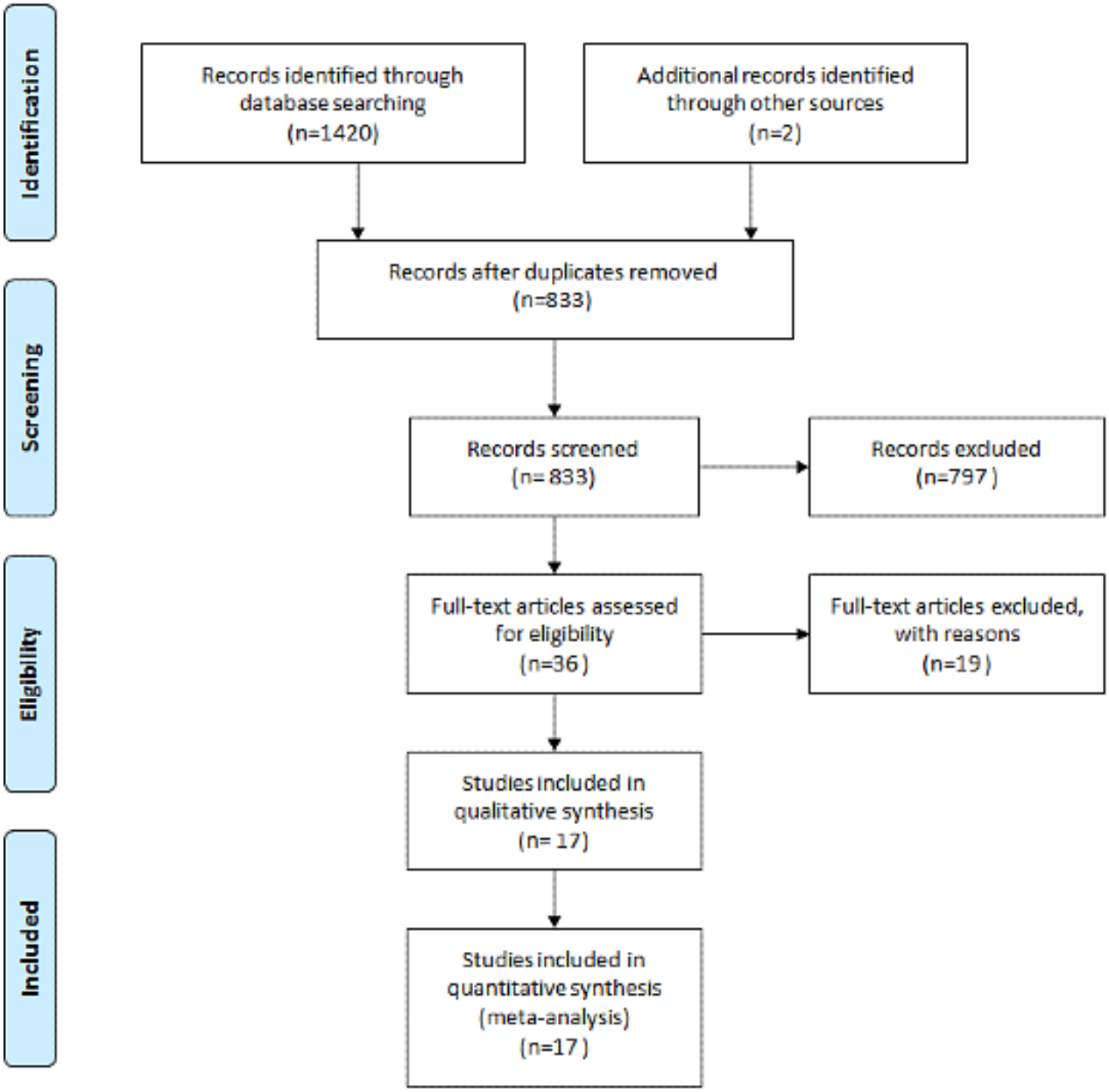
Results

Overview

Use of mHealth solutions in managing chronic conditions is increasing; however, the effective and long-term engagement (ie, attrition rate) has been attributed to various factors.

After database searches were performed, a total of 1420 articles were identified. After excluding duplicates, 831 unique records remained. Of these, 797 were excluded prior to review. A further 2 records were identified through reference screening from included studies, leaving a total of 36 studies to be included in the review (see Figure 1). Of the 36 studies included in the final review, 19 were excluded based on the exclusion criteria of studies including children and studies that only looked at acute or infectious diseases. Studies that were purely online, telephone and texting interventions, and studies that did not have any measurement of rates of dropout and attrition were also excluded. This left 17 studies to be included in the final qualitative and quantitative synthesis.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Characteristics of Included Studies

Included studies were published between 2011 and 2019 [14,16-31]. Of these, most (14/17, 82%) examined a range of chronic diseases, including single studies targeting lower back pain, chronic kidney disease, pain, dysmenorrhea, and HIV medications; the remainder (9/17, 53%) looked at more general lifestyle improvement, such as eating behavior and physical activity. Out of 17 studies, 3 (18%) that were included in the review looked specifically at diabetes. There were 9 RCTs (53%) included in the final synthesis and 8 observational trials (47%). These are summarized in Table 1.

Studies ranged significantly in duration, size, attrition rate, methodology, and other areas. The shortest trial included in this review lasted 2 weeks, and a total of 5 out of 17 (29%) lasted one month or less. Out of 9 RCTs, 2 (22%) looked at 1 year of data, and a number of observational trials were conducted over a period of 6-10 months. The lowest attrition rate in any study was 9% in an RCT lasting 1 year [29]; the highest attrition rate was 82% in an observational trial lasting 6 weeks [26]. The largest trial was an observational study lasting 24 weeks, with nearly 200,000 participants; the smallest trial was a small cohort study including just 20 people.

Table 1. Summary of studies.

Author	Name of the study	Year published	No. of participants	Area of study	Attrition rate	Type of study
Cook et al, [16]	A Counselor in Your Pocket: Feasibility of Mobile Health Tailored Messages to Support HIV Medication Adherence	2015	37	HIV medication adherence	60%	Observational study
Torbjørnsen et al, [17]	A Low-Intensity Mobile Health Intervention With and Without Health Counseling for Persons With Type 2 Diabetes, Part 1: Baseline and Short-Term Results From a Randomized Controlled Trial in the Norwegian Part of RENEWING HEALTH	2014	151	Self-management support for people with type 2 diabetes	18%	RCT[a]
Elbert et al, [18]	A Mobile Phone App Intervention Targeting Fruit and Vegetable Consumption: The Efficacy of Textual and Auditory Tailored Health Information Tested in a Randomized Controlled Trial	2016	342	Fruit and vegetable consumption	55%	RCT
Selter et al, [19]	An mHealth App for Self-Management of Chronic Lower Back Pain (Limbr): Pilot Study	2018	93	Self-management of chronic lower back pain	62%	Pilot study
Lee et al, [20]	Effect of Self-Monitoring on Long-Term Patient Engagement With Mobile Health Applications	2018	1439	Self-monitoring health app	46%	Observational study
Chen et al, [21]	Effects of Journaling Dietary Intake App on the Health Outcomes of Chronic Kidney Disease Stage 3B-5	2016	20	Chronic kidney disease	35%	Observational study
Mak et al, [22]	Efficacy and Moderation of Mobile App-Based Programs for Mindfulness-Based Training, Self-Compassion Training, and Cognitive Behavioral Psychoeducation on Mental Health: Randomized Controlled Noninferiority Trial	2018	2161	Mental well-being	76.5% and 83.9%	RCT
Guertler et al, [23]	Engagement and Nonusage Attrition With a Free Physical Activity Promotion Program: The Case of 10,000 Steps Australia	2015	16,948	Physical activity promotion: 10,000 steps	25% to 75%	Observational study
Helander et al, [14]	Factors Related to Sustained Use of a Free Mobile App for Dietary Self-Monitoring With Photography and Peer Feedback: Retrospective Cohort Study	2014	189,770	Diet self-monitoring	86.39%	Retrospective cohort study
Fukuoka et al, [24]	Identifying Factors Associated With Dropout During Prerandomization Run-in Period From an mHealth Physical Activity Education Study: The mPED Trial	2015	318	Physical activity education	34%	Observational study
Spring et al, [25]	Multicomponent mHealth Intervention for Large, Sustained Change in Multiple Diet and Activity Risk Behaviors: The Make Better Choices 2 Randomized Controlled Trial	2018	212	Diet behavior	17.90%	RCT
Roepke et al, [26]	Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms	2015	283	Reduce depressive symptoms	81.66%	RCT
Druce et al, [31]	Recruitment and Ongoing Engagement in a UK Smartphone Study Examining the Association Between Weather and Pain: Cohort Study	2017	6370	Weather and pain	N/A[b]	Observational study
Hales et al, [27]	Social Networks for Improving Healthy Weight Loss Behaviors for Overweight and Obese Adults: A Randomized Clinical Trial of the Social Pounds Off Digitally (Social POD) Mobile App	2016	51	Weight loss behavior	12%	RCT
Blödt et al, [28]	Effectiveness of App-Based Self-Acupressure for Women With Menstrual Pain Compared to Usual Care: A Randomized Pragmatic Trial	2018	221	App-based self-acupressure for menstrual pain	14%	RCT

Author	Name of the study	Year published	No. of participants	Area of study	Attrition rate	Type of study
Karhula et al, [29]	Telemonitoring and Mobile Phone-Based Health Coaching Among Finnish Diabetic and Heart Disease Patients: Randomized Controlled Trial	2015	517	Health coaching: diabetes and heart disease patients	8.90%	RCT
Holmen et al, [30]	A Mobile Health Intervention for Self-Management and Lifestyle Change for Persons With Type 2 Diabetes, Part 2: One-Year Results From the Norwegian Randomized Controlled Trial RENEWING HEALTH	2014	151	Self-management support for people with type 2 diabetes	21%	RCT

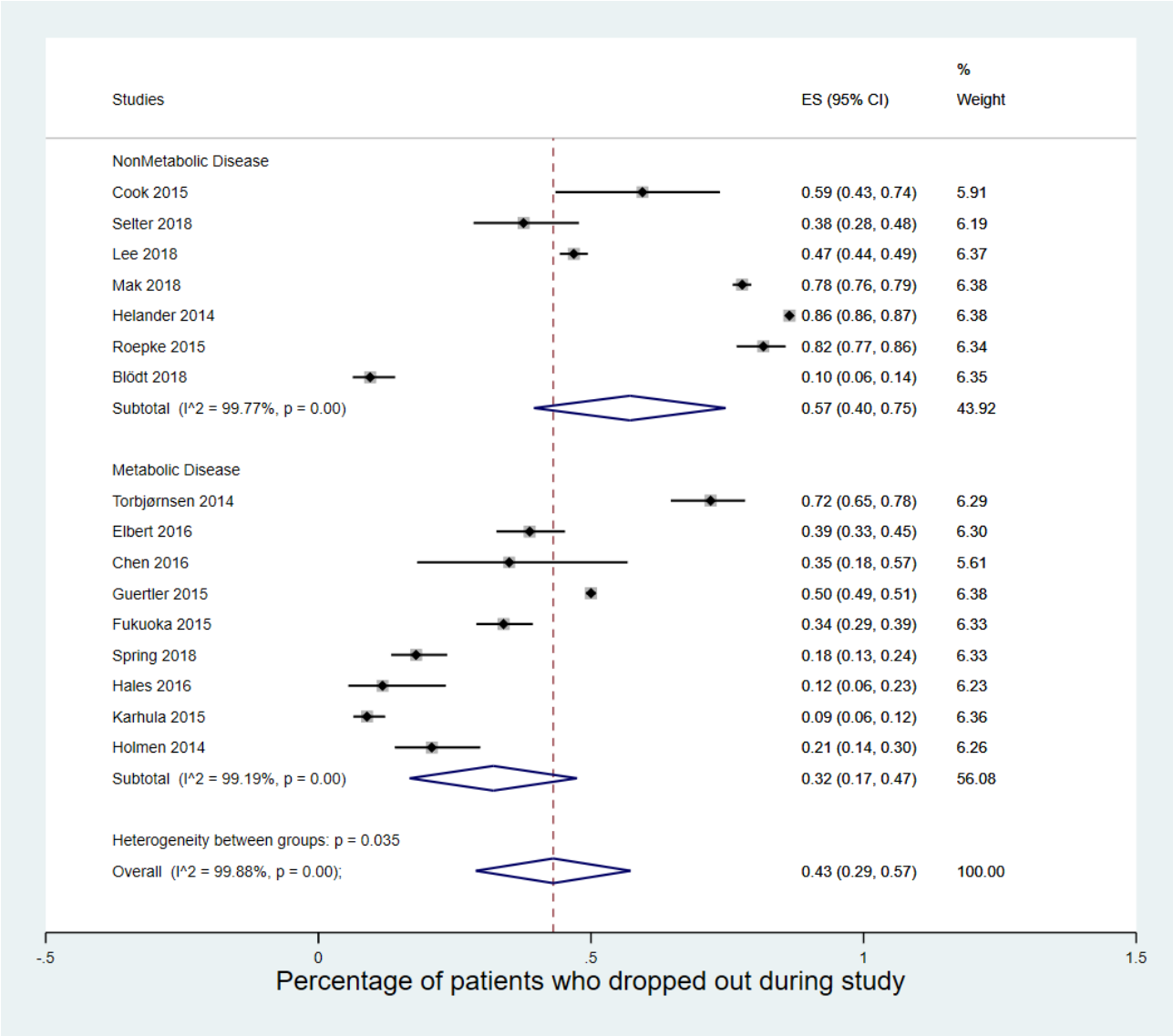
[a]RCT: randomized controlled trial.

[b]N/A: not applicable; this value was not reported.

Meta-Analysis

Results from the meta-analysis are presented in Figure 2. The average attrition rate overall was 43% (95% CI 29-57), with very high between-study heterogeneity indicated by an I^2 statistic of >99%. The very high heterogeneity is not unexpected in this case, as studies were extremely varied in terms of time, implementation, and the disease state that they were examining.

Figure 2. Meta-analysis of attrition rates in app-based intervention studies for chronic disease. ES: effect size.



Looking at the breakdown of results by the type of study, there was a higher degree of attrition in the observational *real-world* studies (49%, 95% CI 27-70) than in the RCTs in more controlled scenarios, with only 40% (95% CI 16-63) dropping out. Sensitivity analyses looking at differences in length of study (ie, short vs long), diabetes versus other chronic diseases, or

whether the studies were numerically large did not find any similar differences in attrition rate between trials.

Attrition Rates

One reason associated with lower attrition rates was the behavioral characteristics of the included participants. Low attrition rates were characterized by reasons such as the perception of own health as poor—thus, incentivizing the need to change [18]—and those who wanted to be involved in their health care [20]. Other factors that were associated with attrition in included studies were health literacy; age, with younger participants dropping out less; and postgraduate education [18,22]. Very low attrition was also reported among those who were on strict diets or who had been healthy eaters prior to the initiation of the study [14]. Another association with low attrition was with those engaged in multiple interventions. Those engaged in internet or phone programs as well as apps were more likely to remain in the research study [23]. Conversely, there did not appear to be much influence on attrition rates in terms of length of study, the disease studied, or the size of the app trial.

Findings from these studies suggest ways to improve attrition rate and long-term engagement by using varying message contents or formats to maintain users' interests; for example, tailored messages may have the potential to improve adherence to a clinically significant degree [16]. Other studies suggest less of a benefit from tailoring messages to maintain users' interest; despite a low attrition rate of 22% at 4 months and 1 year in two studies, respectively, an app and health counseling did not reduce hemoglobin A1c levels between the intervention and usual care groups [17,30]. In addition, self-management skills and the ability to contact health professionals were found to increase engagement, while users' feedback input improved usability of apps and enhanced user experiences for daily self-reports [17,19]. Classifying different types of users may be important in improving long-term engagement. Low retention rate might also be due to an unguided self-help approach, and further engaging those who need self-monitoring remains challenging.

Another issue with attrition was that definitions varied significantly. While some studies reported users who only logged in a single time as *dropouts*, others expanded the definition to include those who only used an app once or twice. For example, the RCT with the lowest dropout rate overall included patients who only sent a single report through the app during the entire follow-up time, which did not indicate sustained, long-term use [29]. While these users may have been nondropouts by the definitions used in that study, they had received significantly lower benefits from the intervention, and would likely be considered dropouts had the analysis been less broad.

Discussion

Principal Findings

Attrition in app-based interventions is an important and yet underresearched element. For these interventions to work, it is a necessary component that people use and continue to use the app; however, there appears to be evidence that this is not always the case. In this systematic review and meta-analysis, the pooled estimate of dropout rates was 43%, with higher rates seen in real-world research and lower rates in highly supported RCTs. This may indicate a very serious underlying issue, as high dropout rates in these interventions will limit their use and uptake in health care across a range of chronic diseases.

While dropout rates in RCTs were notably lower than in observational trials, it is worth noting that attrition was often defined differently in this research. RCTs tended to describe all participants as users of the app unless they had ceased using it entirely; while this is in line with best-practice intention-to-treat analyses, it also presents an important limitation with the pooled results above. Including people in analyses who have been randomized is laudable; however, this also obscures the fact that large proportions of people, even in these randomized trials with detailed patient support and follow-up, barely used the app, if at all. This is worrying, because it implies that even with very high levels of support, apps are not an intervention with substantial staying power for people with various chronic diseases. It is also worth recognizing that combining estimates for both observational and RCT studies may have some drawbacks, given the differences in study design and participant retention. However, the difference in reported dropout rates was not significantly different, with only a small difference in point estimates and overlapping confidence intervals. This is likely because the main divergence was in definition of attrition, rather than the specifics of the study per se. As reported above, even among RCTs with low rates of dropout, there may have been a number of studies that would meet the looser definitions of *attrition* used in some observational research.

It was also concerning that there does not appear to have been much examination of the reasons behind this attrition in many studies. Few studies attempted to explain why people dropped out, with this being attributed to health literacy, age, and education, but it is unlikely that these are the only factors that would be related to attrition in the use of apps. For example, as mentioned in the results, people whose health saw greater improvements were more likely in some studies to keep using the app. It is likely that there are a range of unidentified issues that could potentially be targeted to ameliorate this problem, but thus far there has been little recognition of the issue formally, which may have limited the research that has been done to remedy the situation. Many studies do not even address the possibility that people dropping out of an app-based intervention at alarming rates could be an issue for the intervention's adoption at scale, nor do they address the issue that this could cause in terms of aggravating health inequities depending on the reasons for dropout. This is especially concerning when considering that age and social status are likely to be barriers to app access—as some included studies hinted—which may further compound the issues caused by selection bias of those who use apps in the first place. If younger, healthier people are more likely to use apps overall, which is often the case [32], and are then more likely to use them long term, the apps may be less useful for the very populations that we most want them to help.

This is a common theme among trials, in which attrition or nonusage is barely addressed, or only given very surface-level appraisal. If there is a significant difference in the primary outcome between the intervention and control groups, there is a general attitude that the attrition is unimportant; this appears to be fairly common in RCTs, and may be because the aim of this research is specifically to evaluate the app in an intention-to-treat framework [29,30]. However, there are clear drawbacks to this, not least that we may be seeing a large underestimate in the literature of the efficacy of app-based interventions, caused by their generally low use in the populations who have been studied.

There are a number of very important limitations to this research. Firstly, the estimates produced are certainly not comprehensive. Many studies (N>30) that fit all of the inclusion criteria failed to report dropout or attrition in a way that could be extracted. Given the number of trials on app-based interventions, it was not considered feasible to follow up with every author group that had these figures, but it is worth noting that this best guess represents a relatively small number of trials within the total pool of potential evidence.

There is also the issue with heterogeneity. Given the nature of the included studies, it is not surprising to find very significant levels of heterogeneity statistically, but it is concerning for the meta-analysis as a reasonable estimate of attrition. These studies were conducted across different disease states, with highly variable interventions; the fact that they all included an app is a thin bond that did not overcome the vast differences that they had between them.

It is also worth noting that the research in this space is quickly evolving. We found no published research to be included before 2010, very little in the years leading up to 2015, and then an explosion of studies in the years after. It is likely that redoing this meta-analysis in 2025 will yield a much more reliable estimate of the figures. This may also allow for analyzing by disease state, which could prove to be a more accurate estimation of the rate of attrition.

There are a number of theories that might pertain to attrition in app-based interventions, with several different focuses. One area that could help inform attrition research in the future is behavioral theory, perhaps by examining the sociocognitive aspects of people who do and do not drop out of app-based studies. Integrated behavioral theories might also be useful in examining the relationship between social factors and the behaviors they cause, to come to an understanding of the process by which people decide to use or discontinue using apps.

This would ideally tie in to an examination of the broader social and demographic drivers of attrition. While few of these drivers have been identified in research so far—age, in particular—there remains a large evidential gap pertaining to how society influences behavior to prevent people from using app-based interventions. Future research should combine these two theoretical approaches to define the background reasons for attrition, so that interventions can be designed to minimize it.

Aside from the estimates of app attrition, there are some important implications of this research. Future studies looking at app-based interventions should include attrition as a secondary endpoint and develop methods to prevent it if possible. One important aspect would be to develop a standard measure of minimum use in app-based interventions; a reasonable example is the one used in some of the included trials of one or fewer log-ins to the app in any given period of time (ie, one log-in per month). Lower use than this basic threshold could then be considered *attrition* for the purposes of research studies. As well, there should be trials looking at ways to reduce the rate of dropouts, as well as the potential inequity in the rate of attrition, in app-based interventions. Without such research, we have no way of knowing if apps can be effective in the general population.

Conclusions

This systematic review and meta-analysis found that the pooled estimate for dropout rates in trials of app-based interventions for chronic diseases was 43% over a variety of timelines, with the length of time having little impact on the rate of dropout. Attrition was higher in observational *real-world* studies, with randomized clinical research seeing less than a third of patients drop out before the trials were completed. However, findings were limited by high heterogeneity and the lack of reporting in many trials on attrition rates. Future research should focus on how often patients drop out and examine reasons why, so that this important issue can be addressed in app-based interventions for chronic disease.

Conflicts of Interest

None declared.

References

1. Global Report on Diabetes. Geneva, Switzerland: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;jsessionid=A20B0AE31F0C6512AE42FBDC695F695F?sequence=1 [accessed 2020-09-22]
2. Meyerowitz-Katz G, Seelan S, Gaur P, Francisco R, Ferdousi S, Astell-Burt T, et al. Detecting the hidden burden of pre-diabetes and diabetes in Western Sydney. Diabetes Res Clin Pract 2019 May;151:247-251. [doi: 10.1016/j.diabres.2019.04.019] [Medline: 31004674]
3. Diabetes: The Silent Pandemic and its Impact on Australia. Canberra, Australia: Diabetes Australia; 2015. URL: https://static.diabetesaustralia.com.au/s/fileassets/diabetes-australia/e7282521-472b-4313-b18e-be84c3d5d907.pdf [accessed 2020-05-01]

4. Caperchione CM, Kolt GS, Mummery WK. Physical activity in culturally and linguistically diverse migrant groups to Western society: A review of barriers, enablers and experiences. Sports Med 2009;39(3):167-177. [doi: 10.2165/00007256-200939030-00001] [Medline: 19290674]
5. Colagiuri S. Optimal management of type 2 diabetes: The evidence. Diabetes Obes Metab 2012 Jan;14 Suppl 1:3-8. [doi: 10.1111/j.1463-1326.2011.01506.x] [Medline: 22118704]
6. Funnell MM, Brown TL, Childs BP, Haas LB, Hosey GM, Jensen B, et al. National standards for diabetes self-management education. Diabetes Care 2008 Dec 31;32(Supplement_1):S87-S94. [doi: 10.2337/dc09-s087]
7. Grady PA, Gough LL. Self-management: A comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/ajph.2014.302041]
8. Zhao F, Suhonen R, Koskinen S, Leino-Kilpi H. Theory-based self-management educational interventions on patients with type 2 diabetes: A systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2017 Apr;73(4):812-833. [doi: 10.1111/jan.13163] [Medline: 27681948]
9. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the development of mobile phone apps to support diabetes self-management: Systematic review. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]
10. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
11. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
12. Pfammatter AF, Mitsos A, Wang S, Hood SH, Spring B. Evaluating and improving recruitment and retention in an mHealth clinical trial: An example of iterating methods during a trial. Mhealth 2017;3:49 [FREE Full text] [doi: 10.21037/mhealth.2017.09.02] [Medline: 29184901]
13. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
14. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: Retrospective cohort study. J Med Internet Res 2014 Apr 15;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
15. Walker RJ, Smalls BL, Campbell JA, Strom Williams JL, Egede LE. Impact of social determinants of health on outcomes for type 2 diabetes: A systematic review. Endocrine 2014 Sep;47(1):29-48 [FREE Full text] [doi: 10.1007/s12020-014-0195-0] [Medline: 24532079]
16. Cook PF, Carrington J, Schmiege SJ, Starr W, Reeder B. A counselor in your pocket: Feasibility of mobile health tailored messages to support HIV medication adherence. Patient Prefer Adherence 2015 Sep:1353. [doi: 10.2147/ppa.s88222]
17. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A low-intensity mobile health intervention with and without health counseling for persons with type 2 diabetes, Part 1: Baseline and short-term results from a randomized controlled trial in the Norwegian part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
18. Elbert SP, Dijkstra A, Oenema A. A mobile phone app intervention targeting fruit and vegetable consumption: The efficacy of textual and auditory tailored health information tested in a randomized controlled trial. J Med Internet Res 2016 Jun 10;18(6):e147 [FREE Full text] [doi: 10.2196/jmir.5056] [Medline: 27287823]
19. Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, et al. An mHealth app for self-management of chronic lower back pain (Limbr): Pilot study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179 [FREE Full text] [doi: 10.2196/mhealth.8256] [Medline: 30224333]
20. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
21. Chen J, Wu L, Chou S. Effects of journaling dietary intake app on the health outcomes of chronic kidney disease stage 3B-5. In: Proceedings of the 49th Annual Hawaii International Conference on System Sciences. 2016 Presented at: 49th Annual Hawaii International Conference on System Sciences; January 5-8, 2016; Koloa, HI p. 3379-3388. [doi: 10.1109/HICSS.2016.421]
22. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: Randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
23. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: The case of 10,000 Steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]

24. Fukuoka Y, Gay C, Haskell W, Arai S, Vittinghoff E. Identifying factors associated with dropout during prerandomization run-in period from an mHealth physical activity education study: The mPED Trial. JMIR Mhealth Uhealth 2015 Apr 13;3(2):e34 [FREE Full text] [doi: 10.2196/mhealth.3928] [Medline: 25872754]
25. Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, et al. Multicomponent mHealth intervention for large, sustained change in multiple diet and activity risk behaviors: The Make Better Choices 2 randomized controlled trial. J Med Internet Res 2018 Jun 19;20(6):e10528 [FREE Full text] [doi: 10.2196/10528] [Medline: 29921561]
26. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
27. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: A randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
28. Blödt S, Pach D, von Eisenhart-Rothe S, Lotz F, Roll S, Icke K, et al. Effectiveness of app-based self-acupressure for women with menstrual pain compared to usual care. Obstet Gynecol Surv 2018;73(6):348-350. [doi: 10.1097/ogx.0000000000000558]
29. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: Randomized controlled trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
30. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, Part 2: One-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
31. Druce KL, McBeth J, van der Veer SN, Selby DA, Vidgen B, Georgatzis K, et al. Recruitment and Ongoing Engagement in a UK Smartphone Study Examining the Association Between Weather and Pain: Cohort Study. JMIR Mhealth Uhealth 2017 Nov 01;5(11):e168 [FREE Full text] [doi: 10.2196/mhealth.8162] [Medline: 29092810]
32. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]

Abbreviations

CENTRAL: Central Register of Controlled Trials
MEDLINE: Medical Literature Analysis and Retrieval System Online
mHealth: mobile heath
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 14.05.20; peer-reviewed by C Reis, W Zhang; comments to author 29.07.20; revised version received 30.07.20; accepted 03.08.20; published 29.09.20.

Please cite as:
Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T
Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(9):e20283
URL: http://www.jmir.org/2020/9/e20283/
doi:10.2196/20283
PMID:32990635

Original Paper

Impact of a Blended Periconception Lifestyle Care Approach on Lifestyle Behaviors: Before-and-After Study

Melissa van der Windt[1], MD; Rianne Maria van der Kleij[1,2], MSc, PhD; Katinka Marianne Snoek[1], MD; Sten Paul Willemsen[3], MSc, PhD; Ramon Henny Maria Dykgraaf[1], MD; Joop Stephanus Elisabeth Laven[1], MD, PhD; Sam Schoenmakers[1*], MD, PhD; Régine Patricia Maria Steegers-Theunissen[1*], MD, PhD

[1]Department of Obstetrics and Gynecology, Erasmus University Medical Center, Rotterdam, Netherlands

[2]Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[3]Department of Biostatistics, Erasmus University Medical Center, Rotterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Régine Patricia Maria Steegers-Theunissen, MD, PhD
Department of Obstetrics and Gynecology
Erasmus University Medical Center
PO Box 2040
Rotterdam, 3000CA
Netherlands
Phone: +31 107038255
Email: r.steegers@erasmusmc.nl

Abstract

Background: Periconception lifestyle behaviors affect maternal, paternal, offspring, and transgenerational health outcomes. Previous research in other target populations has shown that personalized lifestyle interventions, in which face-to-face counseling and eHealth ("blended care") are combined, may effectively target these lifestyle behaviors.

Objective: We aimed to assess the effectiveness of a periconceptional lifestyle intervention on the improvement of specific lifestyle components.

Methods: A blended periconception lifestyle care approach was developed, combining the outpatient lifestyle counseling service "Healthy Pregnancy" with the eHealth platform "Smarter Pregnancy" (www.smarterpregnancy.co.uk) in which lifestyle was coached for 24 weeks. All couples contemplating pregnancy or already pregnant (≤12 weeks of gestation) who visited the outpatient clinics of the Department of Obstetrics and Gynecology at the Erasmus University Medical Center (Erasmus MC), Rotterdam, the Netherlands, between June and December 2018, were invited to participate. We measured changes in lifestyle behaviors at weeks 12 and 24 compared with baseline. Generalized estimating equations were used to analyze the changes in lifestyle behaviors over time. Subgroup analyses were performed for women with obesity (BMI ≥30 kg/m^2), women pregnant at the start of the intervention, and those participating as a couple.

Results: A total of 539 women were screened for eligibility, and 450 women and 61 men received the blended periconception intervention. Among the participating women, 58.4% (263/450) were included in the preconception period. Moreover, 78.9% (403/511) of the included participants completed the online lifestyle coaching. At baseline, at least one poor lifestyle behavior was present in most women (379/450, 84.2%) and men (58/61, 95.1%). In the total group, median fruit intake increased from 1.8 to 2.2 pieces/day (P<.001) and median vegetable intake increased from 151 to 165 grams/day (P<.001) after 24 weeks of online coaching. The probability of taking folic acid supplementation among women increased from 0.97 to 1 (P<.001), and the probability of consuming alcohol and using tobacco in the total group decreased from 0.25 to 0.19 (P=.002) and from 0.20 to 0.15 (P=.63), respectively. Overall, the program showed the strongest effectiveness for participating couples. Particularly for vegetable and fruit intake, their consumption increased from 158 grams/day and 1.8 pieces/day at baseline to 190 grams/day and 2.7 pieces/day at the end of the intervention, respectively.

Conclusions: We succeeded in including most participating women in the preconception period. A high compliance rate was achieved and users demonstrated improvements in several lifestyle components. The blended periconception lifestyle care approach seems to be an effective method to improve lifestyle behaviors. The next step is to further disseminate this approach and to

perform a randomized trial to compare the use of blended care with the provision of only eHealth. Additionally, the clinical relevance of these results will need to be substantiated further.

(J Med Internet Res 2020;22(9):e19378) doi:10.2196/19378

KEYWORDS

eHealth; periconception period; lifestyle intervention

Introduction

A healthy lifestyle helps to optimize the well-being and health outcomes of individuals across their life course. Previous research has shown that adhering to a healthy lifestyle is associated with an increase in the quality of life and life expectancy [1-5]. The exact definition of a "healthy lifestyle" varies over time and between authors. Predominantly, a "healthy lifestyle" is referred to as a combination of lifestyle components associated with lower mortality and better health outcomes. Most definitions include components such as nutrition, physical activity, no smoking, and minimal alcohol consumption [6-10]. According to the World Health Organization, healthy nutrition can be further specified as the consumption of at least 400 grams of fruits and vegetables per day, less than 5 mg of salt per day, and limited consumption of (free) sugars (maximum of 25 gram per day), saturated fats (less than 10% of total energy intake), and trans-fats (less than 1% of total energy intake) [11].

The impact of parental lifestyle around the time of conception on maternal, paternal, fetal, and neonatal health is a focus of increasing interest. The periconception period commences 14 weeks before conception and ends 10 weeks after conception. During this period, adherence to a healthy lifestyle is critical for optimizing gamete function, early placentation, and embryonic development [12]. These processes influence both offspring health and future health across generations [13]. Therefore, the periconception period is considered the "window of opportunity" in which the foundation for optimum growth, development, and health across the life course is established.

A lifestyle intervention can be defined as a behavioral intervention method to improve an individual's lifestyle by addressing the optimization of one or more lifestyle components. A lifestyle intervention during the periconception period has the potential to effectively target unhealthy lifestyle components and thereby improve reproductive outcome and the future health of both parents and their offspring [12]. However, several interventions have aimed to influence maternal lifestyle, but the time window to do so has been mostly too short to make an impact, with interventions starting when pregnancy is well underway [14]. Moreover, most interventions were not personalized to the individual couple, making it difficult to address their individual needs and wants. A lifestyle intervention that starts early, preferably in the preconception period, takes couples into account (not only women), and integrates the foundations of personalized medicine is therefore warranted to address the lifestyle of women and men throughout the pregnancy and perinatal periods [14-16].

The Erasmus University Medical Center (Erasmus MC) developed and tested the outpatient lifestyle counseling clinic "Achieving a Healthy Pregnancy," and this intervention was demonstrated to decrease the prevalence of unhealthy lifestyle components in a sample of 419 mostly subfertile couples [17]. Additionally, the Erasmus MC developed and evaluated the eHealth coaching program "Smarter Pregnancy" in a large survey including both a general population cohort and a subfertile (in vitro fertilization/intracytoplasmic sperm injection) cohort [18-20].

Interventions that combine face-to-face counseling with eHealth (referred to as "blended care") are increasingly applied in the context of mental health care [21]. Moreover, promising results have been shown for using blended care in lifestyle medicine [22]. With blended care, intervention effectiveness may be enhanced, as this treatment modality improves user motivation, engagement, and self-management, and decreases intervention resistance (and dropouts) and the number of hospital consultations. As such, blended care has the potential to reduce total treatment cost [23,24]. However, its benefits have mainly been proven in mental health care and have not been confirmed in the specific setting of periconception lifestyle care. We therefore developed and evaluated a blended personalized lifestyle care approach to periconception care, combining an outpatient lifestyle counseling service with the eHealth platform "Smarter Pregnancy" [25]. We aimed to assess the effectiveness of the lifestyle intervention on the improvement of the specific lifestyle components vegetable and fruit intake, folic acid supplement use, tobacco use, and alcohol consumption. Furthermore, intervention compliance rates were determined. Subgroup analyses were performed to evaluate the intervention effectiveness for specific subgroups, such as pregnant women, women whose partners also participated, women suffering from obesity, and men.

Methods

Study Design and Participants

All couples contemplating pregnancy or those already pregnant (≤12 weeks of gestation) who visited the outpatient clinics of the Department of Obstetrics and Gynecology at the Erasmus MC, Rotterdam, the Netherlands, were invited to participate. In our study, we invited couples to participate between June 2018 and December 2018. The exclusion criteria were pre-existing type 1 diabetes mellitus, insufficient knowledge of the Dutch language, and inability to provide informed consent. A woman could be included as a single participant if her partner did not participate.

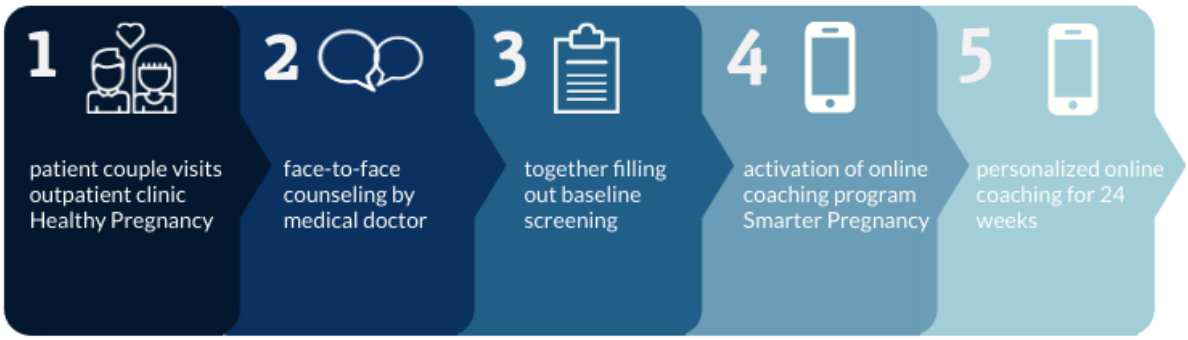
This study was approved by the medical ethics and institutional review board of the Erasmus MC, Rotterdam, the Netherlands (MEC-2018-1232). Written informed consent was obtained from all female and male participants at enrollment.

Intervention

The blended care approach comprised the following two integrated parts: the outpatient clinic Healthy Pregnancy and the online eHealth coaching program Smarter Pregnancy. The study process is presented in Figure 1. The outpatient clinic Healthy Pregnancy was based on the previously proven to be effective outpatient clinic Achieving a Healthy Pregnancy [17]. The online eHealth program included 6 months of personalized coaching on the most prevalent inadequate nutrition and lifestyle behaviors (vegetable, fruit, and folic acid intake, tobacco use, and alcohol consumption). Use of the online coaching platform improved lifestyle components of the participants after 6 months by 26.3% (95% CI 23.0-29.9) for vegetable intake, 38.4% (95% CI 34.5-42.5) for fruit intake, 56.3% (95% CI 48.8-63.6) for folic acid supplement use, 35.1% (95% CI 29.1-41.6) for no tobacco use, and 41.9% (95% CI 35.2-48.9) for no alcohol consumption.

The counselor, a medical doctor trained in motivational interviewing, filled out the baseline screening of Smarter Pregnancy [25] together with the couple during a visit to the outpatient clinic Healthy Pregnancy. Following the results of the baseline screening of Smarter Pregnancy, the counselor provided the couple with tailored lifestyle advice and possible options to alter their lifestyle. The information obtained during the face-to-face session was used to personalize the 6 months of online coaching provided by the eHealth platform Smarter Pregnancy. Participating couples, both women and men, received up to three short motivating and supporting messages per week by email. These messages included vouchers, seasonal recipes, and personalized tips and recommendations to achieve a healthy diet and to increase physical activity. Every 6 weeks, additional questions addressing behavior, diet, and pregnancy status were sent to monitor lifestyle changes. All participants (both women and men) had access to their personal page on the online Smarter Pregnancy platform, which provided additional modules to encourage the performance of physical activity, increase compliance with hospital appointments, and optimize medication adherence. A summary of all individual results can also be extracted or shared with the health care professional for further evaluation and support of preconception and antenatal care.

Figure 1. Diagram of the study process.



Measurement Instruments and Data Collection

The lifestyle indicators weight and height were measured during the face-to-face appointment. The BMI was calculated by dividing weight in kilograms by the square of height in meters. Using a lifestyle questionnaire integrated in the online coaching program Smarter Pregnancy, baseline screening and follow-up screening at weeks 6, 12, 18, and 24 of the program were performed. The follow-up screening was used to monitor the change in lifestyle components and the pregnancy state. At weeks 12 and 24 of the program, participants were invited to fill out a short questionnaire about all lifestyle components and the pregnancy state. At weeks 6 and 18, they received a short questionnaire that only included questions about lifestyle components that were found to be inadequate at baseline. The lifestyle questionnaire included questions regarding vegetable and fruit intake, folic acid supplement use, tobacco use, and alcohol consumption.

Outcomes

The blended care approach focused on the following five lifestyle components: vegetable and fruit intake, folic acid supplement use, tobacco use, and alcohol consumption. Adequate lifestyle behaviors, considered as a healthy lifestyle, were formulated as intake of at least 200 grams of vegetables per day, intake of at least two pieces of fruit per day, low dose (400 µg) folic acid supplement use (only for female participants), no tobacco use, and no alcohol consumption.

Statistical Analysis

Participants who completed the blended lifestyle care approach and those who resigned prematurely were included in the analysis. Compliance was defined as the percentage of participants who completed the 6-month blended care approach. Baseline characteristics of the study participants were computed as medians or percentages and were calculated separately for women who were pregnant at the start of the approach, women who were trying to conceive, and male participants. Vegetable

intake and fruit intake were analyzed as continuous variables. The results of dichotomous variables, folic acid supplement use, tobacco use, and alcohol consumption were shown as probabilities.

Generalized estimating equations were used to analyze the changes in lifestyle behaviors. Subgroup analyses were performed for obese women (BMI ≥30 kg/m^2), pregnant women at the start of the program, and couples. We hypothesized that women with obesity may have more unhealthy lifestyle behaviors, pregnant women may have healthier lifestyle behaviors, and women who participated with their partners may be more responsive to the blended care approach. A *P* value <.05 was considered statistically significant. All analyses were performed using SPSS package 21.0 (IBM SPSS Statistics) and R (R for Windows, version 3.5; R Core Team).

Results

Study Population

A total of 511 patients (450 women and 61 men) received the blended lifestyle care approach. Additionally, 58.4% (263/450) of female participants were included during the preconception period. The median BMI among female participants was 24.8 kg/m^2. Among female participants, 27.6% (124/450) were overweight (BMI 25-30 kg/m^2) and 20.9% (94/450) were obese (BMI ≥30 kg/m^2). The median BMI among male participants was 25.5 kg/m^2. Among male participants, 38.2% (21/55) were overweight and 16.4% (9/55) were obese. The intervention compliance rate was 78.9% (403/511) among all participants. The highest compliance, defined as the percentage of participants who completed the online coaching program Smarter Pregnancy, was achieved in the group of women who were not pregnant at the start of the approach (215/263, 81.7%).

Baseline Lifestyle Behaviors

The median vegetable intake was 157 grams per day for women and 146 grams per day for men (Table 1). Adequate vegetable intake was achieved in 31.6% (137/434) of participating women and 25.0% (14/56) of participating men. The median fruit intake was 1.9 pieces per day for women and 1.4 pieces per day for men. Adequate fruit intake was observed in 49.1% (212/432) and 27.3% (15/55) of women and men, respectively. Folic acid intake was adequate at baseline in 76.7% (345/450) of participating women, and was adequate in 96.8% (181/187) and 62.4% (164/263) of women pregnant at the start of the program and nonpregnant women, respectively. Regular tobacco use was reported by 8.8% (38/431) of all women and 4.6% (8/175) of women who were pregnant at the start of the approach. On the other hand, 20.0% (11/55) of men reported tobacco use, and their partners were predominantly in the preconception period (10/11, 90.9%). Alcohol consumption was reported by 20.5% (88/430) of participating women, 1.1% (2/175) of women who were pregnant at the start of the approach, and 65.5% (36/55) of men. All baseline lifestyle behaviors were worse in nonpregnant women and men compared with women who were pregnant at the start of the approach. Subgroup analyses of lifestyle behaviors between normal weight (n=356) and obese women (n=94) showed no relevant differences at baseline (Table 2).

Table 1. Baseline characteristics of the study participants stratified by sex and pregnancy status.

Characteristic	Women (N=450)	Pregnant women (n=187)	Nonpregnant women (n=263)	Men (N=61)
Age				
Overall value (years), median (IQR)	32.3 (28.5-36.2)	31.8 (28.0-35.2)	32.5 (29.0-36.6)	33.6 (30.9-39.5)
Missing data, n	0	0	0	0
BMI				
Overall value (kg/m^2), median (IQR)	24.8 (22.1-29.1)	24.5 (21.9-29.0)	25.2 (22.3-29.4)	25.5 (22.6-29.0)
Overweight (BMI 25-30 kg/m^2), median (IQR)	27.5 (25.8-28.7)	27.3 (25.7-28.8)	27.5 (26.0-28.7)	27.7 (25.9-29.0)
Overweight, n (%)	124 (27.6%)	46 (24.6%)	78 (29.7%)	21 (38.2%)
Obese (BMI 30-60 kg/m^2), median (IQR)	32.9 (31.5-35.6)	32.0 (31.6-35.9)	33.8 (31.6-35.9)	34.5 (31.2-38.7)
Obese, n (%)	94 (20.9%)	37 (19.8%)	57 (21.7%)	9 (16.4%)
Missing data, n	0	0	0	6
Adequate folic acid intake				
Value, n (%)	345 (76.7%)	181 (96.8%)	164 (62.4%)	N/A[a]
Missing data, n	0	0	0	N/A
Vegetable intake				
Overall value (grams/day), median (IQR)	157 (100-214)	157 (100-207)	150 (100-214)	146 (93-209)
Adequate (≥200 grams/day), n (%)	137 (31.6%)	51 (29.0%)	86 (33.3%)	14 (25.0%)
Missing data, n	16	11	5	5
Fruit intake				
Overall value (pieces/day), median (IQR)	1.90 (0.94-3.11)	2.3 (1.3-3.7)	1.7 (0.8-2.5)	1.4 (0.8-2.1)
Adequate (≥2 pieces/day), n (%)	212 (49.1%)	106 (60.6%)	106 (41.2%)	15 (27.3%)
Missing data, n	18	12	6	6
Smoking				
Smoking (yes), n (%)	38 (8.8%)	8 (4.6%)	30 (11.7%)	11 (20.0%)
Missing data, n	19	12	7	6
Alcohol use				
Alcohol use (yes), n (%)	88 (20.5%)	2 (1.1%)	85 (33.3%)	36 (65.5%)
Missing data, n	20	12	8	6
Completed the program (yes), n (%)	358 (79.6%)	143 (76.5%)	215 (81.7%)	45 (73.8%)

[a]N/A: not applicable.

Table 2. Differences in lifestyle behaviors at baseline between normal weight and overweight women and obese women.

Characteristic	Women (N=450)		*P* value
	Normal weight and overweight (n=356)	Obese (n=94)	
Age			.20
Overall value (years), median (IQR)	32.5 (28.6-36.5)	31.6 (28.4-34.5)	
Missing data, n	0	0	
Pregnancy status			.63
Pregnant (yes), n (%)	150 (42.1%)	37 (39.4%)	
Missing data, n	0	0	
BMI			<.001
Overall value (kg/m^2), median (IQR)	23.8 (21.5-26.0)	32.8 (31.5-35.6)	
Missing data, n	0	0	
Adequate folic acid intake			.10
Value (no), n (%)	84 (23.9%)	36 (38.3%)	
Missing data, n	0	0	
Vegetable intake			.13
Overall value (grams/day), median (IQR)	157 (100-214)	128 (93-214)	
Missing data, n	13	3	
Fruit intake			.07
Overall value (pieces/day), median (IQR)	2.08 (0.95-3.24)	1.6 (0.7-2.5)	
Missing data, n	15	3	
Smoking			.12
Smoking (yes), n (%)	27 (7.6%)	12 (13.2%)	
Missing data, n	16	3	
Alcohol use			.34
Alcohol use (yes), n (%)	72 (20.2%)	15 (16.7%)	
Missing data, n	16	4	

Effectiveness

Figure 2 depicts the improvement in lifestyle behaviors for the total group and the several subgroups during the 24 weeks of online coaching.

An increase in vegetable intake was observed after 24 weeks compared with baseline intake. This improvement was found for all subgroups; however, the strongest effect was found in women whose partners also participated in the blended periconception lifestyle care approach. These women showed a mean vegetable intake of 158 grams per day at baseline, and 185 grams per day after 12 weeks (*P*<.001) and 190 grams per day after 24 weeks of online coaching (*P*<.001). Women with obesity increased their vegetable intake from 144 grams per day at baseline to 145 grams per day after 12 weeks (*P*=.87) and 147 grams per day at the end of the online coaching program (*P*=.72).

Fruit intake increased in all subgroups and this improvement was the greatest in the group of male participants. This group showed almost doubling in fruit intake at the end of the online coaching program compared with baseline (1.4 pieces per day at baseline, 2.1 pieces per day at 12 weeks [*P*<.001], and 2.3 pieces per day at 24 weeks [*P*<.001]). Subgroup analysis in the group of women with obesity showed an increase in fruit intake from 1.6 pieces per day at baseline to 1.7 pieces per day at 12 weeks (*P*=.35) and 1.8 pieces per day at 24 weeks of online coaching (*P*=.05).

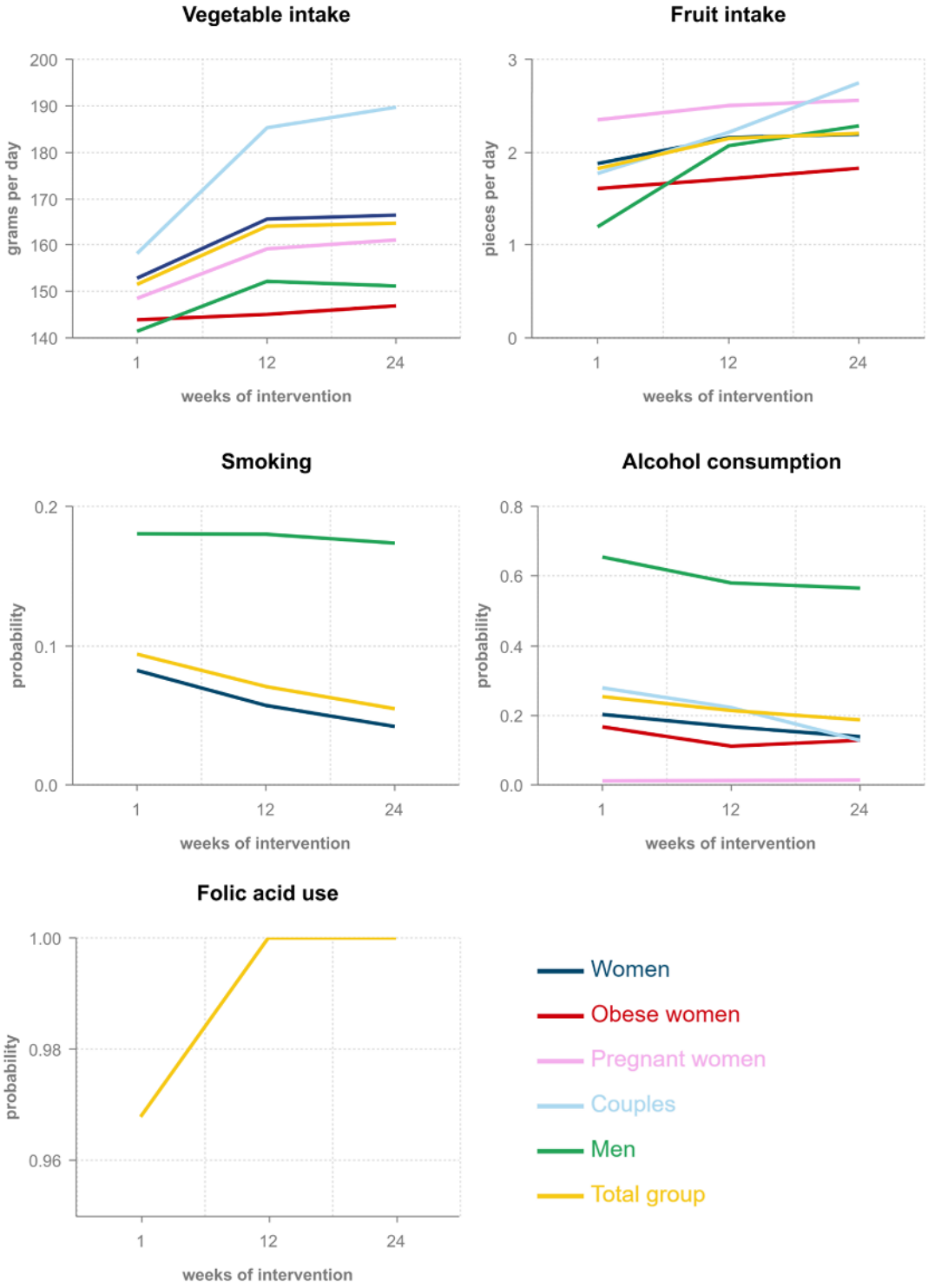
The probability of using tobacco decreased from 0.08 at baseline to 0.06 at 12 weeks (*P*=.004) and 0.04 at 24 weeks (*P*=.012) in the group of female participants. Tobacco use among male participants decreased from a probability of 0.18 at baseline to 0.17 at 24 weeks; however, this effect was not significant (*P*=.84). Additionally, the number of tobacco users was too low to perform further analyses in subgroups.

The probability of consuming alcohol decreased in all subgroups at the end of the coaching program compared with baseline. A significant effect was shown in the total study population and in the subgroup of women, with a decrease in probability from 0.25 to 0.19 (*P*<.001) and from 0.20 to 0.14 (*P*<.001), respectively. Alcohol consumption among male participants

decreased from a probability of 0.95 to 0.82; however, this effect was not significant (P=.08).

The probability of having adequate folic acid supplementation increased from 0.97 at baseline to 1 after 12 weeks (P<.001). This effect was sustained after 24 weeks of online coaching (P<.001).

Figure 2. Improvement in lifestyle behaviors during online coaching in the total group and multiple subgroups.



Discussion

Main Findings

This study confirms previous research reporting a high prevalence of inadequate lifestyle behaviors among women and men in the preconception period as well as during pregnancy [18]. Moreover, findings of this study suggest that a blended lifestyle care approach is an effective method to enhance a healthy lifestyle for couples who are either trying to conceive or who are pregnant. All targeted lifestyle components improved after the 24-week intervention. In particular, vegetable and fruit intake increased evidently among all study participants.

Obesity

The results from women with obesity were less convincing. Vegetable and fruit intake in this group increased only marginally. However, this finding is in contrast to previous results in the study by van Dijk et al that showed positive effects of the Smarter Pregnancy program on lifestyle behaviors in the group of overweight and obese women [18]. In this study, subgroup analyses of lifestyle behaviors at baseline between women with obesity and the rest of the study population (including overweight women) showed no differences. Traditionally, obesity was thought to be a consequence of an individual's poor lifestyle choices resulting in excess energy balance. Recently, the complexity of factors contributing to excess energy balance and weight gain have become more clear, and the following seven factors highly connected to the development of obesity have been identified: individual physiology, social psychology, individual psychology, individual physical activity, physical activity environment, food consumption, and food production [26]. Therefore, an intervention targeting unhealthy lifestyle components among women with obesity should not only focus on food consumption, but also address multiple factors in an integrated manner, including individualized psychological support.

The Black Box of eHealth

The online coaching program Smarter Pregnancy has been proven to be effective for improving lifestyle behaviors [18,20,27]. However, the mechanism by which this intervention affects lifestyle behaviors and the elements that contribute the most are unclear and represent a "black box." Opening this black box is even more relevant for the current blended care approach. Determining the core intervention principles, optimal frequency, and proportion between face-to-face counseling and eHealth could create an even more effective and efficient intervention.

Strengths and Limitations

This is the first study to develop and evaluate a blended personalized lifestyle care approach for the periconception period, designed for both women and their partners. We included a high number of participants (n=511) and achieved a high intervention compliance rate of 78.9%. All lifestyle components improved after the 24-week intervention. However, the clinical relevance of increasing vegetable intake with 20 grams per day and fruit intake with 1 piece per day should be explored further, since no research has been performed to substantiate this effect. Extending follow-up and including data on sustainment of improved parental behavior and on maternal and neonatal pregnancy outcomes could provide insights on the clinical relevance of our findings.

The relatively low number of male participants may reflect their low level of engagement in a healthy lifestyle program. It might even demonstrate the low male involvement in pregnancy itself. However, it could also reflect the low extent to which women actively involve their partners in pregnancy. Future interventions should focus on raising awareness among expecting fathers of their role in reproductive health and pregnancy [28]. Considering that women who participated with their partners showed the strongest improvements in all lifestyle components, increasing the number of men participating in the approach might improve intervention effects. The current blended care approach could be extended with one face-to-face counseling focused on the partner. Another option that would be easy to realize and implement is the development of an informative animation video aimed at men that underlines the shared responsibility and opportunities to promote reproductive health. The reasons for the high effective rates among couples could include mutual support, eating warm meals together, and inspiring one another.

Comparison With Prior Work and Implications for Future Research

A randomized controlled trial (RCT) is often considered to provide the most reliable evidence on the effectiveness of interventions. From this point of view, the absence of a control group (receiving only eHealth) is a possible limitation of this study. However, comparisons with previous research substantiate our findings. The study by Hammiche et al provided couples with tailored face-to-face preconception dietary and lifestyle counseling [17] and reported a compliance rate of only 26%. The survey by van Dijk et al included an even larger number of participants in the preconception or pregnancy period who received only the eHealth intervention Smarter Pregnancy [18]. This intervention study reported a compliance rate of 64.9%. The difference in compliance may be explained by the combination of face-to-face counseling with eHealth, resulting in higher participant engagement. Comparisons of effectiveness between the blended care approach and the survey with only the eHealth intervention should be made with caution, since information on possible confounding factors is missing. Furthermore, in recent years, there has been increasing discussion about the limitation of traditional RCT methodologies for the evaluation of eHealth interventions [29]. Locking down these interventions results in inclusion of possible defects and eliminates the opportunities for quality improvement and adaptation to the changing technological environment, often leading to validation of tools that are outdated by the time trial results are published. Therefore, a design that evaluates the intervention principles (rather than a specific locked-down version of the intervention) is more appropriate. Using the "Trial of Intervention Principle" evaluation method, the design allows for ongoing quality improvement modifications to the behavioral intervention technology based on the core intervention principles, and it continuously improves functionality and maintains technological currency [30].

Further research could be performed on how to continue the steep increase in vegetable and fruit intake after 12 weeks, since our study showed an almost flat curve after this intervention period. The plateau in vegetable and fruit intake might be explained by intervention elements applied in the Smarter Pregnancy program. This intervention particularly focused on increasing external motivation and providing education. It might be effective to add elements that support behavior change in the long term, such as person action plans, goal setting, and maintenance plans [31].

The next step is to further disseminate the approach. Additionally, follow-up of our study participants and their pregnancy outcomes is needed to substantiate the clinical relevance of the results of this study.

Acknowledgments

The authors would like to thank the team of the outpatient clinics of the Department of Obstetrics and Gynecology at the Erasmus University Medical Center in Rotterdam, especially all counselors, and the participating couples for their contributions. This research was funded by ZonMw (VIMP grant) and the Department of Obstetrics and Gynecology of the Erasmus University Medical Center, Rotterdam, the Netherlands.

Conflicts of Interest

None declared.

References

1. Tamakoshi A, Tamakoshi K, Lin Y, Yagyu K, Kikuchi S, JACC Study Group. Healthy lifestyle and preventable death: findings from the Japan Collaborative Cohort (JACC) Study. Prev Med 2009 May;48(5):486-492. [doi: 10.1016/j.ypmed.2009.02.017] [Medline: 19254743]
2. Khaw K, Wareham N, Bingham S, Welch A, Luben R, Day N. Combined impact of health behaviours and mortality in men and women: the EPIC-Norfolk prospective population study. PLoS Med 2008 Jan 8;5(1):e12 [FREE Full text] [doi: 10.1371/journal.pmed.0050012] [Medline: 18184033]
3. Manuel DG, Perez R, Sanmartin C, Taljaard M, Hennessy D, Wilson K, et al. Measuring Burden of Unhealthy Behaviours Using a Multivariable Predictive Approach: Life Expectancy Lost in Canada Attributable to Smoking, Alcohol, Physical Inactivity, and Diet. PLoS Med 2016 Aug;13(8):e1002082 [FREE Full text] [doi: 10.1371/journal.pmed.1002082] [Medline: 27529741]
4. O'Doherty MG, Cairns K, O'Neill V, Lamrock F, Jørgensen T, Brenner H, et al. Effect of major lifestyle risk factors, independent and jointly, on life expectancy with and without cardiovascular disease: results from the Consortium on Health and Ageing Network of Cohorts in Europe and the United States (CHANCES). Eur J Epidemiol 2016 May;31(5):455-468 [FREE Full text] [doi: 10.1007/s10654-015-0112-8] [Medline: 26781655]
5. Li K, Hüsing A, Kaaks R. Lifestyle risk factors and residual life expectancy at age 40: a German cohort study. BMC Med 2014 Apr 07;12:59 [FREE Full text] [doi: 10.1186/1741-7015-12-59] [Medline: 24708705]
6. van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: prospective cohort study in US women. BMJ 2008 Sep 16;337:a1440 [FREE Full text] [doi: 10.1136/bmj.a1440] [Medline: 18796495]
7. McKenzie F, Ellison-Loschmann L, Jeffreys M, Firestone R, Pearce N, Romieu I. Healthy lifestyle and risk of breast cancer for indigenous and non-indigenous women in New Zealand: a case control study. BMC Cancer 2014 Jan 10;14:12 [FREE Full text] [doi: 10.1186/1471-2407-14-12] [Medline: 24410858]
8. Chiuve SE, McCullough ML, Sacks FM, Rimm EB. Healthy lifestyle factors in the primary prevention of coronary heart disease among men: benefits among users and nonusers of lipid-lowering and antihypertensive medications. Circulation 2006 Jul 11;114(2):160-167. [doi: 10.1161/CIRCULATIONAHA.106.621417] [Medline: 16818808]
9. Nechuta SJ, Shu X, Li H, Yang G, Xiang Y, Cai H, et al. Combined impact of lifestyle-related factors on total and cause-specific mortality among Chinese women: prospective cohort study. PLoS Med 2010 Sep 14;7(9):e1000339 [FREE Full text] [doi: 10.1371/journal.pmed.1000339] [Medline: 20856900]
10. Fransen HP, May AM, Beulens JW, Struijk EA, de Wit GA, Boer JM, et al. Association between lifestyle factors and quality-adjusted life years in the EPIC-NL cohort. PLoS One 2014;9(11):e111480 [FREE Full text] [doi: 10.1371/journal.pone.0111480] [Medline: 25369457]
11. Healthy diet. World Health Organization. 2020 Apr 29. URL: https://www.who.int/news-room/fact-sheets/detail/healthy-diet [accessed 2020-09-18]
12. Steegers-Theunissen RP, Twigt J, Pestinger V, Sinclair KD. The periconceptional period, reproduction and long-term health of offspring: the importance of one-carbon metabolism. Hum Reprod Update 2013;19(6):640-655. [doi: 10.1093/humupd/dmt041] [Medline: 23959022]
13. Godfrey KM, Gluckman PD, Hanson MA. Developmental origins of metabolic disease: life course and intergenerational perspectives. Trends Endocrinol Metab 2010 Apr;21(4):199-205. [doi: 10.1016/j.tem.2009.12.008] [Medline: 20080045]

14. Barker M, Dombrowski SU, Colbourn T, Fall CH, Kriznik NM, Lawrence WT, et al. Intervention strategies to improve nutrition and health behaviours before conception. The Lancet 2018 May 05;391(10132):1853-1864 [FREE Full text] [doi: 10.1016/S0140-6736(18)30313-1] [Medline: 29673875]
15. Bogaerts A, Ameye L, Bijlholt M, Amuli K, Heynickx D, Devlieger R. INTER-ACT: prevention of pregnancy complications through an e-health driven interpregnancy lifestyle intervention - study protocol of a multicentre randomised controlled trial. BMC Pregnancy Childbirth 2017 May 26;17(1):154 [FREE Full text] [doi: 10.1186/s12884-017-1336-2] [Medline: 28549455]
16. Oteng-Ntim E, Varma R, Croker H, Poston L, Doyle P. Lifestyle interventions for overweight and obese pregnant women to improve pregnancy outcome: systematic review and meta-analysis. BMC Med 2012 May 10;10:47 [FREE Full text] [doi: 10.1186/1741-7015-10-47] [Medline: 22574949]
17. Hammiche F, Laven JS, van Mil N, de Cock M, de Vries JH, Lindemans J, et al. Tailored preconceptional dietary and lifestyle counseling in a tertiary outpatient clinic in The Netherlands. Hum Reprod 2011 Sep;26(9):2432-2441. [doi: 10.1093/humrep/der225] [Medline: 21752799]
18. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth Platform for Pregnancy on Nutrition and Lifestyle of the Reproductive Population: A Survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
19. van Dijk MR, Oostingh EC, Koster MP, Willemsen SP, Laven JS, Steegers-Theunissen RP. The use of the mHealth program Smarter Pregnancy in preconception care: rationale, study design and data collection of a randomized controlled trial. BMC Pregnancy Childbirth 2017 Jan 26;17(1):46 [FREE Full text] [doi: 10.1186/s12884-017-1228-5] [Medline: 28125970]
20. Gootjes DV, van Dijk MR, Koster MP, Willemsen SP, Steegers EA, Steegers-Theunissen RP. Neighborhood Deprivation and the Effectiveness of Mobile Health Coaching to Improve Periconceptional Nutrition and Lifestyle in Women: Survey in a Large Urban Municipality in the Netherlands. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11664 [FREE Full text] [doi: 10.2196/11664] [Medline: 30973345]
21. Boots LM, de Vugt ME, Kempen GI, Verhey FR. Effectiveness of a Blended Care Self-Management Program for Caregivers of People With Early-Stage Dementia (Partner in Balance): Randomized Controlled Trial. J Med Internet Res 2018 Jul 13;20(7):e10017 [FREE Full text] [doi: 10.2196/10017] [Medline: 30006327]
22. Kouwenhoven-Pasmooij TA, Djikanovic B, Robroek SJ, Helmhout P, Burdorf A, Hunink MG. Design and baseline characteristics of the PerfectFit study: a multicenter cluster-randomized trial of a lifestyle intervention in employees with increased cardiovascular risk. BMC Public Health 2015 Jul 28;15:715 [FREE Full text] [doi: 10.1186/s12889-015-2059-9] [Medline: 26215589]
23. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
24. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
25. Smarter Pregnancy. URL: https://www.smarterpregnancy.co.uk/ [accessed 2020-09-14]
26. Vandenbroeck P, Goossens J, Clemens M. Foresight: Tackling Obesities: Future Choices – Building the Obesity System Map. United Kingdom: UK Government; 2007.
27. van Dijk MR, Koster MP, Willemsen SP, Huijgen NA, Laven JS, Steegers-Theunissen RP. Healthy preconception nutrition and lifestyle using personalized mobile health coaching is associated with enhanced pregnancy chance. Reprod Biomed Online 2017 Oct;35(4):453-460. [doi: 10.1016/j.rbmo.2017.06.014] [Medline: 28688924]
28. Pachauri S. Male involvement in reproductive health care. J Indian Med Assoc 2001 Mar;99(3):138-141. [Medline: 11478756]
29. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating Digital Health Interventions: Key Questions and Approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
30. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies. J Med Internet Res 2015 Jul 08;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
31. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key Components in eHealth Interventions Combining Self-Tracking and Persuasive eCoaching to Promote a Healthier Lifestyle: A Scoping Review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]

Abbreviations

BMI: body mass index
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 15.04.20; peer-reviewed by S Prior, M Bijlholt; comments to author 12.06.20; revised version received 16.07.20; accepted 26.07.20; published 30.09.20.

Please cite as:
van der Windt M, van der Kleij RM, Snoek KM, Willemsen SP, Dykgraaf RHM, Laven JSE, Schoenmakers S, Steegers-Theunissen RPM
Impact of a Blended Periconception Lifestyle Care Approach on Lifestyle Behaviors: Before-and-After Study
J Med Internet Res 2020;22(9):e19378
URL: http://www.jmir.org/2020/9/e19378/
doi:10.2196/19378
PMID:32996885

Original Paper

Intention to use Medical Apps Among Older Adults in the Netherlands: Cross-Sectional Study

Marjan Askari[1], BSc, MSc, PhD; Nicky Sabine Klaver[1], BSc, MSc; Thimon Johannes van Gestel[1], BSc, MSc; Joris van de Klundert[1,2], BSc, MSc, PhD

[1]Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, Netherlands

[2]Prince Mohammad Bin Salman College of Business & Entrepreneurship, King Abdullah Economic City, Saudi Arabia

Corresponding Author:
Marjan Askari, BSc, MSc, PhD
Erasmus School of Health Policy & Management
Erasmus University
PO Box 1738
Rotterdam, 3000 DR
Netherlands
Phone: 31 104088641
Email: askari@eshpm.eur.nl

Abstract

Background: The increasing health service demand driven by the aging of the global population calls for the development of modes of health service delivery that are less human resource–intensive. Electronic health (eHealth) and medical apps are expected to play an important role in this development. Although evidence shows mobile medical apps might be effective in improving the care, self-management, self-efficacy, health-related behavior, and medication adherence of older adults, little is known about older adults' intention to use these technologies when needed, or the factors influencing this intention.

Objective: The objective of this study was to investigate the relationship of technology acceptance factors and intention to use mobile medical apps among community-dwelling older adults.

Methods: Data was collected using questionnaires. The factors selected from the literature have been validated using Cronbach α and tested for significance using logistic regressions.

Results: Almost half (49.7%) of the included older adults reported no intention to use medical apps. Adjusted logistic regression analysis per factor showed that the factors Attitude toward use (odds ratio [OR] 8.50), Perceived usefulness (OR 5.25), Perceived ease of use (OR 4.22), Service availability (OR 3.46), Sense of control (OR 3.40), Self-perceived effectiveness (OR 2.69), Facilities (OR 2.45), Personal innovativeness (OR 2.08), Social relationships (OR 1.79), Subjective norm (OR 1.48), and Feelings of anxiety (OR 0.62) significantly influenced the intention to use mobile medical apps among older adults, whereas the factor Finance (OR 0.98) did not. When considered together, a controlled multivariate logistic regression yielded high explained variances of 0.542 (Cox-Snell R^2) and 0.728 (Nagelkerke R^2).

Conclusions: The high odds ratios and explained variance indicate that the factors associated with the intention to use medical apps are largely understood and the most important factors have been identified. To advance the evidence base, experimental controlled research should investigate the causality between the factors, intention to use, and actual use. For this purpose, our evidence suggests that policies designed to improve Attitude toward use appear most effective, followed by policies addressing Perceived usefulness, Perceived ease of use, Service availability, and Sense of control.

(J Med Internet Res 2020;22(9):e18080) doi:10.2196/18080

KEYWORDS

Senior Technology Acceptance Model; intention to use; elderly; older adults; medical apps; mHealth; adoption

Introduction

The number of adults over 65 years of age worldwide is expected to triple from 562 million in 2012 to 1.6 billion in 2050 and comprise 16.7% of the growing global population, up from the current 8.0% [1]. This aging of society is caused by increased longevity, decreased fertility, and the aging of the "baby boom" generation [1-3]. Older adults tend to make more

use of health services compared to other age categories [4]. In the Netherlands, for instance, older adults form about 20% of the total population while accounting for approximately 80% of total health care expenditure [4]. The abovementioned three-fold global increase in the number of older adults may therefore indeed be expected to significantly increase global health services utilization and expenditure, exposing governments and societies to a wide range of social and economic challenges [5].

An important consequence of this aging-driven increase in health service needs is the additional demand for human resources it will create. For instance, the Dutch health care sector is expected to face a shortage of between 100,000 and 125,000 health care professionals by 2022 [6]. The largest shortages will involve nurses, geriatric specialists, and psychiatrists. Instead of resolving the increased health service needs by expanding the factor of production labor (ie, increasing human resources), policy makers are considering increasing human resource productivity, as facilitated by the factor of production technology [7]. Technology may, for instance, enable older adults to live in their own homes more independently and for a longer period of time [4].

Electronic health (eHealth), defined by the World Health Organization as the use of information and communication technologies for health, is widely considered to be a promising technological advancement to address the challenges presented above [8]. Its potential for health service delivery innovation and for service expansion without increasing human resource capacity is viewed as essential to addressing the increasing needs of the aging population with limited extra burden on the health care system [9].

Mobile medical apps provide an easily and widely accessible form of eHealth. Medical apps are defined as apps that run on electronic consumer devices such as smartphones and tablets [10-12]. These apps can, for example, be used to gather information about one's health, disease, or condition; help monitor health; or support users in activities concerning their health [13-15]. Medical apps have been shown to improve the care, self-management, and self-efficacy of older adults, as well as promote better behavior and medication adherence [16-19].

While medical apps have been shown to be effective tools in supporting or substituting conventional health service delivery, evidence reveals that older adults tend to be more resistant to accepting new information technology apps [20,21] and to be apprehensive toward novel technologies [22]. In view of the aging-related challenges outlined above, and the contribution medical apps may have in resolving them, it is imperative to better understand medical app adoption by older adults. The widely accepted and validated Theory of Reasoned Action (TRA) and Theory of Planned Behavior (TPB), as well as more specific theoretic models introduced below, posit that medical app adoption is primarily determined by the intention to use such apps [23]. Our research aim is to advance the evidence base on factors influencing the intention to use medical apps among older adults.

The Technology Acceptance Model (TAM) derived from the TRA and TPB posits that Perceived ease of use, Perceived usefulness, and Social influence are the main factors of technology adoption, via the Intention to use factor [24,25]. Subsequent research has revealed a variety of additional factors influencing the intended and actual use of technology, such as Subjective norm, Voluntariness, Image (TAM2), Self-efficacy, Social norms, Trust, and Compatibility [16,25-29].

More recently, several studies have been conducted to advance understanding of the intention to use and actual use of medical technology [28,30-32] in general and by older adults in particular [23,33-35]. These studies, among which are studies specifically addressing intention to use medical apps, had quite small sample sizes [31,35] and/or took a qualitative approach [23,30,36]. Hence, while these studies have advanced toward a more specific Senior Technology Acceptance Model (STAM) [32], evidence on the factors determining intention to use medical apps by older adults is quite limited. Therefore, we set out to assess the validity and significance of factors determining intention to use medical apps in a quantitative study involving a large sample of older adults.

Methods

Study Design and Data Collection

A cross-sectional study was designed to study the relationship between the intention to use medical apps and proposed factors derived from literature. For this purpose, we developed a questionnaire and administered it both digitally as well as on paper to facilitate the inclusion of older adults with limited computer experience. Assistance and explanations were given to participants who needed help filling out the questionnaire, via telephone, email, or personal assistance when requested. The data was collected by 4 data assistants from November 2018 to June 2019 in cooperation with different types of organizations, such as living facilities and leisure activity clubs for older adults, general practitioners, and a hospital. To further strengthen data triangulation, we used online questionnaires, which were distributed across the Netherlands in cooperation with health service provider organizations and wellness organizations via different online channels and mailing lists. The reporting of the online questionnaire follows the CHERRIES checklist (Checklist for Reporting Results of Internet E-Surveys), which can be found in Multimedia Appendix 1 [37].

The inclusion criteria for participants were as follows: the participant is 65 years of age or older; the participant does not have cognitive impairments (as assessed by caregivers or those who distributed the questionnaires on paper); and the participant lives alone or with other people in a regular or senior living facility, the rationale being that older adults living in care facilities are already receiving care and therefore have little to no need for medical apps that promote time- and location-independent care.

At the start of the questionnaire, the following information was given in writing: the purpose of the project; information and instructions regarding the questionnaire; the expected duration of the survey, and the names of the main researchers. In addition, information about data management and privacy of the

participants was provided to them. Before the participants filled out the questionnaire, an informed consent form was signed to give permission to use the data for research purposes. To alleviate potential concerns regarding privacy for the paper-based version, an envelope was provided to participants to ensure no one other than the data assistants would see the completed questionnaire. Data assistants entered the completed questionnaires into a SPSS database (IBM Corp) and pseudonymized the data to ensure anonymity.

Senior Technology Acceptance Model

To analyze the association between acceptance factors and intention to use medical apps, an adapted and expanded version of the TAM for older adults is used. The TAM suggests that the Perceived usefulness and Perceived ease of use are key factors in explaining the intention to use, and subsequent use of a technological system [24,32]. We also included a number of factors from the STAM [32] and TAM2 [25], from which STAM is derived. In addition, we included specific acceptance factors for the use of medical apps among older adults from the literature, as indicated in Table 1, for a total of 12 factors. In addition, Table 1 shows the description of each factor and an example of a statement included in the questionnaire. For each factor, we included 1 to 4 statements to measure different aspects and strengths of the factor. These statements were answered using a 5-point Likert-scale (1=completely disagree, 2=disagree, 3=neutral, 4=agree, 5=completely agree). We computed a factor score by calculating the average score of all statements per factor.

The internal consistency of the items within the factors was investigated using Cronbach α [38,39]. The Cronbach α is expressed as a number between 0 and 1. The higher this number is, the lower the error variance is within the measuring instrument. The Cronbach α was acceptable if above 0.7 [40]. Items that, when removed, increased the Cronbach α of the acceptance factor by 0.1 or more were excluded from the acceptance factor and further analysis.

Table 1. Description of the included factors with an example statement and literature references.

Factor (number of statements)	Operational definition	Example of a statement	References
Perceived usefulness (3)	The extent to which a person believes that using the medical app will improve his or her quality of life	Using medical apps for remote health care would make my life easier.	[24,29,32,34]
Perceived ease of use (4)	The extent to which a person believes that using medical apps will be free of effort	It is easy to use medical apps for remote health care.	[24,25,29,32,34]
Attitude toward use (4)	An individual's positive or negative feelings or appraisal about using medical apps	Using medical apps for remote care would be a good idea.	[24,29,32,34,41-43]
Subjective norm (3)	The person's perception that most people who are important to them think they should or should not use medical apps	People who are important to me think that I should use medical apps.	[24,25,29,41-43]
Sense of control (2)	The perceptions of internal and external constraints on using medical apps	Using medical apps for remote health care is entirely within my control.	[29,42-44]
Feelings of anxiety (2)	An individual's apprehension when he or she is faced with the possibility of using technology	I feel anxious to start using medical apps for remote health care.	[29,32]
Personal innovativeness (4)	Personal tendency to innovate, or introduce something new or different	In general, I do not hesitate to try out new information technology.	[20]
Social relationships (3)	An individual's satisfaction with personal relationships and support from friends and family	I am satisfied with my personal relationships.	[32,45,46]
Self-perceived effectiveness (2)	Judgment of one's ability to use medical apps to accomplish a particular job or task	I could perform a task on a medical app if I have just the instruction manual for assistance.	[29,32]
Service availability (3)	The obtainability and accessibility of medical apps	Medical apps for remote health care are always available whenever I need them.	[47]
Facilities (2)	Objective factors in the environment that can make technology usage easy. Included indicators are basic knowledge and available help	I have the knowledge needed to use medical apps.	[28,29,32]
Finance (1)	Having the financial resources to make technology usage easy	My financial situation stops me from using medical apps.	[28,29,32]

Statistical Analyses

Descriptive statistics were used to analyze the composition of the research sample. For continuous variables, the mean and standard deviation were calculated; for categorical variables, percentages were used. The Assessment of Activities of Daily Living, Self-Care, and Independence (ADL) [48-50] and the Identification of Seniors at Risk – Primary Care (ISAR-PC) questionnaires [51,52] were used and scores were calculated as follows. ADL consists of 16 items and for every item, the participant answered whether they needed help doing the mentioned activities (such as showering, dressing, or walking). The ADL score was computed by counting the number of activities for which no help was needed [50]. The ISAR-PC was used to measure the increased risk of functional decline and consisted of three elements. The first two elements, household help required and repeated aptness to forget, are yes/no questions where answering with yes increased the ISAR-PC score by 2.5 and 2, respectively. The last element is an ordinal scale of 3 age groups: 65 to 74 years, 75 to 84 years, and 85 years or older. For individuals in the first group, the ISAR-PC score was increased by 0; it was increased by 1.5 for the second group and by 3 for the last group. An ISAR-PC score greater or equal to 2 shows an increased risk of functional decline [51].

An individual's living situation was a categorical variable, consisting of 4 options: living independently alone, living independently with others, living in a senior living facility alone, or living in a senior living facility with others. Previous internet experience was a binary question. Perceived quality of life involved respondents rating their quality of life between 0 and 100. General health consisted of 5 categories: excellent, very good, good, fair, and poor. Health care utilization was the sum of the number of times a participant had visited the GP, the emergency department, and the hospital in the last 6 months.

Univariate and Multivariate Logistic Regression

The calculated acceptance factor score served as input for the univariate and multivariate logistic regression analysis to examine the relationship between each of the acceptance factors as independent variables and intention to use medical apps as a dependent variable. Age, sex, and education served as control variables and were always included in the multivariate logistic regression model. Other candidate controls were marital status, living situation, ISAR-PC score, ADL score, previous internet experience, perceived quality of life, general health, and health care utilization. The controls that were measured on a continuous scale were tested on multicollinearity and one of two items that had an absolute correlation larger than 0.8 was removed, using expert opinion. The controls were iteratively added to the multivariate logistic regression model. Those that changed the odds ratio (OR) of any independent variable (the acceptance factor) by at least 10% when added to the multivariate logistic regression model were retained and incorporated in the final model [39,53]. Finally, we reported the measures of explained variance (the Cox-Snell R^2 and the Nagelkerke R^2) for the multivariate logistic regression models.

Validity and Reliability

To increase the internal validity, we included common, standardized, and validated instruments such as ADL and ISAR-PC [25,29,32,48,49,52]. Moreover, the questionnaire has been validated with the assistance of 4 older adults and several experts, including a geriatric nurse and 2 eHealth experts. The questionnaire is available on request. The database was checked for completeness and input errors, where a sample of paper questionnaires was compared to the database counterpart to check if they were identical.

To increase the external validity, we collected data for 40% of the respondents on paper, thus ensuring all eligible older adults were included. Moreover, data collection took place in several different geographical locations within The Netherlands. Lastly, Cronbach α was calculated for each factor to test the reliability. The study was approved by the medical ethical committee of Erasmus Medical Center (number MEC-2018-120).

Results

Population Characteristics

Our data set consisted of 364 older adults with an average age of 75 years (SD 7 years). Overall, 42.6% (n=155) of the 364 participants were male. Although 85.2% (n=310) of participants had experience using the internet, only 15.9% (n=58) had experience with medical apps. Despite the low proportion of participants that had ever used a medical app prior to filling out the questionnaire, more than 50.3% (n=183) stated an intent to use medical apps. Table 2 provides an overview of population characteristics.

A Cronbach α score was calculated for each of the acceptance factors to validate the internal consistency of the items within that factor [38]. The Cronbach α scores of the acceptance factors are shown in Table 3. All factors have an acceptable value of above 0.7 and none of the statements, when deleted, increased the Cronbach α by 0.1 or more [40].

Table 2. Baseline characteristics of the study cohort.

Characteristics	Participants (n=364)
Age (years), mean (SD)	74.9 (7.1)
Sex (male), n (%)	155 (42.6)
Education, n (%)	
No education	9 (2.5)
Lower education	57 (15.7)
Intermediate education	160 (44.0)
Higher education	125 (34.3)
Marital status, n (%)	
Married	190 (52.2)
Divorced	51 (14.0)
Widowed	89 (24.5)
Single	22 (6.0)
Living with partner	8 (2.2)
Living arrangement, n (%)	
Living independently, alone	129 (35.4)
Living independently, with others	161 (44.2)
Senior living facility, alone	34 (9.3)
Senior living facility, with others	35 (9.6)
Identification of Seniors at Risk – Primary Care questionnaire score, mean (SD)	1.4 (1.7)
Assessment of Activities of Daily Living, Self-Care, and Independence score, mean (SD)	14.6 (2.3)
Quality of life, mean (SD)[a]	7.6 (7.0)
Prior experience with internet, n (%)	310 (85.2)
Prior experience with medical apps, n (%)	58 (15.9)
Intention to use, n (%)	183 (50.3)

[a]This measure is scored on a scale from 0 to 10.

Table 3. Cronbach α of the technology acceptance factors.

Factors[a]	Cronbach α
Perceived usefulness (n=3)	0.922
Perceived ease of use (n=4)	0.950
Attitude toward use (n=4)	0.955
Subjective norm (n=3)	0.974
Sense of control (n=2)	0.890
Intention to use (n=3)	0.969
Feelings of anxiety (n=2)	0.913
Personal innovativeness (n=4)	0.950
Social relationships (n=3)	0.716
Self-perceived effectiveness (n=2)	0.742
Service availability (n=3)	0.923
Facilities (n=2)	0.746
Finance (n=1)	N/A

[a]The n value refers to the number of statements within a construct.

Univariate and Multivariate Analysis

To analyze the relationship between each acceptance factor and intention to use medical apps, univariate and multivariate analyses were performed. The results of the univariate logistic regression analysis showed that all factors were significantly associated with Intention to use medical apps, except for Finance. As expected, these results showed all factors to be positively associated with Intention to use, except for the factor Feelings of anxiety.

The multivariate logistic regression analyses largely confirmed the results of the univariate analysis. None of the controls displayed multicollinearity and therefore none of the variables had to be excluded in the multivariate logistic regression. Controlling for age, sex, and education modestly reduces or increases the ORs. The other candidate controls mainly impact the original TAM key factor Attitude toward use. Table 4 presents these results.

The two right-most columns of Table 4 present two measures of explained variance (the Cox-Snell R^2 and the Nagelkerke R^2) for the models, which control for age, sex, education, and controls increasing the OR by more than 10%. When including all factors and none of the controls, the explained variances are 0.486 (Cox-Snell R^2) and 0.651 (Nagelkerke R^2). Adding the standard controls age, sex, and education did not have much of an impact on the explained variance. When adding all controls to the model, which includes all factors, the explained variances increase to 0.542 (Cox-Snell R^2) and 0.728 (Nagelkerke R^2), respectively.

Table 4. Association between acceptance factors and intention to use medical apps.

Factors	Univariate OR[a] (95% CI)	*P* value	Multivariate OR (95% CI)[b]	*P* value	Multivariate OR (95% CI)[c]	*P* value	Included controls	Cox-Snell R^2	Nagelkerke R^2
Perceived usefulness	5.94 (3.88-9.08)	<.001	5.25 (3.41-8.07)	<.001	—	—	—	.340	.456
Perceived ease of use	4.43 (3.01-6.52)	<.001	4.22 (2.78-6.40)	<.001	4.71 (3.02-7.36)	<.001	ISAR-PC[d]	.298	.400
Attitude toward use	9.19 (5.52-15.30)	<.001	8.50 (5.03-14.38)	<.001	11.24 (6.08-20.79)	<.001	ISAR-PC, ADL[e], marital state, health care use	.440	.591
Subjective norm	1.47 (1.17-1.84)	.001	1.48 (1.15-1.90)	.002	—	—	—	.129	.173
Sense of control	3.59 (2.64-4.87)	<.001	3.40 (2.45-4.72)	<.001	—	—	—	.293	.394
Feelings of anxiety	0.56 (0.44-0.70)	<.001	0.62 (0.47-0.81)	.001	—	—	—	.135	.181
Personal innovativeness	2.38 (1.85-3.06)	<.001	2.08 (1.58-2.73)	<.001	—	—	—	.182	.243
Social relationships	1.76 (1.21-2.56)	.003	1.79 (1.18-2.71)	.006	—	—	—	.131	.175
Self-perceived effectiveness	2.84 (2.10-3.82)	<.001	2.69 (1.93-3.76)	<.001	3.05 (2.14-4.36)	<.001	Living situation	.240	.321
Service availability	3.71 (2.61-5.26)	<.001	3.46 (2.37-5.06)	<.001	—	—	—	.245	.329
Facilities	2.70 (2.03-3.59)	<.001	2.45 (1.78-3.35)	<.001	—	—	—	.178	.239
Finance	0.93 (0.74-1.16)	.51	0.98 (0.76-1.28)	.90	—	—	—	.097	.129

[a]OR: odds ratio.

[b]Adjusted for age, sex, and education.

[c]Adjusted for age, sex, and education and all controls that increase the OR by at least 10%.

[d]ISAR-PC: Identification of Seniors at Risk – Primary Care.

[e]ADL: Assessment of Activities of Daily Living, Self-Care, and Independence.

Discussion

Principal Results

In this study, we aimed to provide the first robust quantitative evidence on the intention to use medical apps among community-dwelling older adults and the factors identified in the literature that could assist in determining intention to use. We found that almost half of the respondents (49.7%, n=181) had no intention to use medical apps. This first descriptive finding is relevant due to the important contribution medical apps are anticipated to make in keeping health services affordable as the number of older people worldwide triples, resulting in an increase in health service demand. Hence, it is important to understand the factors determining the intention to use medical apps. Here, we first synthesize the evidence as identified in our study results.

All but one of the proposed factors were very significantly related to Intention to use, with *P* values <.01. More specifically, the multivariate logistic regression analyses showed that the following acceptance factors are significantly related to the intention to use medical apps in this population: Perceived usefulness, Perceived ease of use, Attitude toward use, Subjective norm, Sense of control, Feelings of anxiety, Personal innovativeness, Social relationships, Self-perceived effectiveness, Service availability, and Facilities. All these factors were positively associated with Intention to use, except for the factor Feelings of anxiety, which was negatively associated with Intention to use. Finance was the only identified factor not significantly related to Intention to use. As our results showed, having feelings of anxiety about using new technology may negatively affect the intention to use medical apps. This might be caused by factors such as a lack of self-efficacy, a desire for a greater sense of control, privacy issues, or a lack of trust. Future studies are needed to study the underlying causal factors.

The factor Attitude toward use stood out with an OR of 8.50 in the multivariate model (11.24 when controlled), indicating that a positive Attitude toward use roughly increases the Intention to use ten-fold. The other two original TAM factors, Perceived usefulness and Perceived ease of use, had the second- and third-highest ORs, albeit much lower than that for Attitude toward use. Sense of control is another classic factor [29,42-44] that has an OR above 3; the same is true for the lesser known factor Service availability [47]. The lack of significance of the factor Finance, originally included in STAM [32] and confirmed in subsequent studies [28,29], might be explained by the relatively low cost of medical apps compared to other technologies and the relatively generous mandatory health insurance coverage in The Netherlands.

The variables ISAR-PC, ADL, marital status, and recent health care use significantly impacted the Attitude toward use. Otherwise, the added controls had little (<10%) effect on ORs, with only two exceptions (Table 4). This indicates that, perhaps with the exception of Attitude toward use, measures to improve factor scores do not need to distinguish among subpopulations but can target the entire population of older adults. Standing out for its high OR, the factor Attitude toward use appears to be a prime candidate for interventions to increase intention to use among older adults by making attitudes more favorable.

Comparison with Prior Work

All the technology acceptance factors considered in this study are taken from the literature and have been positively associated with intention to use various forms of (medical) technology in other contexts, sometimes specifically for older adults [23,28,30-35]. Our research confirms the validity of these factors in explaining older Dutch adults' intention to use medical apps.

Our findings are akin to the findings that Cajita et al [23] obtained for older adults with heart failure. They found Perceived usefulness, Perceived ease of use, Subjective norm, Social relationships, and Social influence to be significantly related to intention to use mobile technology [23]. These similarities arise despite the differences in patient populations, which include individuals from different countries, who have different morbidity and are of different sizes. Moreover, our findings strongly confirm the three original TAM factors as the main factors driving intention to use medical apps among older adults [24]. This robust finding is notable as these factors date back more than 30 years and technology has advanced considerably. Medical apps did not exist when TAM was developed. It is possible that the aging population has simply carried the factors associated with their generation into the future since 1989. In addition, our analyses confirm all but one of the factors of the recent STAM [32]. Altogether, these similarities suggest that our main findings have validity outside the Netherlands and beyond a near-term horizon.

Limitations

A first limitation of our study may be the length of the questionnaire used. Although steps were taken to minimize the impact of the length of the questionnaire, such as printing out the questionnaire so participants could take breaks or sitting with the participants while they filled out the questionnaire, some participants still showed signs of response fatigue [54]. To minimize the impact of response fatigue, participants could take breaks and save their answers online to continue later on. Second, we noticed that some of the participants, especially those aged >75 years, struggled to understand the use and utility of medical apps. To address this situation, the questionnaires and interviewers provided additional explanations about medical apps. Due to the cross-sectional design of this study, no claims of causality can be made [55] and the results might suffer from self-report bias [56]. Lastly, while the data was collected from a variety of contexts in The Netherlands, we cannot claim validity in other countries, where for instance the factor Finance may be of larger significance.

Recommendations and Future Research

The main contribution of this study is to provide the first large-scale quantitative evidence of the relationships between the proposed acceptance factors and the intention to use medical apps among older adults in the Netherlands. As noted, due to the research design, we cannot confirm causality among the identified relationships. Hence, a first recommendation is to advance research on the most significant factors using controlled experiments rather than large-scale cohort studies to confirm or refute any potential causality of the relationships found. Such studies may target older adults who do not yet intend to use medical apps; this is the most urgent group to include in such initiatives in view of the challenges related to the aging of societies. Moreover, even though behavioral intention has been shown to predict actual technology adoption [57], such experiments might study actual technology use, rather than intention to use. In addition, we recommend qualitative research to advance understanding regarding the nature of the relationships between the most significant factors and the intention to use. Meanwhile, policy designed at improving Attitude toward use appear most effective, accompanied by policies addressing Perceived usefulness and Perceived ease of use, as well as Service availability and Sense of control.

Acknowledgments

MA designed the research project and developed the questionnaire. NSK, MA, and TG collected the data with the help of data assistants. NSK and TG performed the analyses under the supervision of MA and JvK. All authors interpreted the results. MA wrote the initial version of the manuscript. All authors revised the paper critically. We would like to thank all the experts and older adults who helped us validate our questionnaire, as well as the participants and data assistants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CHERRIES checklist.
[DOCX File , 23 KB - jmir_v22i9e18080_app1.docx]

References

1. He W, Goodkind D, Kowal P. An Aging World: 2015. International Population Reports. Washington, DC: U.S. Government Publishing Office; 2016. URL: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p95-16-1.pdf [accessed 2019-09-17]
2. United Nations, Department of Economic and Social Affairs, Population Division. World Population Ageing 2017 (ST/ESA/SER.A/408). 2017. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf [accessed 2019-10-06]
3. Mathers CD, Stevens GA, Boerma T, White RA, Tobias MI. Causes of international increases in older age life expectancy. The Lancet 2015 Feb;385(9967):540-548. [doi: 10.1016/s0140-6736(14)60569-9]
4. Nederlandse Zorgautoriteit. Monitor - Zorg voor ouderen 2018. 2018. URL: https://www.rijksoverheid.nl/documenten/rapporten/2018/04/19/monitor-zorg-voor-ouderen-2018 [accessed 2019-10-06]
5. Jeurissen P. Beheersing zorgkosten. Tijds gezondheidswetenschappen 2014 Apr 6;92(3):99-100. [doi: 10.1007/s12508-014-0039-6]
6. Ministerie van Volksgezondheid, Welzijn en Sport. Actieprogramma Werken in de Zorg. 2018 Mar. URL: https://www.rijksoverheid.nl/documenten/jaarplannen/2018/03/14/actieprogramma-werken-in-de-zorg [accessed 2019-10-06]
7. Scholte R, Kok L. Economische groei en de vraag naar zorg: macro-economische tegenkrachten in het ramingsmodel van het Capaciteitsorgaan. Amsterdam: SEO Economisch Onderzoek; 2013 Jun. URL: https://capaciteitsorgaan.nl/app/uploads/2016/02/SEO-Economische-groei-en-de-vraag-naar-zorg.pdf [accessed 2019-10-06]
8. Global Observatory for eHealth. World Health Organization. 2017. URL: https://www.who.int/goe/en/ [accessed 2019-10-06]
9. Wouters M, Swinkels I, van Lettow B, de Jong J, Sinnige J, Brabers A, et al. E-health in verschillende snelheden. eHealth-monitor 2018. Den Haag and Utrecht: Nictiz and het Nivel; 2018 Nov. URL: https://www.ehealth-monitor.nl/wp-content/themes/nictiz/assets/pdf/ehealth-monitor-2018.pdf [accessed 2019-10-06]
10. Global diffusion of eHealth: making universal health coverage achievable. Report of the third global survey on eHealth. Geneva: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf;jsessionid=C9C5ECC47AC7E602C18FDF7DE10AE136?sequence=1 [accessed 2019-01-08]
11. Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage. WHO Global Observatory for eHealth. Geneva; 2016. URL: https://www.who.int/publications/i/item/atlas-of-ehealth-country-profiles-the-use-of-ehealth-in-support-of-universal-health-coverage [accessed 2019-01-08]
12. Adibi S. Mobile Health: A Technology Road Map. Switzerland: Springer International Publishing; 2015.
13. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
14. Oliveira JGRD, Askari M, Silva Junior GBD, Freitas Filho RAD, Vasconcelos Filho JE. Renal Health: An Innovative Application to Increase Adherence to Treatment Through Self-monitoring for Patients With CKD and Provide Information for the General Population. Kidney Int Rep 2019 Apr;4(4):609-613 [FREE Full text] [doi: 10.1016/j.ekir.2019.01.008] [Medline: 30993237]
15. Meulendijk MC, Spruit MR, Willeboordse F, Numans ME, Brinkkemper S, Knol W, et al. Efficiency of Clinical Decision Support Systems Improves with Experience. J Med Syst 2016 Apr 20;40(4):76 [FREE Full text] [doi: 10.1007/s10916-015-0423-z] [Medline: 26791992]
16. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. mHealth 2017 Nov 27;3:51-51. [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
17. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. Electron Markets 2012 Dec 11;23(1):49-61. [doi: 10.1007/s12525-012-0112-4]

18. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5(S5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
19. mHealth Technologies: Applications to Benefit Older Adults.: Center for Technology and Aging; 2011. URL: https://2mjt5a2emh374130j5vkxw9g-wpengine.netdna-ssl.com/wp-content/uploads/2011/01/ghcah59qtuhe4iqhf3h7kp12v7q8xv15quh6u99569k1zuzce7.pdf [accessed 2019-03-14]
20. Agarwal R, Prasad J. Are Individual Differences Germane to the Acceptance of New Information Technologies? Decision Sciences 1999 Mar;30(2):361-391. [doi: 10.1111/j.1540-5915.1999.tb01614.x]
21. Harrison AW, Rainer RK. The Influence of Individual Differences on Skill in End-User Computing. Journal of Management Information Systems 2015 Dec 16;9(1):93-111. [doi: 10.1080/07421222.1992.11517949]
22. Tams S, Grover V, Thatcher J. Modern information technology in an old workforce: Toward a strategic research agenda. The Journal of Strategic Information Systems 2014 Dec;23(4):284-304. [doi: 10.1016/j.jsis.2014.10.001]
23. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to Use mHealth in Older Adults With Heart Failure. The Journal of Cardiovascular Nursing 2017;32(6):E1-E7. [doi: 10.1097/jcn.0000000000000401]
24. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
25. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
26. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Information & Management 2003 Jan;40(3):191-204. [doi: 10.1016/s0378-7206(01)00143-4]
27. de Vries H, Mesters I, van de Steeg H, Honing C. The general public's information needs and perceptions regarding hereditary cancer: an application of the Integrated Change Model. Patient Educ Couns 2005 Feb;56(2):154-165. [doi: 10.1016/j.pec.2004.01.002] [Medline: 15653244]
28. Ryu M, Kim S, Lee E. Understanding the factors affecting online elderly user's participation in video UCC services. Computers in Human Behavior 2009 May;25(3):619-632. [doi: 10.1016/j.chb.2008.08.013]
29. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
30. van Velsen L, Evers M, Bara C, Op den Akker H, Boerema S, Hermens H. Understanding the Acceptance of an eHealth Technology in the Early Stages of Development: An End-User Walkthrough Approach and Two Case Studies. JMIR Form Res 2018 Jun 15;2(1):e10474 [FREE Full text] [doi: 10.2196/10474] [Medline: 30684434]
31. Hendrikx H, Pippel S, van de Wetering R, Batenburg R. Expectations and attitudes in eHealth: A survey among patients of Dutch private healthcare organizations. International Journal of Healthcare Management 2013 Nov 15;6(4):263-268. [doi: 10.1179/2047971913y.0000000050]
32. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014 Mar 24;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
33. Renaud K, van Biljon J. Predicting technology acceptance and adoption by the elderly: A Qualitative study. 2008 Presented at: Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists; 6-8 Oct 2008; Wilderness, South Africa. [doi: 10.1145/1456659.1456684]
34. van Biljon J, Renaud K. A qualitative study of the applicability of technology acceptance models to senior mobile phone users. 2008 Presented at: International Workshop on Modeling Mobile Applications and Services; 20-23 Oct 2008; Barcelona, Spain p. 228-237. [doi: 10.1007/978-3-540-87991-6_28]
35. Barutçu S, Barutçu E, Adigüzel DU. A Technology Acceptance Analysis for mHealth Apps: The Case of Turkey. Balkan and Near Eastern Journal of Social Sciences 2018;04(04):104-113 [FREE Full text]
36. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han H. Facilitators of and Barriers to mHealth Adoption in Older Adults With Heart Failure. CIN: Computers, Informatics, Nursing 2018:1. [doi: 10.1097/cin.0000000000000442]
37. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
38. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
39. Heale R, Twycross A. Validity and reliability in quantitative studies. Evid Based Nurs 2015 Jul 15;18(3):66-67. [doi: 10.1136/eb-2015-102129] [Medline: 25979629]
40. Cortina JM. What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology 1993;78(1):98-104. [doi: 10.1037/0021-9010.78.1.98]
41. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research Internet. Reading, MA: Addison-Wesley; 1975.
42. Taylor S, Todd P. Assessing IT Usage: The Role of Prior Experience. MIS Quarterly 1995 Dec;19(4):561. [doi: 10.2307/249633]
43. Taylor S, Todd PA. Understanding Information Technology Usage: A Test of Competing Models. Information Systems Research 1995 Jun;6(2):144-176. [doi: 10.1287/isre.6.2.144]

44. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
45. Leung KF, Wong WW, Tay MSM, Chu MML, Ng SSW. Development and validation of the interview version of the Hong Kong Chinese WHOQOL-BREF. Qual Life Res 2005 Jun 1;14(5):1413-1419. [doi: 10.1007/s11136-004-4772-1] [Medline: 16047516]
46. McDowell I. Measuring Health: A guide to rating scales and questionnaires. New York: Oxford University Press; 2006.
47. Wu I, Li J, Fu C. The adoption of mobile healthcare by hospital's professionals: An integrative perspective. Decision Support Systems 2011 Jun;51(3):587-596. [doi: 10.1016/j.dss.2011.03.003]
48. Laan W, Zuithoff NPA, Drubbel I, Bleijenberg N, Numans ME, de Wit NJ, et al. Validity and reliability of the Katz-15 scale to measure unfavorable health outcomes in community-dwelling older people. J Nutr Health Aging 2014 Nov 8;18(9):848-854. [doi: 10.1007/s12603-014-0558-5] [Medline: 25389963]
49. Mlinac ME, Feng MC. Assessment of Activities of Daily Living, Self-Care, and Independence. Arch Clin Neuropsychol 2016 Sep 29;31(6):506-516. [doi: 10.1093/arclin/acw049] [Medline: 27475282]
50. Mueller-Schotte S, Zuithoff NPA, van der Schouw YT, Schuurmans MJ, Bleijenberg N. Trends in Risk of Limitations in Instrumental Activities of Daily Living Over Age in Older Persons With and Without Multiple Chronic Conditions. J Gerontol A Biol Sci Med Sci 2020;75(1):197-203. [doi: 10.1093/gerona/glz049]
51. Askari M, Eslami S, van Rijn M, Medlock S, Moll van Charante EP, van der Velde N, et al. Assessment of the quality of fall detection and management in primary care in the Netherlands based on the ACOVE quality indicators. Osteoporos Int 2016 Feb 21;27(2):569-576 [FREE Full text] [doi: 10.1007/s00198-015-3235-6] [Medline: 26194490]
52. Suijker JJ, Buurman BM, van Rijn M, van Dalen MT, ter Riet G, van Geloven N, et al. [Identification of seniors at risk--primary care: a validated questionnaire predicting functional decline]. Tijdschr Gerontol Geriatr 2015 Apr 8;46(2):113-121. [doi: 10.1007/s12439-015-0128-2] [Medline: 25850542]
53. Askari M, Eslami S, Scheffer AC, Medlock S, de Rooij SE, van der Velde N, et al. Different risk-increasing drugs in recurrent versus single fallers: are recurrent fallers a distinct population? Drugs Aging 2013 Oct 20;30(10):845-851. [doi: 10.1007/s40266-013-0110-z] [Medline: 23959914]
54. Lavrakas PJ. Encyclopedia of survey research methods. Thousand Oaks, California, United States of America: Sage Publications, Inc; 2008.
55. Reichenheim ME, Coutinho ES. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. BMC Med Res Methodol 2010 Jul 15;10(1):66 [FREE Full text] [doi: 10.1186/1471-2288-10-66] [Medline: 20633293]
56. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. IJBHR 2011;2(4):320. [doi: 10.1504/ijbhr.2011.043414]
57. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

Abbreviations

ADL: Activities of Daily Living
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
eHealth: electronic health
ISAR-PC: Identification of Seniors at Risk – Primary Care
mHealth: mobile health
OR: odds ratio
STAM: Senior Technology Acceptance Model
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 02.02.20; peer-reviewed by R Yin, T Wyatt; comments to author 23.03.20; revised version received 15.05.20; accepted 03.06.20; published 04.09.20.

Please cite as:
Askari M, Klaver NS, van Gestel TJ, van de Klundert J
Intention to use Medical Apps Among Older Adults in the Netherlands: Cross-Sectional Study
J Med Internet Res 2020;22(9):e18080
URL: https://www.jmir.org/2020/9/e18080
doi:10.2196/18080
PMID:32624465

Original Paper

Exploring the Use of Mobile Health to Improve Community-Based Health and Nutrition Service Utilization in the Hills of Nepal: Qualitative Study

Ajay Acharya[1*], MPH; Kenda Cunningham[2*], DrPH; Shraddha Manandhar[2], MPH; Niva Shrestha[1], MPH; Mario Chen[3], PhD; Amy Weissman[4*], DrPH

[1]Family Health International 360, Kathmandu, Nepal

[2]Helen Keller International, Kathmandu, Nepal

[3]Global Health, Population, and Nutrition, Family Health International 360, Durham, NC, United States

[4]Asia Pacific Regional Office, Family Health International 360, Bangkok, Thailand

[*]these authors contributed equally

Corresponding Author:
Amy Weissman, DrPH
Asia Pacific Regional Office
Family Health International 360
19th Floor, Tower 3, Sindhorn Building
130-132 Wireless Road, Lumpini, Phatumwan
Bangkok, 10330
Thailand
Phone: 66 2 263 5200 ext 20226
Email: aweissman@fhi360.org

Abstract

Background: With mobile phone coverage and ownership expanding globally, mobile health (mHealth) interventions are increasingly being used to improve coverage and quality of health and nutrition services. However, evidence on mHealth intervention feasibility and factors to consider during program design is limited in low- and middle-income countries like Nepal.

Objective: This study aimed to examine the potential of using text messages to improve health and nutrition services by exploring mobile phone ownership and sharing; mobile phone use and skills; and interest, preferences, and limitations regarding mHealth interventions.

Methods: We conducted 35 in-depth interviews with 1000-day women (the period from conception to a child's second birthday), health facility staff, and female community health volunteers (FCHVs), as well as eight focus group discussions with health facility staff, FCHVs, and 1000-day household decision-makers (ie, husbands, mothers-in-law, and fathers-in-law). We also conducted a mobile phone skills test. We employed thematic analysis using framework matrices and analytical memos.

Results: The study included 70 study participants, of whom 68 (97%) had a mobile phone, and phone sharing was uncommon. Use of text messages was most commonly reported by 1000-day women and health facility staff than household decision-makers and FCHVs. More than 8 in 10 participants (54/64, 84%) could dial numbers, and the majority (28/34, 82%) of 1000-day women, health facility staff, and male decision-makers could also read and write text messages. We found that 1000-day women preferred educational and reminder messages, whereas health facility staff and FCHVs desired educational and motivational messages. Participants suggested different types of texts for 1000-day women, families, FCHVs, and health facility staff, and reported less value for texts received from unknown phone numbers.

Conclusions: A text message–based mHealth intervention is acceptable in the hills of Nepal and has the potential to improve community health and nutrition service utilization, particularly by sending meeting reminders and by providing information. Our findings contribute to text message–based mHealth intervention design in under-resourced settings.

(J Med Internet Res 2020;22(9):e17659) doi:10.2196/17659

KEYWORDS

mobile health; text messages; SMS text message; qualitative study; Nepal; health and nutrition services; health mothers' group; female community health volunteers; mobile phone

Introduction

Mobile health (mHealth) involving the use of mobile phone technologies, such as text messaging, voice services, global navigation satellite systems, and mobile apps, is increasingly being adopted as a means of delivering public health interventions [1]. With expanding mobile phone ownership and accessibility in low- and middle-income countries (LMICs), interest in using mHealth interventions to overcome some of the health service delivery barriers has grown in recent years [1,2].

Text messages delivering reminders and other information can encourage recipients to adopt or maintain healthy behaviors. A previous meta-analysis showed an increase in antenatal care (ANC) visits and delivery by skilled birth attendants among women who received text message reminders about ANC appointments and health information during their pregnancy [3]. Despite the increased use of mHealth to improve health service uptake in LMICs [4], little is known about its feasibility and effectiveness. Although systematic reviews of mHealth in LMICs showed positive effects on maternal and child health service utilization [5-8], the reviewed studies mostly employed text messaging and varied widely in intervention and implementation details, study design, and context. Further, a systematic review on the use of mHealth for behavior change communication (BCC) interventions showed no promising results regarding effectiveness [9]. All of these systematic reviews identified the need to continue building the mHealth evidence base and highlighted the specific need for high-quality implementation research [5-8].

In LMICs like Nepal, where there is nearly universal mobile phone ownership and coverage of mobile phone service, the potential for mHealth interventions to extend the reach and quality of health services is encouraging. As of 2016, nearly all Nepali households (93%) owned at least one mobile phone (94% in urban areas, 91% in rural areas). Among 20 to 24-year-old individuals (the age group most likely to possess a mobile phone), 85% of women and 96% of men owned a mobile phone [10]. In Nepal, prepaid service, where customers pay before accessing the service offered by two leading mobile network service providers (Nepal Telecom and Ncell), is common.

Despite measurable progress in health and nutrition outcomes in the last few decades, Nepal retains the highest rates of maternal mortality in South-East Asia [11], and 36% of children under 5 years are chronically malnourished (stunted) [10]. The reasons for this include poor health service delivery, with financial constraints, unequal distribution of the health workforce, challenges in service extension and integration, and poor access to and inconsistent quality of health services [12,13].

Many development partners and the government of Nepal are committed to improving the availability and uptake of quality health services. For instance, Helen Keller International and Family Health International (FHI) 360 (authors' organizations) and five other partners are implementing *Suaahara II* (2016-2021), a United States Agency for International Development–funded multisector nutrition program, which is now operational in 42 of Nepal's 77 districts. *Suaahara II* involves both household and community-level interventions to improve health and nutrition-related behaviors and increase demand for quality health services, as well as system-level interventions to improve the quality of health services both at facilities and in communities [14].

Believing that mHealth interventions may help address some of the most challenging structural barriers to accessing community health and nutrition services [15], development partners and the government are beginning to implement mHealth activities in Nepal. For example, in January 2018, *Suaahara II* started an age- and stage-specific text message campaign of reminders and motivational messages targeting 1000-day households (the period from conception to a child's second birthday), which involves messages to improve health and nutrition practices, as a new social and behavior change intervention to complement its ongoing home visits, group meetings, food demonstrations, and mass media radio programs. This campaign is based on monitoring data pointing to the high prevalence of mobile phone ownership and use [16] and two rounds of qualitative formative research exploring the potential for sending text messages to provide information on health and nutrition. *Suaahara II* has also been interested in designing a similar text message intervention for service providers to help address supply-side health service constraints.

Evidence on the feasibility of using text messages to improve uptake of government-led health and nutrition services in LMICs, however, is scant. Given these gaps, we conducted a qualitative study with both 1000-day households and service providers to determine the feasibility of using mHealth within the context of *Suaahara II* to improve the reach and quality of health and nutrition services in Nepal. Specifically, this study aimed to understand the potential of using text messages with 1000-day households and/or service providers to improve health and nutrition services by exploring mobile phone ownership and sharing; mobile phone use and skills; and interest, preferences, and limitations regarding mHealth interventions.

Methods

We conducted a qualitative study employing in-depth interviews (IDIs) and focus group discussions (FGDs) with health facility staff, female community health volunteers (FCHVs), 1000-day women, and other household decision-makers (ie, husbands, mothers-in-law, and fathers-in-law of the 1000-day women). The study was conducted in December 2018 in Syangja, a hilly district of Nepal, comprising 11 urban and rural municipalities made up of 97 wards with a population of about 290,000. The district has 97 public health facilities and 612 FCHVs [17]. We purposefully selected Syangja because it is a *Suaahara II* district, is geographically similar to other hilly districts, and has

good mobile network coverage, a prerequisite for a text message intervention.

Within Syangja, the research team purposively selected one rural (Kaligandaki) and one urban (Chapakot) municipality. In consultation with *Suaahara II* staff and based on the implementation status of service-related interventions, the team selected one ward in each municipality. Kaligandaki's sampled ward had implemented the Self Applied Technique for Quality Health (SATH) approach, a *Suaahara II* participatory approach that supports health mothers' groups (HMGs) to conduct live social mapping with community members, encourages mothers to seek health services, and tracks 1000-day women's health service utilization, whereas Chapakot's sampled ward had not received this intervention.

All health facility staff and FCHVs of the selected wards were included in the study. Families of women in the 1000-day period were purposively selected. To identify 1000-day women, each health facility in charge selected one more and one less effective FCHV (performance assessed via monthly reports and evaluations) who in turn provided the names of five 1000-day women from their FCHV-led monthly HMG meeting registry. We asked FCHVs to ensure selection of mothers with diverse backgrounds, including different caste/ethnicity groups, age, level of education, and mobile phone access. Five additional 1000-day women per municipality were recruited by asking the five selected 1000-day women to each refer one additional mother from the community. Researchers also invited other 1000-day family members to participate in the study by asking each 1000-day woman who completed the interview to identify all adult decision-makers (husband, mother-in-law, or father-in-law) currently living in the same household and available to participate in the study.

We conducted 9-day training for 10 data collectors and pretested all data collection tools. The trained data collectors conducted IDIs and FGDs. IDIs were conducted with 10 FCHVs (five per ward) and 20 1000-day women (10 per ward). Eight FGDs were conducted as follows: four with male and female household decision-makers separately (two from each ward), two with FCHVs (one from each ward), and two with health facility staff (one from each ward). Researchers perceived this sample size to be adequate to reach saturation [18] but agreed to continue data collection until saturation was achieved. No new themes emerged in the final IDIs or FGDs.

Topics explored during IDIs included access to a mobile phone; phone and network coverage quality; and phone and text messaging use and preferences such as timing, type, frequency, recipient, and intention to reply to or share the text messages received. Similarly, wider community perception on mobile phone use/SIM card change, network access, phone sharing, text message information sharing with others, text message interest and preferences, and thoughts on using text messages as a BCC intervention were explored during FGDs.

Researchers also assessed the ability of all 1000-day women, FCHVs, and health facility staff to use a mobile phone. First, researchers asked the participants if they could dial on the phone and read and/or write a text message. Thereafter, researchers asked the participants to dial a phone number (usually that of the researcher) and then to read any existing text message on their mobile phone. If there were no text messages on the phone, the researcher sent a sample message in Nepali and asked the participant to open and read it. The researcher also asked each participant to write and send a text message from their phone to the study moderator's phone number.

IDIs and FGDs were audio-recorded and then transcribed into Nepali by trained data collectors. Two translators were hired to translate the transcripts into English. Two researchers selected and independently open coded the same eight transcripts using NVivo 12 (QSR International). These were then used to develop the initial themes and codebook, which were used to code the remaining transcripts. On occasion, researchers added new themes and codes generated during coding. The two researchers developed a framework matrix using NVivo and independently analyzed the data producing analytical memos and then triangulated the results. Themes were identified and analyzed, and any similarities/differences among data collection sites and as per the categories of participants were noted. Following this round of analysis, in July 2019, researchers visited the study sites and gathered all of the health facility staff and FCHVs who had been study participants to share and validate the emerging findings and explore their perceptions and agreement/disagreement vis-a-vis the findings related to 1000-day families.

The study was approved by the Nepal Health Research Council on July 2, 2018, and FHI 360's Protection of Human Subject Committee on November 21, 2018. All study participants provided written informed consent for both the interview/discussion and recording.

Results

Demographics

We conducted 35 IDIs and 8 FGDs, with a total of 70 participants (Table 1). The majority of health facility staff were young (mean age 28 years), and 5 out of 8 (63%) were from the Janajati socially disadvantaged caste in Nepal, with some Brahmins (socially advantaged caste). All the health workers possessed the qualifications required for their positions. Their service duration ranged from just 4 days to more than 15 years. The majority of FCHVs were in their early fifties (mean age 51 years), and 11 out of 17 (65%) were Brahmins, with some Janajatis. While slightly more than one-third (6/17, 35%) of FCHVs had completed 10 years of schooling, 53% (9/17) had not completed secondary school (below 8 years of schooling), and 2 out of 17 (12%) had no formal education. Nearly all had been FCHVs for more than 16 years, but two FCHVs were appointed only 4 months before data collection.

Table 1. Distribution of study participants.

Approach	Value, n	Description
In-depth interviews	35	
Health facility staff	6	Three per site
FCHVs[a]	9	Four in Kaligandaki and five in Chapakot
1000-day women[b]	20	10 per site
Focus group discussions	8	
Health facility staff	2	One per site with four participants per FGD[c]
FCHVs	2	One per site with four participants per FGD
Male decision-makers	2	One per site with four to five participants per FGD
Female decision-makers	2	One per site with five participants per FGD

[a]FCHVs: female community health volunteers.

[b]The period from conception to a child's second birthday.

[c]FGD: focus group discussion.

Most 1000-day women were in their mid-twenties. Of the 20 1000-day women, 35% (7/20) had completed at least 10 years of schooling, and 4 out of 20 (20%) had not completed primary school. The majority (14/20, 70%) were Brahmins, with some Janajatis and Dalits. Most women were living in a joint family with their in-laws. All female decision-makers who participated in the study were mothers-in-law (n=10), with a mean age of 50 years. The majority of male decision-makers in the study were husbands (n=6), with a mean age of 29 years. A few decision-makers were fathers-in-law (n=3), with a mean age of 55 years.

Mobile Phone Ownership and Sharing

All study participants, except two of the 10 mothers-in-law, had a mobile phone and reported high mobile phone ownership among adult family members. The majority (48/64, 75%) of participants had a smartphone. We found no differences in phone ownership by age, sex, or caste among the majority of study participants. A few participants mentioned that ownership among women has increased to enable them to contact male relatives, many of whom emigrate for work.

The majority of the 1000-day women reported that they are allowed to access and read text messages, and most decision-makers confirmed this lack of restriction. Phone sharing was uncommon at the study sites, with a few exceptions. The majority (15/19, 79%) of 1000-day women mentioned accessing another phone, mainly that of their husbands, to make urgent calls when their phones had insufficient credit, when their phones stopped working, or when wanting to watch videos or check Facebook and this was not possible on their own phones. Some husbands also mentioned using the same phone as their wives (1000-day women). Some mothers-in-law expressed hesitation to ask their daughters-in-law to use their phones but felt comfortable sharing their phones with their daughters-in-law.

Being out of network coverage was uncommon at the study sites. The majority of study participants reported never having changed the SIM card. In the FGD with Kaligandaki health facility staff, age differences in SIM card behaviors were reported as follows:

> *The aged people use the old number; we find the newer generation changing the SIM frequently.* [Health facility staff, Kaligandaki, FGD]

Mobile Phone Use and Skills

Participants reported using mobile phones for various purposes, but primarily to receive calls. Text messaging was more frequently reported by 1000-day women and health facility staff compared with other male and female decision-makers and FCHVs. The majority (14/20, 70%) of 1000-day women reported using their phones to access the internet, particularly Facebook, Imo (video call and chat app), and YouTube, to make video calls, chat via text/messenger, and play games. Nearly all health facility staff, but only few FCHVs, reported accessing the internet on their phones. Some participants also reported differences in phone use among their household members, with fathers-in-law and mothers-in-law using the phone primarily to make calls and younger family members (eg, mothers and fathers of young children) using the phone to access the internet and for messaging.

More than 8 in 10 study participants (54/64, 84%) could dial phone numbers, but only about 72% (46/64) were able to read and 56% (36/64) were able to write text messages. The majority of those who could read and write text messages were 1000-day women, their husbands, and health facility staff (Table 2). Five of the 10 FCHVs reported a preference for voice compared to text messages, stating that they prefer to listen than read. This was supported by the mobile skills test, which found that three of the 10 were illiterate. Some of the FCHVs and health facility staff reported a preference for receiving important messages via both voice and text. In a health facility staff FGD, participants reported that text messages would be fine for health workers because of their capacity to read; however, a combined form would be more effective because reading would be bothersome especially when text messages are long. One health facility staff member made the following statement:

Person may feel bored in the middle of messages and feel why should I read this. So, it would be effective to listen, as well. The things that someone listened to, that person can present it in another place. [Health facility staff, Kaligandaki, FGD]

Table 2. Mobile phone use skills.

Category	Phone dial[a]	SMS text message writing and reading[a]
1000-day women[b] (N=20)	20 (100%) could dial the phone number	16 (80%) could read and 14 (70%) could write text messages in Nepali
FCHVs[c] (N=17)	15 (88%) could dial the phone number	14 (82%) could read and 8 (47%) could write text messages in Nepali
Health facility staff (N=8)	8 (100%) could dial the phone number	8 (100%) could read and write text messages in both Nepali and English
Male decision-makers (husbands) (N=6)	6 (100%) could dial the phone number	6 (100%) could read and write text messages in Nepali
Male decision-makers (fathers-in-law)[d] (N=3)	3 (100%) could dial the phone number	1 (33%) replied that they could read text messages in Nepali
Female decision-makers (mothers-in-law)[d] (N=10)	2 (20%) could dial the phone number	1 (10%) replied that they could read text messages in Nepali

[a]Data are provided as n (%).

[b]The period from conception to a child's second birthday.

[c]FCHVs: female community health volunteers.

[d]Assessment based on response.

Interest and Preferences for Text Message Interventions

Nearly all participants expressed interest in receiving text messages with health and nutrition information and about community-provided services, such as HMG meetings. Some topics of interest mentioned were nutrition, timing of ANC visits, postnatal care checkups, child immunization, growth monitoring and promotion, family planning, deworming, iron and calcium consumption, HMG meeting times and discussion topics, drinking water purification, and personal, family, and environmental sanitation.

Many 1000-day women also mentioned a desire for messages on the topics discussed in the HMG meeting to help participants retain information. The 1000-day women also expressed an interest in receiving text messages because the content would give them something to discuss when they are with their peers. Women made the following statements:

Some things could be confused or forgotten or due to noise could not be understood properly in the meetings. We can get information about that thing by reading [a] message. [1000-day woman, Chapakot, IDI]

SMS is good in that we can read the message, we talk about the message we read, and we can talk to each other when we go to cut grass and leaves. [1000-day woman, Kaligandaki, IDI]

Health facility staff and FCHVs were also interested in health and nutrition-related text messages and perceived them as a type of job aid. They explained that the new information would update and/or refresh their knowledge, and they would be able to store and access the information repeatedly on their phones. They also mentioned a desire for motivational messages that would inspire them to improve their work, as described by one IDI participant as follows:

If messages are sent regarding how a health worker should work, we will have to think about those things. We will realize if we are out of track, the things we need to improve. I will get a chance to improve those things. [Health facility staff, Kaligandaki, IDI]

Health facility staff and FCHVs also highlighted that text messages to families may encourage greater HMG participation (for example, by reminding mothers of the meeting time and communicating with mothers who they are unable to reach because of their heavy workloads). Health facility staff made the following statements:

If people get [a] reminder message today about tomorrow's meeting, then they can show that message to the household head and tell [them] that they have to go tomorrow. [Health facility staff, Chapakot, IDI]

If the message goes to the family, there can be family discussions about the meeting, which might help create a more convenient environment within the family to allow women to attend meetings. [Health facility staff, Chapakot, FGD]

Health facility staff also noted that such messages would provide information that would tempt women to join HMGs to learn more and improve women's perception of the importance of HMG meetings, thus encouraging more active participation. One health facility staff member made the following statement:

If topics of discussion are sent through SMS, they will realize how appropriate is it for them to know about it. There might be some things they might have known and some that they might not have known. They will

> *develop a feeling that they should go.* [Health facility staff, Kaligandaki, IDI]

Health facility staff from both sites suggested that text messages should also be sent to FCHVs because they are the main organizers of the HMG meetings and the text messages can serve as a reminder and a guide in meeting preparation. FCHVs in Chapakot identified health facility staff as the priority text message recipients so health facility staff can teach the FCHVs new information learned from the messages. FCHVs in Kaligandaki prioritized health facility staff and FCHVs themselves but recommended different content for each group, since health facility staff are perceived to have more knowledge than FCHVs. Health facility staff in Chapakot noted that 1000-day mothers are the most crucial text message target.

Most FCHVs and household decision-makers from both study sites prioritized 1000-day women as primary text message recipients. This preference was because the women in the 1000-day period were perceived to be more interested in these matters, more educated, and more able to use the information as they were considered less likely to migrate away for work.

Some participants in both IDIs and FGDs expressed preferences regarding the frequency and timing of text messages. For example, in both settings, the majority (18/20, 90%) suggested that it would be most appropriate to receive a reminder text message 1 to 2 days before the HMG meeting. Furthermore, the majority (17/20, 85%) of 1000-day women preferred receiving text messages in the evening, but before 8 PM. Some women, however, reported being comfortable receiving text messages any time during the day. The majority (19/20, 95%) suggested that an SMS campaign send text messages once a month. One woman made the following statement:

> *Instead of making people feel bored by sending many messages, it would be good to send once or twice a month. Many messages should not be sent.* [1000-day woman, Chapakot, IDI]

When asked about their response to receiving text messages, the majority (17/24, 71%) reported that they would read a message from an unknown phone number out of curiosity. Yet, some participants, particularly FCHVs and 1000-day mothers, mentioned that they would not read it, delete it, or read it but not reply. The majority (15/20, 75%) of 1000-day women also showed a high level of interest in interactive messaging that allowed them to reply to received text messages. They mentioned that they would respond to questions, reply with thanks, or provide feedback on the messages. Furthermore, nearly all (19/20, 95%) 1000-day participants reported that they would share information from the text messages with family members (eg, husband and mother-in-law), friends, and neighbors. Mothers-in-law from Kaligandaki, however, mentioned that when a daughter-in-law receives a text message, she would prefer not to share it with her mother-in-law because of differences in thoughts and attitudes. A health worker and a male decision-maker from Chapakot mentioned that they would like to share the text messages via Facebook to spread the word. Additionally, a health worker from Chapakot pointed out that sharing depends upon the nature and source of the message as follows:

> *If there is a message from Suaahara, you should also read. It was about what we should feed our children, and how should we take care of our children. We have to take care and feed the children.* [Health facility staff, Chapakot, FGD]

Discussion

This study describes the interest, preferences, and skills related to a text message–based mHealth intervention among both 1000-day households and health and nutrition service providers. Our findings confirm high mobile phone ownership with sufficient skills among study participants to engage in a text message–based mHealth intervention. Participants preferred reminder messages that would provide information about the date and time of community-based services, such as HMGs. They also preferred informational text messages related to health and nutrition and interactive text messages from known phone numbers. The prioritized recipients were 1000-day women and FCHVs.

Prior research has shown promising use of text messages to improve uptake of health facility maternal and child health services in LMICs [5,6] and FCHV mobile phone use for data collection and disease surveillance following adequate training and support in Nepal [19,20]. Our study findings suggest that text messages could also be used to increase uptake of community health and nutrition services, such as HMGs meetings led by FCHVs, in Nepal and in similar low-resource settings.

Our study found that although 1000-day women have access to other phones in their families, they prefer to use their own phones and phone sharing is uncommon, which is important because previous studies have found reliance on shared phones a challenge for mHealth interventions. This is particularly true for retrieved voice messages [21] and for sensitive topics, such as family planning [22], because the information conveyed does not remain confidential [23]. Our findings suggest that information conveyed via an mHealth intervention in the hilly areas of Nepal will remain confidential and will reach the intended recipient. This highlights that phone sharing should be considered when designing mHealth interventions in LMICs [23].

Understanding target populations' message delivery preferences is important [9] as mHealth interventions are more likely to be effective when message frequency, content, and style are tailored to these needs and preferences [15]. In our study, the majority of participants preferred receiving messages once a month and not early in the morning or at night. This is consistent with a qualitative study on text messaging preferences among adolescent females in the United States, where participants reported a desire for text messages in the afternoon or evening [24]. According to a meta-analysis, interventions using user-driven tailored message frequency were more beneficial compared with interventions that applied fixed or varying frequency [25].

In our study, we found a mixed preference for message delivery among health facility staff and FCHVs. Nearly half (5/10, 50%)

of the FCHVs favored voice messages, whereas some FCHVs and health staff favored a combined approach of both text and voice messages. The preference of FCHVs for voice messages, given that they were slightly older and that some were illiterate, is consistent with findings from a cluster randomized controlled trial in the United States that found younger and higher-skilled users favored text messages [26]. Meanwhile, a study in rural Malawi determined that voice messages are more desirable by illiterate users [21]. These studies concluded that mHealth interventions need to include multiple modalities to ensure more participants benefit. This aligns with our findings and reflects that mHealth interventions in LMICs may be on the cusp of change owing to increased literacy levels.

We also found that although participants would read messages from unknown phone numbers, they would not value them. This suggests that before starting an mHealth intervention or sending text messages, programs should prepare the target audience and ensure the phone number to be used is communicated. Participants also expressed a strong interest in replying to text messages, suggesting that the mHealth intervention may be more acceptable if it requests a reply. This would also help to monitor whether messages have been received and read or to check whether messages were understood. Literature to date, however, points to contradictory findings on the benefits of this type of interaction. A meta-analysis on the efficacy of text messaging showed no difference between interventions that text participants and ask them to text back and those that text participants only [25]. Meanwhile, another systematic review of periodic messaging interventions showed that mHealth interventions using participants' feedback were more effective compared with interventions not using feedback [27].

Owing to the purposive sampling of the data collection area, selection bias may have influenced the study results. Yet, when selecting study participants, we ensured diversity in age, education, caste, and ethnicity, so we believe our results reflect a broad range of people in the hilly areas of Nepal. The mountains and lowland plains, however, have not been reflected in this study. It is also important to note that the study was conducted in a *Suaahara II* implementation district, and thus, the study participants may have more exposure to study topics than populations residing in districts without at-scale health and nutrition interventions. We anticipate that if the intervention we designed can be successfully implemented and proven effective in improving health outcomes in the context of the *Suaahara II* program, it can be scaled up to additional geographies in Nepal. This would require additional validation of the findings across a broader range of settings beyond the one in this study. Furthermore, additional research is needed to deepen our understanding of the opportunities to use mHealth interventions to improve service reach and delivery in Nepal and other LMICs. For example, further implementation research on the effectiveness and costs associated with different implementation modalities (eg, text only, voice only, and combined messages as well as one-way versus interactive messaging) is needed.

Text messaging appears to be a promising cost-effective way to deliver health and nutrition information in remote and geographically dispersed populations like those in Nepal. Unlike other mHealth interventions, such as the use of applications, text messaging is easy to implement and does not require a special phone. This study provides important insights for the design of a text message–based mHealth intervention and suggests that mHealth has the potential to improve community health and nutrition service utilization in Nepal, particularly by sending meeting reminders and by providing information. Different groups had different preferences and requirements for potential text messages. Text message–based mHealth interventions in under-resourced settings could be encouraged and may result in higher service coverage and ultimately improved health and nutrition practices.

Acknowledgments

This study was supported by Family Health International 360's Ward Cates Emerging Scientific Leader Award. The authors would like to acknowledge FHI 360, and *Suaahara II* Kathmandu and Syangja teams for their support and collaboration throughout this study. Finally, the authors would like to thank all the participants and the data collection, transcription, and translation teams for their time and effort in this study.

Authors' Contributions

AA, KC, and AW designed the study, conceptualized the manuscript, guided the analysis, and supported writing of multiple drafts. AA and AW conducted the analysis and the literature review, and drafted the manuscript. SM and NS conducted the analysis, supported the design of the qualitative data collection tools and methodology, and provided editorial support. MC provided study document review and general guidance throughout the study, guided the analysis, and helped with manuscript development. All authors reviewed multiple drafts of the manuscript, and read and approved the final version.

Conflicts of Interest

None declared.

References

1. WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2019-10-13]

2. Qiang CZ, Yamamichi M, Hausman V, Altman D. Mobile Applications for the Health Sector. World Bank. 2012. URL: http://documents1.worldbank.org/curated/en/751411468157784302/pdf/726040WP0Box370th0report00Apr020120.pdf [accessed 2019-10-13]
3. Wagnew F, Dessie G, Alebel A, Mulugeta H, Belay YA, Abajobir AA. Does short message service improve focused antenatal care visit and skilled birth attendance? A systematic review and meta-analysis of randomized clinical trials. Reprod Health 2018 Dec 22;15(1):191 [FREE Full text] [doi: 10.1186/s12978-018-0635-z] [Medline: 30466453]
4. Bassi A, John O, Praveen D, Maulik PK, Panda R, Jha V. Current Status and Future Directions of mHealth Interventions for Health System Strengthening in India: Systematic Review. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11440 [FREE Full text] [doi: 10.2196/11440] [Medline: 30368435]
5. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the Effect of mHealth Interventions in Improving Maternal and Neonatal Care in Low- and Middle-Income Countries: A Systematic Review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
6. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. BMC Health Serv Res 2017 Dec 07;17(1):704 [FREE Full text] [doi: 10.1186/s12913-017-2664-7] [Medline: 29115992]
7. Colaci D, Chaudhri S, Vasan A. mHealth Interventions in Low-Income Countries to Address Maternal Health: A Systematic Review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]
8. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: Systematic review and meta-analysis. J Glob Health 2016 Jul;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
9. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
10. Nepal Demographic and Health Survey. The DHS Program. 2016. URL: https://www.dhsprogram.com/pubs/pdf/fr336/fr336.pdf [accessed 2019-10-13]
11. World Health Organization. Country Office for Nepal. Nepal – WHO country cooperation strategy: 2018-2022. World Health Organization. URL: https://apps.who.int/iris/handle/10665/272476 [accessed 2019-10-17]
12. WCO Nepal. Health System in Nepal: Challenges and strategic options. World Health Organization. 2007. URL: https://apps.who.int/iris/handle/10665/205257 [accessed 2019-10-18]
13. Ranabhat CL, Kim C, Singh A, Acharya D, Pathak K, Sharma B, et al. Challenges and opportunities towards the road of universal health coverage (UHC) in Nepal: a systematic review. Arch Public Health 2019;77:5 [FREE Full text] [doi: 10.1186/s13690-019-0331-7] [Medline: 30740223]
14. SUAAHARA II Good Nutrition Program-Second Annual Report. USAID. 2018. URL: https://pdf.usaid.gov/pdf_docs/PA00TD9X.pdf [accessed 2019-10-18]
15. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
16. SUAAHARA II Good Nutrition Program Annual Survey, Year One. Care Evaluations. 2018. URL: https://www.careevaluations.org/evaluation/suaahara-ii-good-nutrition-program-annual-survey-year-one-2017/ [accessed 2019-10-18]
17. Central Bureau of Statistics, Government of Nepal. National Population and Housing Census 2011. UN Statistics. 2011. URL: https://unstats.un.org/unsd/demographic-social/census/documents/Nepal/Nepal-Census-2011-Vol1.pdf [accessed 2019-10-21]
18. Namey E, Guest G, McKenna K, Chen M. Evaluating Bang for the Buck. American Journal of Evaluation 2016 Jul 09;37(3):425-440. [doi: 10.1177/1098214016630406]
19. Meyers DJ, Filkins M, Harsha Bangura A, Sharma R, Baruwal A, Pande S, et al. Management challenges in mHealth: failures of a mobile community health worker surveillance programme in rural Nepal. BMJ Innov 2017 Jan 17;3(1):19-25. [doi: 10.1136/bmjinnov-2015-000102]
20. Meyers DJ, Ozonoff A, Baruwal A, Pande S, Harsha A, Sharma R, et al. Combining Healthcare-Based and Participatory Approaches to Surveillance: Trends in Diarrheal and Respiratory Conditions Collected by a Mobile Phone System by Community Health Workers in Rural Nepal. PLoS One 2016;11(4):e0152738 [FREE Full text] [doi: 10.1371/journal.pone.0152738] [Medline: 27111734]
21. Crawford J, Larsen-Cooper E, Jezman Z, Cunningham SC, Bancroft E. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. Glob Health Sci Pract 2014 Jan 28;2(1):35-46. [doi: 10.9745/ghsp-d-13-00155]
22. Laidlaw R, Dixon D, Morse T, Beattie TK, Kumwenda S, Mpemberera G. Using participatory methods to design an mHealth intervention for a low income country, a case study in Chikwawa, Malawi. BMC Med Inform Decis Mak 2017 Jul 05;17(1):98 [FREE Full text] [doi: 10.1186/s12911-017-0485-6] [Medline: 28679428]
23. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013 Dec 05(12):CD007458 [FREE Full text] [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]

24. Chernick LS, Schnall R, Stockwell MS, Castaño PM, Higgins T, Westhoff C, et al. Adolescent female text messaging preferences to prevent pregnancy after an emergency department visit: a qualitative analysis. J Med Internet Res 2016 Oct 29;18(9):e261 [FREE Full text] [doi: 10.2196/jmir.6324] [Medline: 27687855]
25. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Dec;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
26. Greaney ML, Puleo E, Sprunck-Harrild K, Bennett GG, Cunningham MA, Gillman MW, et al. Electronic reminders for cancer prevention: factors associated with preference for automated voice reminders or text messages. Prev Med 2012 Aug;55(2):151-154 [FREE Full text] [doi: 10.1016/j.ypmed.2012.05.014] [Medline: 22659227]
27. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]

Abbreviations

ANC: antenatal care
BCC: behavior change communication
FCHV: female community health volunteer
FGD: focus group discussion
FHI: Family Health International
HMG: health mothers' group
IDI: in-depth interview
LMIC: low- and middle-income country

Edited by G Eysenbach; submitted 01.01.20; peer-reviewed by A Ecker, S Burchert; comments to author 18.06.20; revised version received 28.06.20; accepted 26.07.20; published 11.09.20.

Please cite as:
Acharya A, Cunningham K, Manandhar S, Shrestha N, Chen M, Weissman A
Exploring the Use of Mobile Health to Improve Community-Based Health and Nutrition Service Utilization in the Hills of Nepal: Qualitative Study
J Med Internet Res 2020;22(9):e17659
URL: http://www.jmir.org/2020/9/e17659/
doi:10.2196/17659
PMID:32915151

Original Paper

Operability, Usefulness, and Task-Technology Fit of an mHealth App for Delivering Primary Health Care Services by Community Health Workers in Underserved Areas of Pakistan and Afghanistan: Qualitative Study

Shehla Zaidi[1], MBBS, MSc, PhD; Abdul Momin Kazi[2], MBBS, MPH; Atif Riaz[2], MSc; Ammarah Ali[1], MSc; Rabia Najmi[1], MSc; Rawshan Jabeen[1], MSc; Umerdad Khudadad[1], MSc; Saleem Sayani[3], MNCM

[1]Department of Community Health Sciences, Aga Khan University, Karachi, Pakistan
[2]Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan
[3]Digital Health Resource Centre, Aga Khan Development Network, Karachi, Pakistan

Corresponding Author:
Shehla Zaidi, MBBS, MSc, PhD
Department of Community Health Sciences
Aga Khan University
Karachi
Pakistan
Phone: 92 3062016825
Email: shehla.zaidi@aku.edu

Abstract

Background: The recent proliferation of digital health technology in low- and middle-income countries has made it possible for community health workers (CHWs) to use mobile health (mHealth) to perform tasks such as data collection and training. Although most studies focus on the prospect of digital apps to motivate and connect CHW, only a few have captured end-user experiences with mobile-based apps. We examined the experience of frontline health workers with a move towards digitalized real-time data to record maternal and childcare services in remote areas of Afghanistan and Pakistan.

Objective: Our study aimed to explore CHW perceptions on the operability of the mHealth app in a community setting, usefulness of the app in the delivery of assigned maternal and childcare functions, and the task-technology fit with monitoring information systems.

Methods: The Hayat app, designed to digitalize and facilitate electronic record keeping, was evaluated to be embedded into mainstream health systems. The app had 2 components: smartphone app for data entry and web dashboard for visualization of the maternal, newborn, and child health reports. Using a qualitative exploratory study design, we conducted a total of 8 focus group discussions with purposively selected lady health workers (LHWs) and CHWs in 3 districts of Pakistan and 3 hamlets of Afghanistan, respectively. Focus group discussions were conducted in the local language, audio recorded, and converted into expanded notes for thematic analysis.

Results: Although a majority of LHWs used the app with ease, some initially faced difficulties in operating it and requested a longer duration of training. Contrary to LHWs, the CHWs were able to use the app without difficulty, as they were using it only to register clients. Overall, use of the mHealth app in both countries resulted in a positive impact on health education sessions, easier communication with parents or clients, tracking of routine immunization defaulters and follow-ups, improved data validity, easily accessible vaccination schedules, and faster registration. In addition to building up their image in the community and personal development, the improved reporting and monitoring mechanisms also set the stage for the LHWs to get recognized for their hard work. CHWs in Afghanistan also reported the app provided immediate access to information when requested by their supervisor. Although the Hayat app eliminates the need to carry multiple registers and helps in recalling client information at the touch of a button, technical issues around connectivity and data inputting tabs were highlighted by the participants.

Conclusions: The digitization of records not only provided CHWs support in their daily routine but also strengthened monitoring mechanisms and improved motivation. We recommend conducting end user experience studies before embedding apps into mainstream health systems as high acceptability does not always result in high uptake of digital technology.

(J Med Internet Res 2020;22(9):e18414) doi:10.2196/18414

KEYWORDS

mHealth; community health workers; usability; usefulness; task-technology fit

Introduction

The vital role of community health workers (CHWs) as a bridge between the community and health care services is getting renewed attention in low- and middle-income countries (LMICs). CHWs have collectively helped in reducing the maternal and child mortality burden and assisted in decreasing the burden and costs of tuberculosis and malaria [1-3]. Pakistan and Afghanistan have well-structured CHW programs known as the lady health worker (LHW) program in Pakistan and CHW program in Afghanistan, which have provided primary health care services for more than a decade [4]. In both programs, CHWs perform similar functions ranging from health education and awareness; diagnosis and treatment of prevalent diseases such as diarrhea, malaria, respiratory illnesses, and intestinal worms; directly observed therapy for tuberculosis; either antenatal or postnatal health care; and referrals to the health care facilities [5,6]. As Pakistan and Afghanistan still face some of the highest maternal and child mortality burdens and fertility rates in the region [7], the role of frontline health workers is critical to meet the sustainable development goal targets. A supportive work environment, motivation, and competency building are the key factors identified by the framework for strengthening CHW performance [8,9]. However, LHWs often work under little supervision and outreach support in Pakistan [10,11]. Similarly, in Afghanistan, capacity building and establishing support systems have been reported to improve service delivery by CHWs [12].

Cellphone penetration >90% in LMICs, coupled with falling call prices and increased network connectivity options, has improved the feasibility of mobile health (mHealth) programs [13] in remote areas of LMICs, creating the possibility for strengthening weak health systems. CHWs around the globe are using mHealth technology for data collection, training, communication, mobile job aids, decision support tools, and behavior change communication in the community. The few examples of successful mHealth-based facilitation of service provision by CHWs in remote and hard-to-reach areas in LMICs include an increase in the registration and uptake of family planning services in rural Tanzania [14] and an 85% reduction in the average number of days for overdue visits in Dar es Salam, Tanzania [15]. In Pakistan, an app called “e-Vaccs” was introduced in the province of Punjab to track the movement of vaccinators using GPS. The app was implemented to improve vertical accountability through digitalization of records and was supported by the government of Punjab (Punjab Information Technology Board) [16]. This led the way for other provinces to adopt digital immunization solutions to improve vaccination coverage, a prime example being the introduction of the Teeko app in the rural districts of the province of Sindh [13].

The majority of mHealth studies have focused on the prospect of a digital app to motivate and connect CHWs among themselves, their supervisors, and around the facilities [17]. However, few studies have captured end-user experiences with a mobile-based app [18,19]. User experience takes a broader look at the individuals’ entire interaction with the app that includes thoughts, feelings, and perceptions resulting from that interaction [20]. If the intended aim of the app is user satisfaction and improved work efficiency, the understanding of end-user experiences can provide valuable details on the features and functions most needed for the intended task, make available detailed knowledge of issues to be addressed, and offer an experience comparable with the traditional methods of working [21]. Concerns of not having enough data regarding the implementation and evaluation of the app before recommending the embedding of the app into mainstream health systems have been highlighted by academics globally. They fear a graveyard of poor-quality, unproven apps contributing to a fragmented mHealth landscape [22]. Some experts recommend assessing the barriers and challenges faced by the targeted end users before designing the app to be able to provide mHealth as a good-fit solution [20,23]. A study attributed the failure of technology to its lack of regard for user requirements and experience [24]. Another recent mHealth study aiming to inform how digital technology can be embedded within district health systems in LMICs highlighted the need for co-option by end users and district stakeholders. It also reported that ease of operability, satisfaction with reliable data, personal recognition, links to field support, and empowerment are powerful enablers of the shift towards digitalization [13].The purpose of our study was to explore the necessary facets of the initial version of the Hayat app revolving around user centeredness to provide an evidence base informed by the understanding of experiences and needs of CHWs in remote areas of Pakistan and Afghanistan [25].

In this paper, we examine the end-user experiences of CHWs in Afghanistan and LHWs in Pakistan with moving towards digitalized real-time data for recording maternal and childcare services provided in the community. Evidence is drawn from experiences with implementing the Hayaat app, a mobile-based app for CHWs piloted in select remote areas of Pakistan and Afghanistan to track community-based health care delivery. The app was launched with an end goal of scalability throughout the regions. As compared with apps previously introduced in Pakistan, the Hayat app had a scaled-up immunization component from the Teeko app as well as a pilot maternal care component. The theory of change underpinning and guiding this intervention assumes that digital intervention will improve the validity and timeliness of reporting by health workers, allowing for accountability of CHW and LHW performance and on-site support resulting in improved visit frequency and quality. This improved visit frequency and quality are theorized to translate into increased household awareness and practices regarding maternal, newborn, and child health (MNCH) leading to improved health outcomes. Key features of the app included GPS to track CHWs during outreach visits, registering clients in the health facility and during outreach, and MNCH data

collection. We draw on the experiences of frontline primary health workers using the mHealth app to inform the shift towards digitalization. The objectives of this study were to explore the perceptions of CHWs on the operability of the app in the community setting, usefulness of the app in the delivery of assigned MNCH functions, and task-technology fit with CHW's monitoring information systems. We aimed to inform the embedding of the digital app into the mainstream health systems of both countries. We also hoped the findings of this evaluation could contribute to app development in other LMICs and inform the process of end-user evaluations.

Methods

Study Setting

This study was nested in a larger ongoing quasiexperimental study assessing the effectiveness of the Hayat app for improving maternal and child health in Pakistan and Afghanistan. In Pakistan, the app was implemented in 9 health care facility catchment areas of the Chitral District in Khyber Pakhtunkhwa Province and in the catchment areas of the districts Astore and Ghizer in the Gilgit-Baltistan Province covering a total population of 310,012 people, of which 49% are women. For this study, 4 union councils (UCs), 2 each from Khyber Pakhtunkhwa and Gilgit-Baltistan, were randomly selected from a list of UCs in which the Hayat app was implemented.

In Afghanistan, the Hayat app was implemented in 4 health facility catchment areas (Iragh, Shunbul, Khandaq, and Kalo) in Bamyan Province and 3 catchment areas (Ghaaran, Baharak, and Ishkashim) in Badakshan Province (total population of 117,878, 49% women). The health facilities in the catchment areas of these 2 provinces are operated by Aga Khan Health Services Afghanistan in collaboration with the Ministry of Public Health. For this study, we selected all the catchment areas from the 2 provinces in which the Hayat app was implemented (Table 1).

Table 1. Districts in which the Hayat ap was implemented.

Country	Catchment areas in which the Hayat app was implemented	Target population	Selected areas for this study
Pakistan	Kosht, Karimabad, Mulkhow, Charun, Lotkoh, Shoghore, Chatorkhand, Ishkoman, Bubar, Hatoon, Sherqill, Singal, Gupis, Phander, Pingal, Sumal, Teru, Hundur Silgan, Taus, Thoi, Yasin, Gahkuch	310,012	Buni, Garam Chashma, Gahkuch, Gupis
Afghanistan	Ishkashem, Baharak, Gharaan, Iragh, Shunbul, Khandaq, Kalo	117,878	Ishkashem, Baharak, Gharaan, Kalo, Shunbul, Khandaq, Iragh

mHealth Intervention (Hayat App)

The Hayat app was designed to digitalize record keeping, to make it easy for CHWs and shift from paper-based to mobile-based electronic record keeping. The app had two components: smartphone app for data entry and dashboard (web portal) for visualization of the MNCH reports at a later stage. In both countries, the CHWs were provided smartphones with the Hayat app. The app was designed according to the contextual role and scope of the CHWs in the respective country. For further details on the workflow of the Hayat app, refer to Figures 1 and 2.

The Hayat app features include client registration, data collection, GPS, capabilities, and awareness content.

Figure 1. Workflow of the Hayat app for community health workers. ANC: antenatal care; MNCH: maternal, newborn, and child health; PNC: postnatal care.

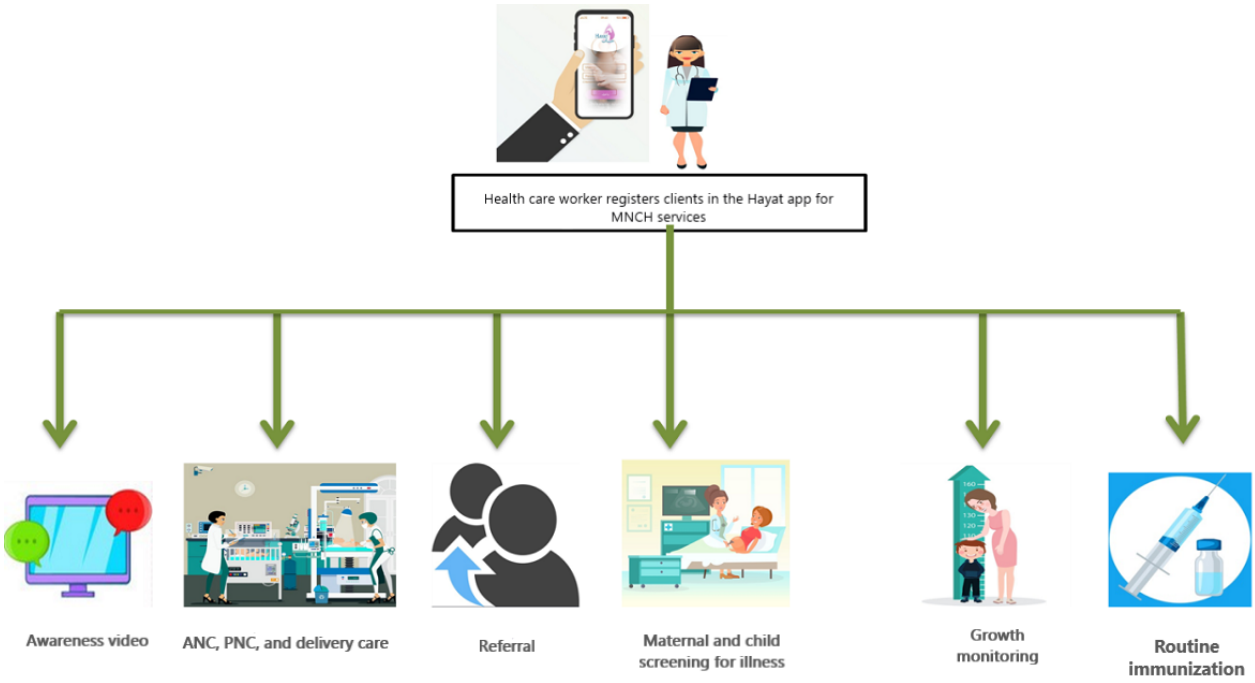
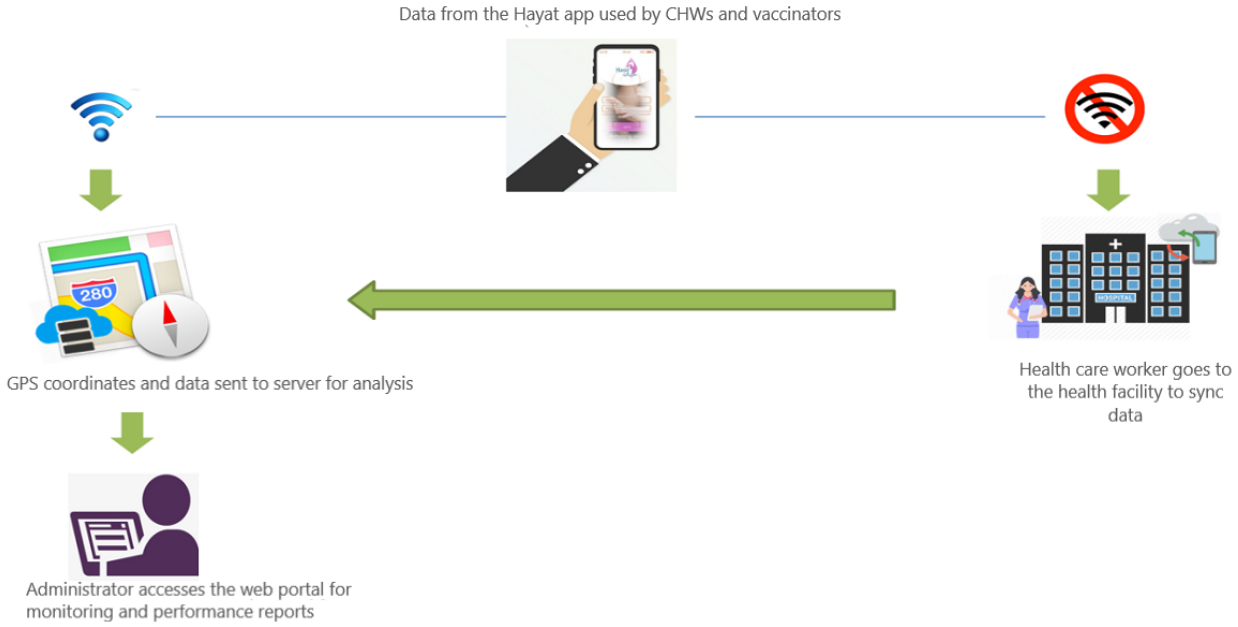


Figure 2. Workflow of the Hayat app for health managers. CHW: community health worker.



Client Registration

Clients can be registered, with their demographic profile, by CHWs in 2 categories: >2 years of age and <2 years of age. To find the client record, the app used multiple means such as fingerprint, QR card, ID card number, or household number.

Data Collection

LHWs can record the demographic and health-related data of pregnant women, women of childbearing age, and children <2 years old. Along with the demographic profile of each registered client, data can be recorded for antenatal care (ANC), delivery, postnatal care (PNC), family planning, illnesses, and maternal and child vaccinations.

GPS

GPS capabilities allowed real-time tracking of CHWs during outreach visits.

Awareness Content

The app has awareness videos to facilitate the CHWs in health education sessions and outreach activities (Figure 3).

Figure 3. Awareness content being shown to the community.

Framework

We used an adapted framework derived from the task-technology fit model [26] and technology acceptance model [27] to explore the operability, usefulness, and task-technology fit of the mHealth app centered on the end users' experiences. For this evaluation, we defined the assessment areas as follows.

Operability was defined as the perceptions and barriers such as ease of use, technical competency, and accessibility. This measure aimed to understand the characteristics and capabilities of end users to capture maternal and child health indicators in a resource-constrained setting using digital app.

Usefulness was defined as the perceived satisfaction of health managers and health workers in achieving their goals, including the results and consequences of use.

The task-technology fit of the app was defined as the capability of the app to facilitate the designated tasks of the end users including how the technology interacted with the management information system interface.

In line with the purpose of our study, we used a deductive form of analysis using 3 broad a priori themes of operability, usefulness, and task-technology fit. These themes guided the construction of interview guides, as our goal was to explore these areas in detail to assess whether the app works for our intended users (ie, CHWs) and help tease out any issues with the app.

Data Collection

Data were collected in July and August 2019 from selected districts of Afghanistan and Pakistan in which the Hayat app was introduced in April 2019. We used a qualitative exploratory study design and conducted focus group discussions (FGDs) with LHWs and CHWs in Pakistan and Afghanistan, respectively. FGD is a widely used tool for collecting data on the perceptions and experiences of health care workers and provides an opportunity for in-depth probing [25]. The moderator started out by asking the participants questions related to their own thoughts and experiences to establish rapport and hear the extent to which our a priori topics were already part of the participants' thinking before we cued them to think along those terms. Our intention was to provide the necessary user-centered feedback to inform the development cycle of the app for future use, easy embedding into mainstream health systems, and ensuring its sustainability.

A total of 8 FGDs were conducted, including 4 FGDs with LHWs in Pakistan and 4 FGDs with CHWs in Afghanistan, achieving theoretical saturation.

Selection of Study Participants

Pakistan

The participants for this study were purposively selected and approached for FGDs. One FGD was conducted in each selected UC with 10-12 participants each.

Afghanistan

Altogether, 4 FGDs were conducted with 8-12 purposively selected participants in each session: 2 in Shebar district, 1 in Baharak district, and 1 in Ishkesham district. The sample size and participant characteristics are provided in Table 2.

Table 2. Characteristics of the community health workers (CHWs) and lady health workers (LHWs).

Characteristics	Pakistan (n=54)	Afghanistan (n=42)
Age (years), mean (SD)	39 (5.7)	39 (13.3)
Level of education, n (%)		
<10 years	0	25 (60)
At least matriculation (10-14 years)	34 (63)	14 (33)
Graduate (≥16 years)	20 (37)	3 (7)
Years of service, mean (SD)	18.6 (7.1)	10 (5.7)

A semistructured interview guide was developed for the FGDs for both countries, using relevant literature and adapted to the local context. Questions were framed around the a priori identified themes for the mobile app including operability, perceived usefulness, and task-technology fit. Probes were added where necessary to steer the discussions. Guides were developed in the English language and translated into local languages (ie, Urdu for Pakistan and Dari for Afghanistan). Experienced data collectors were hired and trained to conduct FGDs, manage data, and maintain professional and ethical conduct throughout the data collection. A free flow of discussion was ensured during the FGDs, and all participants were encouraged to contribute to the discussion. Notes were taken by the interviewer during the interviews and FGDs. The main points were summarized by the interviewer at the end of each interview and FGD to get confirmation from the participants. The interviews were conducted by one researcher in each country. These researchers belonged to the country assigned and were proficient in the local language. This ensured cultural sensitivity and familiarity with the area. The FGDs were arranged at a time and place convenient for participants, and each FGD lasted 40-50 minutes. To limit bias, the research team that undertook tool development, collection, and analyses was not part of the intervention team.

The interviews and FGDs were recorded using a digital recorder. All electronic data were stored on encrypted computers, and consent forms and recordings were stored in a locked drawer in the Community Health Sciences department at Aga Khan University.

Data Analysis

Recordings were transcribed verbatim in Urdu and Dari and translated into English at the Community Health Sciences department, Aga Khan University Pakistan campus by the research teams from Pakistan and Afghanistan, respectively. Translations were reviewed for accuracy by experienced researchers who had command of both the local and English languages. Personal identifiers were removed, and unique IDs were assigned to each participant in the FGDs. Transcripts were reviewed repeatedly to gain familiarity with the data. A thematic content analysis of the textual data was conducted using Nvivo 10 by 2 trained researchers from the team. Transcripts from Pakistan and Afghanistan were coded separately, and similar codes were categorized under the aforementioned thematic areas (ie, operability, perceived usefulness, and task-technology fit). Within each thematic area, further deductive coding was conducted based on emerging concepts from the narrative.

Ethical Considerations

Approval was obtained from the ethical review committee of Aga Khan University Hospital (ERC#2018-0375-951) prior to the start of the study, and institutional review board approval was obtained from the Afghanistan National Public Health Institute (IRB# AIRB-0419-0014). Written informed consent was obtained from all participants after explaining the purpose of the study.

Results

Operability

Pakistan

Despite the concerns shared by the LHWs regarding the short training period for the app, the majority stated that they were able to use the app with ease. A few mentioned that they faced difficulties in operating the app initially; however, they were able to work with it after a bit of trial and error and ongoing facilitation by the supervisors. Regarding the language of the app, in one of the FGDs, the LHWs suggested that it should be changed to Shina (a local language) as some of them had to translate the app features into Shina for some of the mothers to understand. The LHWs that were unable to translate were helped by other LHWs. It was also noted by the LHWs that the monthly call package helped them in paying the costs of the calls, which provided an efficient way of communicating with mothers and caretakers regarding their child's health. They were able to connect through cell phones with the mothers and caretakers, reducing extra unnecessary household visits and more efficient use of their time. There was a reluctance to make the calls without a prepaid call package, as was the case during initial days of the app implementation.

> *At first, we made many mistakes such as writing a female name in the male section. It was because we were not experienced at all. It took time but we learned.* [LHW Booni]

> *Now, we don't have to go everywhere. We get a phone package, and we call them [mothers]. Earlier, we used to call them using our own phones.* [LHW Gahkuch-Ghizer]

> *It would be great if the language of the app is changed to Shina.* [LHW Booni]

Afghanistan

The CHWs expressed that the app was user friendly, and they were able to use it without any trouble because the app was being used for registration of clients only and not for documentation of health status. They also mentioned being supported in the outreach by the Aga Khan Health Services Afghanistan personnel that provided technical support in operating the app. The CHWs were operating the app in their local language (ie, Dari).

> *Some people did not give us the tazkira number, which was a problem. But it is resolved now. We talked to the people who gave us this mobile phone about this, and they have asked us to just put 0799 in place of the tazkira number. This is not a problem now.* [CHW Baharak]

Usefulness

Pakistan

The majority of the LHWs agreed that the awareness videos for the community outreach education sessions were very useful. However, they pointed out that additional videos on other topics such as depression and cancer are also needed. According to one LHW, to change the behavior of the community, variation in the methods of health education sessions is needed. Furthermore, some of the LHWs noted an increase in the communities' response to the follow-up on vaccination and reported that the community viewed the videos with interest and asked questions regarding child health.

> *They (videos) are beneficial for mothers because the videos attract them, and they watch them with interest.* [LHW Garam Chasma]
>
> *We give health education in health care centers, schools, and mosques too, so if we have videos of other diseases, that would be great.* [LHW Booni]
>
> *What we used to tell them verbally wasn't that effective. Now, conducting sessions with videos makes it really interesting, and they understand better.* [LHW Gupis]

The digitalization of health records improved the work efficiency of the LHWs, as they were able to complete an hour's worth of tasks in just a few minutes. Most of the LHWs reported that digitalization made the procedure of vaccination easier as, instead of entering data in multiple registers, LHWs can enter information in a single mobile device. They are also able to contact mothers for the scheduled vaccines; previously, they did not have enough calling credits. The LHWs noted that the app showed not only defaulter children by changing the colors of the entry but also the vaccination schedule of each child, which helped them list and contact mothers. The app was also reported to be helpful for monitoring child growth. The LHWs shared that, before the app, they did not weigh the children; however, with the app, they are required to weigh them for monitoring growth. The LHWs claimed they did not have growth charts for the last 3-4 years. However, now they have a growth chart in the app, which they can show mothers to inform them about their child's nutrition and health status

> *It has made our work easier now. Before, for vaccination, we had to do entry in different registers, but now we only need to enter in the mobile.* [LHW Garam Chasma]
>
> *This [growth chart] is also useful for the mother, as they get to know about the health of the children; if they are skinny, they concentrate on the lines because they don't want their children to be skinny.* [LHW Gahkuch Ghizer]

The LHWs shared that the app had a positive effect in building up their image in the community. They felt that, with the introduction of the app, the community now trusted them and showed respect. The LHWs were pleased about the possibility of getting recognized for their hard work. They shared that the department was unhappy with their collective work and called them out for it despite their efforts, even during the campaigns. They hoped that the digitalization of records would allow for individual performance monitoring and give them their due credit. The LHWs also shared their changed attitude towards their work. Before the app, they had grown tired of working with multiple heavy registers for a long time. The replacement of these registers by a mobile phone brought excitement into their work and motivated them. Many stated that this was their first mobile phone and reported developing skills in using it and considered it as a skill-building opportunity. They reported improved self-esteem and confidence.

> *Our department is not happy with us. They criticize us for not working. We are lucky that we have mobile phones now so that we can get credit for all the work we do. We are very happy with this app.* [LHW Booni]
>
> *We work very passionately. Because of the mobile phone, we have changed. Before, we didn't know how to use a mobile phone, but we got it because of this project.* [LHW Gahkuch Ghizer]
>
> *We are very happy with mobile; it has made our work very easy. We enjoy our work now. We are happy.* [LHW Booni]

Afghanistan

The CHWs shared that they were most excited about using the video feature in the app for conducting health sessions at different locations in the community. This facilitated them in addressing communities' concerns and raising awareness regarding the side effects of vaccination. The CHWs noted these videos to be beneficial for they created awareness and harnessed the community's interest.

> *When we just speak, it doesn't have a big impact. People remember things by watching. The good thing about this app is the videos. People can see it as well as listen.* [CHW Ishkashim]
>
> *Once a month, there is a meeting where all the community people are invited — sometimes in the mosque or at the home of community leaders. We have the responsibility of health education, so we use these videos to educate them about immunization, and they become more motivated for vaccination after watching this.* [CHW Baharak]

> *Now people have become aware. They ask us to give them information not just on vaccination but also about nutrition, tuberculosis, and diabetes…every kind of disease.* [CHW Baharak]

The CHWs reported increased efficiency of their work due to the digital app. They were able to comparatively register more children for vaccination and report their information to the vaccinators in a timely manner so that they can complete vaccinations. They also were able to share the child vaccination status with the parents and connect them to a nearby health care facility. The CHWs felt that the process of MNCH services was accelerating, as the app helped them focus on their catchment areas only. This helps vaccinators find the exact location of the child for routine immunization. Moreover, the saved information and vaccination schedule allowed them to track defaulter children. The digitalization of records eliminated duplication of work because the data are now saved in the app; earlier, they had to be careful to avoid losing information through torn, missing, and wet pages. It also facilitated the reporting of data to the clinic because they did not have to carry multiple registers.

> *For vaccination, the parents should be connected with the nearby health facility, which was missing previously. Parents keen to get their children vaccinated didn't have enough information about the vaccination status of their children and which health facility to go for vaccination. It is resolved through this app.* [CHW Shebar]

> *Before, when we came across someone who asked for the vaccination status of their child, it was difficult for us because the register was at home. But now the same register is in our pocket. We can register children on the spot and share any information if they need it.* [CHW Baharak]

The positive influence of the app on the community was pointed out by the CHWs. In case the outreach team was unable to vaccinate a child, the mother could bring the child to the facility for vaccination. The app also helped the CHWs respond to community questions. The CHWs were pleased about improved reporting mechanisms and having information at hand through the mobile phone when asked by their supervisors. The CHWs reported personal development through the app. The app not only helped them get acquainted with digital technology but also improved their knowledge base regarding MNCH services.

> *My knowledge about the MNCH services was low before the introduction of this app, and I wasn't able to respond to some of the mothers' queries. But now, I have gone through the videos many times, and it has helped me to understand MNCH services such as ANC, PNC, breastfeeding, immunization, and family planning.* [CHW-Shebar 2]

> *From the perspective of reporting, it has helped us in a great way. We were unable to give the right report to the health facility before this app. They weren't even happy with our work. But now, the reporting is good, and we are not losing any data.* [FGD 4 Ishkashim]

Task-Technology Fit

Task-technology fit was defined as the capability of the app to facilitate the designated tasks of the end users including interoperability (ie, how the technology interacts with the management information system interface).

Pakistan

The LHWs expressed that all the work and data entry they did in multiple registers was now being done in a single mobile phone, eliminating the need to carry multiple heavy registers in the field for data recording. They also used the app to identify defaulter children for routine immunization. The LHWs were happy with the online data synchronization, as it will highlight their hard work for higher management. Nevertheless, the LHWs noted their frustration over not being able to edit or delete the client records. This required them to reenter data, which then created duplicate registrations. One LHW shared that because the baby was not named at the time of registration, she had to register the child again as the app would not allow her to edit the baby's name. Also, some features present in the LHW register were reported missing in the app. The app did not include the monthly plans made by the LHWs nor the community chart that includes the total population, number of kids, and death and birth records. They also raised concerns regarding the limited options to choose from in certain categories such as nutrition and adverse outcomes in pregnant mothers.

> *It's an online system, and it will highlight our hard work to the top.* [LHW Booni]

> *There are fewer options than our registers — like if a pregnant woman has a mishap, we enter that in our register, but we can't enter it in the app.* [LHW Gahkuch Ghizer]

Afghanistan

The CHWs expressed appreciation for being able to recall client information at the touch of a button. Earlier, this level of recall was only possible by going through multiple registers. They also felt that the data are now more secure in the central server than in the registers. However, the CHWs voiced their apprehensions regarding the lack of connectivity and synchronization of data with the central server located at the Aga Khan Health Service office. They had to travel long distances every week for data transfer. The CHWs pointed to the lack of network coverage in their regions as the major hindrance to data transfer, as it requires internet connectivity. Their coordination with the health workers in facilities was also affected as they were not able to inform them of their referred cases in time. They also raised concerns regarding the lack of editing and deleting options in the app. This was a problem, as this did not allow them to edit mistakes and information. The CHWs also wanted to have the option to generate a list of children according to their gender.

> *This app is an ever-ready resource for us. Anytime we want something — information on the vaccine status of child or anything else — we can easily find it.* [CHW Ishkashim]

The negative thing about this app is that the data are not synchronized with the portal system, and we have to travel a distance of 3 hours every week to transfer the data into the portal manually. [CHW Shebar]

Discussion

Principal Findings

Although the importance of including end users' perspective for effective implementation of mHealth interventions has been emphasized in the past [28], the contribution of qualitative evaluation of mHealth apps to supplement their implementation has not received much attention in most South Asian countries. This study adds important insights on the function and applicability of the Hayat mHealth app in facilitating routine work of LHW and CHWs in Pakistan and Afghanistan, respectively.

Our findings show that most of the LHWs and CHWs found the app easy to use and it facilitated day-to-day tasks in Pakistan and Afghanistan. Table 2 compares the characteristics of the sampled participants. While the mean age of participants was 39 years in both countries, the level of education and years of experience were higher in participants from Pakistan compared with participants in Afghanistan. The reason for these differences is that LHWs in Pakistan are paid government employees whereas CHWs in Afghanistan work as volunteers. Furthermore, the LHWs are required to have completed at least matriculation, whereas there is no such requirement for recruitment of CHWs. Even though the 2 populations differ significantly, the findings related to the operability and usefulness of the app were largely similar in both countries. However, the app was being used for the full range of MNCH services in Pakistan in contrast to Afghanistan, where its use was limited to client registration, education, referral, and immunization. The app was found to be useful on many fronts; some of the encouraging results highlighted in this study included improvement of service delivery planning, time management and efficiency, accuracy of collected data, communication with caregivers, caregiver's compliance with instructions, individual LHW performance accountability, and simplification of the work routine. CHWs in Uganda and Mozambique also felt that an mHealth app had the potential to improve their work efficiency, planning, and communication with supervisors [29]. Offline features of the app and awareness videos were reported to be most useful when conducting outreach visits, supporting the use of videorecorded health messages for raising awareness in communities. Findings reported from a study conducted in Pakistan indicated the usefulness of an offline feature; however, awareness videos on routine immunization were reported as a less-used feature by the vaccinators [13].

Simultaneously, findings of this study also highlighted the challenges faced by end users in remote and hard-to-reach areas that will require addressing to maximize the utilization of the app in resource-constrained settings. First, network coverage was limited, and while the offline feature prevented data from being lost, the users still felt frustration at not being able to sync data in real-time as it then became a pending or additional task requiring their attention later. In addition, issues in communicating referrals with health facilities and traveling long distances to upload data were some of the other constraints. Second, the capacity building of digital literacy skills for those in need should be carried out on a regular basis to maximize the output of an mHealth intervention. Third, because the CHWs were fully aware of the capabilities of the dashboard and app to monitor individual performance, there was a strong sense that acknowledgment of individual performance was expected. They were not satisfied with the current supervision mechanisms, which masked individual efforts. They were judged on cumulative performance, without individual acknowledgment, creating grounds for favoritism and unjust admonishment. The study also highlighted that the CHWs hoped that the app would help address the need to constantly prove their efforts to their supervisors and dedication to the program, especially in Afghanistan, where they are volunteers [6].

Our study showed that management of data was facilitated, reducing missing and erroneous entries and improving transparency and accountability. Multiple ways of registration allowed for many ways to recall data if one form of information was lost. The CHWs were also excited about the possibility of abolishing registers that were a hassle to work with. CHWs using the Geohealth app in Brazil also reported that replacing paper with the app greatly reduced the load they carried into the field [30].

Technical issues highlighted during the study need to be considered because they strongly contribute to the usability of the app; if not corrected, these issues ultimately threaten uptake. A study in São Paulo, Brazil highlighted that the technical characteristics of an app should align with the task features of CHWs for it to be operable, useful, and sustainable. It also highlighted the need to involve end users in assessing the technical barriers in hardware and software to maximize the utilization of an mHealth app [30].

The benefits of equipping health workers with data management technologies to improve the continuity of care in hard-to-reach populations have been established in the existing literature. Global examples revealed high levels of acceptance and willingness to learn despite lack of experience with such tools [31-33]. It also shows that the use of mHealth apps by CHWs has improved outreach services, data collection and management, and outreach activities [34,35]. Moreover, a digital app has the potential to improve users' knowledge, skills, and performance [36-38]. The literature shows that the contribution of an app to the improvement of health outcomes depends on its perceived usefulness and usability by the users. It highlights the role of the users in the successful implementation of an mHealth app [39,40]. The usefulness of the mHealth app for CHWs is centered on the ease of integration into the routine workflow, improved capacity to deliver follow-up services, less paperwork, organization of data, and abolishing the paper-based registers [30-41]. Moreover, a boost in the CHWs motivation when provided with mobile phones was reported in a study conducted by Madon et al [37].

Strengths and Limitations

Iterative qualitative methods used to explore the experiences of the end users helped identify factors influencing the uptake of the digital app.

The participants' responses may have been skewed towards being socially acceptable, hence creating a social desirability bias. Because this study involved cell phones (desired product), the bias might be significant. Several strategies were adopted to avoid or limit any social desirability bias. The FGD participants were volunteers, and the FGDs were conducted at a time and place convenient to the participants to make them feel comfortable [42]. Interviewees were only provided a brief overview of the study at the beginning to avoid preparing respondents to answer in a socially acceptable way [43]. Moreover, the interviews were conducted by an experienced moderator that strived to create an open and comfortable atmosphere for the participants to share their viewpoints and experiences [44]. The participants were assured that there was no incorrect answer, their responses would not affect their participation or lead to their mobile phones (given as part of the study) being taken away, and the FGD was about hearing their diverse views and understanding their feelings. They were also encouraged to support their views with personal experiences [45]. These strategies collectively provided confidence that any social desirability bias was reduced.

Conclusions

Involving end users is necessary for successful integration of an mHealth app into existing programs. Qualitatively evaluating the digital app from the lens of operability, usefulness, and task-technology fit provided us with a broader picture of the factors that could affect its uptake in the longer term. Our findings indicate that the app being in the local language, supplemented with a simple interface, made it easy for the CHWs to adopt and use it. It saved time, organized the work routine, removed the need to work with manual data registers, and promoted accountability. The relevance of this study extends beyond the 2 countries to similar LMIC settings.

Acknowledgments

The study was funded by Grand Challenge Canada (GCC) and Aga Khan Foundation Canada. The funders had no role in the design of the study; collection, analysis, and interpretation of data; and writing the manuscript.

The authors would like to extend their gratitude to the National Ministry of Health Services Regulation & Coordination, Pakistan and National Ministry of Public Health, Afghanistan for providing support during the evaluation. In addition, we would like to acknowledge the efforts and extend our gratitude to the participants of the Lady Health Worker Program, Pakistan and Community Health Worker program, Afghanistan. We would also like to thank Aga Khan Health Services (AKHS) Pakistan and Afghanistan for their facilitation during the data collection process.

Authors' Contributions

SZ conceptualized, designed, and supervised the study. MK and AR led the data collection and analysis. AR and UK collected the data, while AA and RJ conducted the analysis. AA, RJ, and RN conducted the literature review and refined the analysis. SZ, AR, AA, RN, and SS drafted the paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Christopher JB, Le May A, Lewin S, Ross DA. Thirty years after Alma-Ata: a systematic review of the impact of community health workers delivering curative interventions against malaria, pneumonia and diarrhoea on child mortality and morbidity in sub-Saharan Africa. Hum Resour Health 2011 Oct 24;9:27 [FREE Full text] [doi: 10.1186/1478-4491-9-27] [Medline: 22024435]
2. Langston A, Weiss J, Landegger J, Pullum T, Morrow M, Kabadege M, et al. Plausible role for CHW peer support groups in increasing care-seeking in an integrated community case management project in Rwanda: a mixed methods evaluation. Glob Health Sci Pract 2014 Aug;2(3):342-354 [FREE Full text] [doi: 10.9745/GHSP-D-14-00067] [Medline: 25276593]
3. Scott K, Beckham SW, Gross M, Pariyo G, Rao KD, Cometto G, et al. What do we know about community-based health worker programs? A systematic review of existing reviews on community health workers. Hum Resour Health 2018 Aug 16;16(1):39 [FREE Full text] [doi: 10.1186/s12960-018-0304-x] [Medline: 30115074]
4. Najafizada SAM, Labonté R, Bourgeault IL. HRH dimensions of community health workers: a case study of rural Afghanistan. Hum Resour Health 2019 Feb 06;17(1):12 [FREE Full text] [doi: 10.1186/s12960-019-0347-7] [Medline: 30728062]
5. Folz R, Ali M. Task sharing in health workforce: An overview of community health worker programmes in Afghanistan, Egypt and Pakistan. East Mediterr Health J 2018 Dec 09;24(9):940-950 [FREE Full text] [doi: 10.26719/2018.24.9.940] [Medline: 30570127]
6. Najafizada SAM, Labonté R, Bourgeault IL. Community health workers of Afghanistan: a qualitative study of a national program. Confl Health 2014;8:26 [FREE Full text] [doi: 10.1186/1752-1505-8-26] [Medline: 25904976]

7. Akseer N, Kamali M, Arifeen SE, Malik A, Bhatti Z, Thacker N, et al. Progress in maternal and child health: how has South Asia fared? BMJ 2017 Apr 11;357:j1608. [doi: 10.1136/bmj.j1608] [Medline: 28400481]
8. Community systems strengthening framework. Geneva: The Global Fund; 2010.
9. Jaskiewicz W, Tulenko K. Increasing community health worker productivity and effectiveness: a review of the influence of the work environment. Hum Resour Health 2012 Sep 27;10:38 [FREE Full text] [doi: 10.1186/1478-4491-10-38] [Medline: 23017131]
10. Lehmann U, Sanders D. Community health workers: what do we know about them. The state of the evidence on programmes, activities, costs and impact on health outcomes of using community health workers Geneva: World Health Organization. 2007 Jan. URL: https://www.who.int/hrh/documents/community_health_workers.pdf [accessed 2020-09-02]
11. Rabbani F, Shipton L, Aftab W, Sangrasi K, Perveen S, Zahidie A. Inspiring health worker motivation with supportive supervision: a survey of lady health supervisor motivating factors in rural Pakistan. BMC Health Serv Res 2016 Aug 17;16(1):397 [FREE Full text] [doi: 10.1186/s12913-016-1641-x] [Medline: 27535743]
12. Edward A, Branchini C, Aitken I, Roach M, Osei-Bonsu K, Arwal SH. Toward universal coverage in Afghanistan: A multi-stakeholder assessment of capacity investments in the community health worker system. Soc Sci Med 2015 Nov;145:173-183 [FREE Full text] [doi: 10.1016/j.socscimed.2015.06.011] [Medline: 26141453]
13. Zaidi S, Shaikh SA, Sayani S, Kazi AM, Khoja A, Hussain SS, et al. Operability, Acceptability, and Usefulness of a Mobile App to Track Routine Immunization Performance in Rural Pakistan: Interview Study Among Vaccinators and Key Informants. JMIR Mhealth Uhealth 2020 Feb 13;8(2):e16081 [FREE Full text] [doi: 10.2196/16081] [Medline: 32053115]
14. Using Mobile Health Applications to Improve Family Planning Services. The David and Lucile Packard Foundation. 2020. URL: https://www.packard.org/awardee/m-health-improve-quality-family-planning-services-tanzania/ [accessed 2020-09-02]
15. DeRenzi B, Findlater L, Payne J, Birnbaum B, Mangilima J, Parikh T. Improving community health worker performance through automated SMS. 2012 Mar Presented at: Fifth International Conference on Information and Communication Technologies and Development; March 2012; Atlanta, GA p. 25-34 URL: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [doi: 10.1145/2160673.2160677]
16. Tracking Vaccinators (E-VACCS). Punjab Information Technology Board, Government of the Punjab. URL: https://www.pitb.gov.pk/tracking_evaccs [accessed 2020-09-02]
17. Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of Mobile Health (mHealth) Technologies and Interventions Among Community Health Workers Globally: A Scoping Review. Health Promot Pract 2019 Nov;20(6):805-817. [doi: 10.1177/1524839919855391] [Medline: 31179777]
18. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
19. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AHA, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
20. Albert W, Tullis T. Measuring the user experience: collecting, analyzing, and presenting usability metrics. Burlington, MA: Morgan Kaufmann; 2013.
21. Lavallee DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. Health Aff (Millwood) 2016 Apr;35(4):575-582. [doi: 10.1377/hlthaff.2015.1362] [Medline: 27044954]
22. Roess A. The Promise, Growth, and Reality of Mobile Health - Another Data-free Zone. N Engl J Med 2017 Nov 23;377(21):2010-2011. [doi: 10.1056/NEJMp1713180] [Medline: 29116869]
23. Kazi AM, Ahsan N, Khan A, Jamal S, Kalimuddin H, Ghulamhussain N, et al. Personalized Text Messages and Automated Calls for Improving Vaccine Coverage Among Children in Pakistan: Protocol for a Community-Based Cluster Randomized Clinical Trial. JMIR Res Protoc 2019 May 30;8(5):e12851 [FREE Full text] [doi: 10.2196/12851] [Medline: 31148544]
24. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
25. O.Nyumba T, Wilson K, Derrick CJ, Mukherjee N. The use of focus group discussion methodology: Insights from two decades of application in conservation. Methods Ecol Evol 2018 Jan 11;9(1):20-32. [doi: 10.1111/2041-210x.12860]
26. Goodhue DL, Thompson RL. Task-Technology Fit and Individual Performance. MIS Quarterly 1995 Jun;19(2):213. [doi: 10.2307/249689]
27. Davis F, Bagozzi R, Warshaw P. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
28. Chaiyachati KH, Loveday M, Lorenz S, Lesh N, Larkan L, Cinti S, et al. A pilot study of an mHealth application for healthcare workers: poor uptake despite high reported acceptability at a rural South African community-based MDR-TB treatment program. PLoS One 2013;8(5):e64662 [FREE Full text] [doi: 10.1371/journal.pone.0064662] [Medline: 23724075]

29. Thondoo M, Strachan DL, Nakirunda M, Ndima S, Muiambo A, Källander K, InSCALE Study Group. Potential Roles of Mhealth for Community Health Workers: Formative Research With End Users in Uganda and Mozambique. JMIR Mhealth Uhealth 2015 Jul 23;3(3):e76 [FREE Full text] [doi: 10.2196/mhealth.4208] [Medline: 26206419]
30. Schoen J, Mallett JW, Grossman-Kahn R, Brentani A, Kaselitz E, Heisler M. Perspectives and experiences of community health workers in Brazilian primary care centers using m-health tools in home visits with community members. Hum Resour Health 2017 Sep 29;15(1):71 [FREE Full text] [doi: 10.1186/s12960-017-0245-9] [Medline: 28962569]
31. Medhanyie AA, Little A, Yebyo H, Spigt M, Tadesse K, Blanco R, et al. Health workers' experiences, barriers, preferences and motivating factors in using mHealth forms in Ethiopia. Hum Resour Health 2015 Jan 15;13:2 [FREE Full text] [doi: 10.1186/1478-4491-13-2] [Medline: 25588973]
32. Sukums F, Mensah N, Mpembeni R, Kaltschmidt J, Haefeli WE, Blank A. Health workers' knowledge of and attitudes towards computer applications in rural African health facilities. Glob Health Action 2014;7:24534 [FREE Full text] [doi: 10.3402/gha.v7.24534] [Medline: 25361721]
33. Zakane SA, Gustafsson LL, Tomson G, Loukanova S, Sié A, Nasiell J, et al. Guidelines for maternal and neonatal "point of care": needs of and attitudes towards a computerized clinical decision support system in rural Burkina Faso. Int J Med Inform 2014 Jun;83(6):459-469. [doi: 10.1016/j.ijmedinf.2014.01.013] [Medline: 24613689]
34. Blanas DA, Ndiaye Y, MacFarlane M, Manga I, Siddiqui A, Velez O, et al. Health worker perceptions of integrating mobile phones into community case management of malaria in Saraya, Senegal. Int Health 2015 May;7(3):176-182. [doi: 10.1093/inthealth/ihu075] [Medline: 25316707]
35. Schuttner L, Sindano N, Theis M, Zue C, Joseph J, Chilengi R, et al. A mobile phone-based, community health worker program for referral, follow-up, and service outreach in rural Zambia: outcomes and overview. Telemed J E Health 2014 Aug;20(8):721-728 [FREE Full text] [doi: 10.1089/tmj.2013.0240] [Medline: 24926815]
36. Long L, Pariyo G, Kallander K. Digital Technologies for Health Workforce Development in Low- and Middle-Income Countries: A Scoping Review. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):S41-S48 [FREE Full text] [doi: 10.9745/GHSP-D-18-00167] [Medline: 30305338]
37. Madon S, Amaguru JO, Malecela MN, Michael E. Can mobile phones help control neglected tropical diseases? Experiences from Tanzania. Soc Sci Med 2014 Feb;102:103-110 [FREE Full text] [doi: 10.1016/j.socscimed.2013.11.036] [Medline: 24565147]
38. Mc Kenna P, Babughirana G, Amponsah M, Egoeh SG, Banura E, Kanwagi R, et al. Mobile training and support (MOTS) service-using technology to increase Ebola preparedness of remotely-located community health workers (CHWs) in Sierra Leone. Mhealth 2019;5:35 [FREE Full text] [doi: 10.21037/mhealth.2019.09.03] [Medline: 31620462]
39. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
40. Opoku D, Stephani V, Quentin W. A realist review of mobile phone-based health interventions for non-communicable disease management in sub-Saharan Africa. BMC Med 2017 Feb 06;15(1):24 [FREE Full text] [doi: 10.1186/s12916-017-0782-z] [Medline: 28162090]
41. Rothstein JD, Jennings L, Moorthy A, Yang F, Gee L, Romano K, et al. Qualitative Assessment of the Feasibility, Usability, and Acceptability of a Mobile Client Data App for Community-Based Maternal, Neonatal, and Child Care in Rural Ghana. Int J Telemed Appl 2016;2016:2515420 [FREE Full text] [doi: 10.1155/2016/2515420] [Medline: 28070186]
42. Brunk KH. Reputation building: beyond our control? Inferences in consumers' ethical perception formation. Journal of Consumer Behaviour 2010 Jul 20;9(4):275-292 [FREE Full text] [doi: 10.1002/cb.317]
43. Steenkamp L, Venter D, Walsh C, Dana P. Socio-economic and demographic factors related to HIV status in urban informal settlements in the Eastern Cape, South Africa. Afr J AIDS Res 2014 Sep;13(3):271-279. [doi: 10.2989/16085906.2014.952651] [Medline: 25388981]
44. Hollander J. The Social Contexts of Focus Groups. Journal of Contemporary Ethnography 2016 Jul 26;33(5):602-637 [FREE Full text] [doi: 10.1177/0891241604266988]
45. Bergen N, Labonté R, Asfaw S, Mamo A, Abebe L, Kiros G, et al. Social Desirability Bias in Qualitative Research: What is it and what can researchers do about it? 2019 Presented at: QE Symposium; October 9-11, 2019; Brazil.

Abbreviations

ANC: antenatal care
CHW: community health worker
FGD: focus group discussion
LHW: lady health worker
LMIC: low- and middle-income country
mHealth: mobile health
MNCH: maternal, newborn, and child health
PNC: postnatal care
UC: union council

Edited by G Eysenbach; submitted 25.02.20; peer-reviewed by A Kanter, A Khoja; comments to author 11.04.20; revised version received 06.06.20; accepted 11.06.20; published 17.09.20.

Please cite as:
Zaidi S, Kazi AM, Riaz A, Ali A, Najmi R, Jabeen R, Khudadad U, Sayani S
Operability, Usefulness, and Task-Technology Fit of an mHealth App for Delivering Primary Health Care Services by Community Health Workers in Underserved Areas of Pakistan and Afghanistan: Qualitative Study
J Med Internet Res 2020;22(9):e18414
URL: http://www.jmir.org/2020/9/e18414/
doi:10.2196/18414
PMID:32940612

Original Paper

Adherence of HIV Self-Testing Among Men Who Have Sex With Men in China: Longitudinal Study

Xiangyu Yan[1,2*], PhD; Hexuan Su[1,3*], MD; Bo Zhang[1,2], PhD; Yongjie Li[1,2], PhD; Lingling Zhang[4], ScD; Zhongwei Jia[1,2,5,6], PhD

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China

[2]National Institute on Drug Dependence, Peking University, Beijing, China

[3]Medical Informatics Center, Peking University, Beijing, China

[4]College of Nursing and Health Sciences, University of Massachusetts Boston, Boston, MA, United States

[5]Center for Drug Abuse Control and Prevention, National Institute of Health Data Science, Peking University, Beijing, China

[6]Center for Intelligent Public Health, Institute for Artificial Intelligence, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhongwei Jia, PhD
National Institute on Drug Dependence
Peking University
38 Xueyuan Road, Haidian District
Beijing,
China
Phone: 86 10 82802457
Email: urchinjj@163.com

Abstract

Background: The World Health Organization recommended HIV self-testing (HIVST) for individuals practicing unsafe sexual behaviors; however, the adherence to HIV testing has not been reported.

Objective: In this study, we attempted to determine the adherence to HIVST among men who have sex with men (MSM), as well as the impact factors and potential effects of their adherence.

Methods: We conducted a longitudinal study among MSM in Harbin, Heilongjiang province, China from July 1, 2017 to June 30, 2018. A mobile app system was used to provide the "Mailing rapid test reagent kit" for the HIVST service. The proportion of those who adhered to HIV testing every 3 months was calculated. Logistic regression was used to explore the impact factors related to adherence to HIVST. Rates of HIV infection between MSM who adhered to HIVST and those who did not were compared using Cox proportional hazards regression. Changes of condom use behaviors between the two groups were also compared using the chi-square test.

Results: A total of 1315 MSM who received the HIVST service through the app were included in the study. Overall, 10% of the MSM adhered to HIVST, and the proportion of adhering tests was only 34.9%. Adherence of HIVST was associated with marital status (adjusted odds ratio $[OR]_{\text{unmarried vs married}}$ 2.31, 95% CI 1.13-4.71) and the number of HIV tests they received (adjusted $OR_{\text{3 times vs 2 times or below}}$ 3.36, 95% CI 2.01-5.63; adjusted $OR_{\text{4 times or above vs 2 times or below}}$ 7.30, 95% CI 4.67-11.42). Twenty HIV seroconversions were observed during 1-year follow up. The rate of HIV infection in the adherence group (17.10 per 100 person years, 95% CI 8.80-30.84) was significantly higher than that in the nonadherence group (4.80 per 100 person years, 95% CI 2.77-7.88; adjusted hazard ratio 3.33, 95% CI 1.35-8.20). Those who adhered to HIV testing were more likely to improve condom use behaviors, although the difference was not statistically significant.

Conclusions: Regular HIV testing is necessary for early detection of HIV infection among MSM. Given the poor adherence, a new internet-based management paradigm for MSM is needed to raise their health awareness to optimize the implementation of HIVST.

(J Med Internet Res 2020;22(9):e19627) doi:10.2196/19627

KEYWORDS

HIV self-testing; adherence; men who have sex with men; HIV infection; condom use; mobile app

Introduction

Promoting HIV testing is an essential way to achieve the first 90% of the World Health Organization (WHO)'s "90-90-90" target (90% detection, 90% treatment, 90% viral suppression) [1]. However, in 2017, the gap to reach this first 90% remained large, with a worldwide detection rate of HIV infection of 75%, and about 30% of HIV-infected patients in China did not know their status [2,3]. The proportion was even larger among men who have sex with men (MSM); a systematic review indicated that about half of the MSM in China had not been tested for HIV [4]. In 2018, the theme of World AIDS Day was "Know Your Status," which aimed to close the gap and expand HIV testing [5].

For key populations at high risk of HIV exposure, such as MSM, only paying attention to whether or not testing is performed is insufficient because it cannot reflect whether the frequency and regularity of HIV testing are appropriate [6,7]. Regular HIV testing is necessary for the early detection and treatment of HIV infection among MSM [8]. Considering the time window of 3-12 weeks of HIV detection along with the technical advances in HIV rapid test reagents, the Chinese Center for Disease Control and Prevention (CDC) recommended that sexually active MSM undergo HIV testing every 3 months, which is consistent with the recommendation of the US CDC and more frequent than the interval of 6-12 months recommended by the WHO in 2007 [8-10]. Despite the recommendation, it is important to examine the actual adherence to HIV testing among MSM to provide guidance for interventions. The Chinese government has made substantial effort in expanding HIV testing, with the number of HIV testing services provided by medical institutions increasing from 100 million in 2012 to 230 million in 2019 [11]. However, due to the anonymity in providing HIV testing services, it has not been possible to calculate the number of individuals who received HIV testing, let alone the testing frequency and adherence. Moreover, previous epidemiological studies rarely revealed the adherence to HIV testing but rather focused on the HIV testing rate or HIV testing uptake by examining whether MSM received HIV testing or not during the study period [12,13]. In addition, previous cross-sectional studies were not able to evaluate the real-world HIV testing adherence, as calculating this index requires follow-up observation, and participants in cohort studies and trials were influenced by the researchers' reminders for receiving HIV testing and other services to avoid loss to follow up. In addition, questionnaire-based frequency of HIV testing might be a biased estimation, because the reporting bias was hard to measure.

Considering the inconvenience and perceived stigma raised from conventional facility-based HIV testing, HIV self-testing (HIVST) has been supported by 23 countries and was strongly recommended in the 2016 WHO guideline. HIVST is a process in which a person collects his/her own specimen (oral fluid or blood), performs a test, and interprets the result, often in a private setting, either alone or with a trusted person [14]. The "Thirteenth Five-year Plan" (2017-2022) for HIV prevention and control in China also supports HIVST [15,16]. Currently, with the increasing popularity of mobile devices and the rise in online sexual partners–seeking behaviors through mobile apps among MSM, a new method of HIVST named "Mailing rapid test reagent kit" has been conducted around the world. The implementation of this testing method, which combines internet technology with HIV testing and counseling services of community-based organizations, showed good acceptability and feasibility among MSM [17-19]. These findings shed light on the importance of the adherence to HIVST for development of HIVST policy and intervention methods. In the present study, we conducted a real-world longitudinal analysis of use of the "Mailing rapid test reagent kit" app for HIVST. We aimed to examine the adherence to HIV testing among MSM, as well as the impact factors and potential effects of their adherence.

Methods

Study Design and Setting

A longitudinal study was undertaken in Harbin, Heilongjiang province, China from July 1, 2017 to June 30, 2018. The participants were enrolled by the KangTong clinic, which is a community-based organization for MSM in Harbin. This site was chosen because it is the largest and most professional community-based organization in Harbin that could provide HIV testing services for MSM, and was considered to be the most acceptable site among MSM. In addition, the KangTong clinic uses the "Mailing rapid test reagent kit" as the main HIV testing service.

The inclusion criteria of participants were: (1) biologically male, (2) had anal sex with a man at least once in their life, (3) 15 years or older, (4) no difficulty in using the HIVST app, (5) willing to complete questionnaires of the study, and (6) willing to provide written informed consent. Participant recruitment was conducted continuously for a 1-year period. During recruitment, researchers and trained staff of the KangTong clinic publicized the HIVST app to MSM through electronic media (such as WeChat, QQ, and other geosocial networking apps) and offline venues (such as in clinics, parks, and bars) with the help of publicity posters and pictures. The HIVST app's functions were introduced to participants by trained staff. The privacy policy was specifically explained to the participants, including: (1) the encrypted test result reports are only sent to the user through the app by an authorized person at the KangTong clinic; (2) the information collected will be kept confidential and anonymized; (3) if participants need consulting services related to the test results, the consulting service will be implemented in a one-to-one manner by trained staff of the community-based organization, and the process of consulting will also protect participants' privacy. At the beginning of the registration, an electronic informed consent form was provided via the app. After completing registration and becoming familiar with HIVST functions, the community-based organization's staff routinely informed participants of the benefits of HIV

testing and encouraged them to receive the HIVST service through the app at any time as needed; however, no specific interventions on the frequency of using the HIVST service were implemented. Thus, we observed the frequency of using HIVST services based on individual behavior so as to best reflect adherence in a real-world situation.

Based on the frequency and interval between two adjacent HIVSTs, participants who met the following criteria were defined as the adherence group: (1) no interval between two adjacent HIV tests exceeded 100 days (including 100 days) and (2) the period between the date of the last HIV test and endpoint of this study (June 30, 2018) was less than or equal to 100 days. Participants who did not meet at least one of the above criteria were defined as the nonadherence group. However, for those with HIV seroconversion, if the last HIV test result was positive, and no further tests were required once an HIV infection was confirmed, the second criterion of adherence/nonadherence grouping was no longer considered. The threshold of 100 days was set according to the 3-month HIV testing interval recommended by the Chinese CDC, with consideration of weekends and holidays when the mail might be delayed. The primary outcome was adherence to HIV testing among MSM, and we also explored the impact factors of HIV testing adherence. The secondary outcomes were the rate of HIV infection and changes of condom use behaviors between the adherence and nonadherence groups.

Ethics approval for this study was obtained from the Peking University Institutional Review Board (IRB00001052-16016).

HIVST Procedures

A new method of delivering the HIVST service "Mailing rapid test reagent kit" was supported by the KangTong clinic and provided to MSM through a mobile app. The procedure for using the app for HIVST was as follows: (1) click the "Get the test kit" button, and fill in a questionnaire related to sociodemographic characteristics, sexual behaviors, and condom use information; (2) fill in the phone number and address information used to mail the test kit; (3) pay a deposit (¥100, approximately US $14.22) through the app; (4) KangTong clinic sends a package including the HIV rapid test reagent kit (Alere Determine@) and two condoms to the user (the kit was provided free of charge, but an express shipping fee of ¥10 [approximately US $1.42] was required); (5) complete the HIVST using the test kit (a "How to use" button is available on the app including a tutorial video); (6) take a clear picture of the test kit and upload it to the KangTong clinic as required through the app interface (using the "Upload the picture" button); (7) the KangTong clinic generates an HIV testing report for the user in the app system showing the result ("Negative" or "Positive") and date of this HIVST; (8) after receiving the report, the user can click the "Return your deposit" button, and get their deposit back (see Figure 1 and Figure 2). In addition, if a user's test result was HIV positive, the staff of KangTong clinic would contact the user by phone initially. After checking basic information of the user to avoid contacting the wrong person, the staff would explain the meaning of the test result to the user and persuade him to come to the local CDC to confirm the HIV status accompanied by a staff member of the clinic. If the result was confirmed to be HIV-positive, the community-based organization's staff would provide the psychological counseling service and help the user contact the local CDC for further treatment. The participant was informed of the availability of these follow-up services by community-based organization staff at the time of recruitment.

To protect user privacy, the encrypted test results were only sent to the user through the app by an authorized person at the KangTong clinic. The user could view the report after logging into the app. The user's phone number and address were only used to mail the test kit and are kept confidential.

Figure 1. Interface of the "Mailing rapid test reagent kit" app.

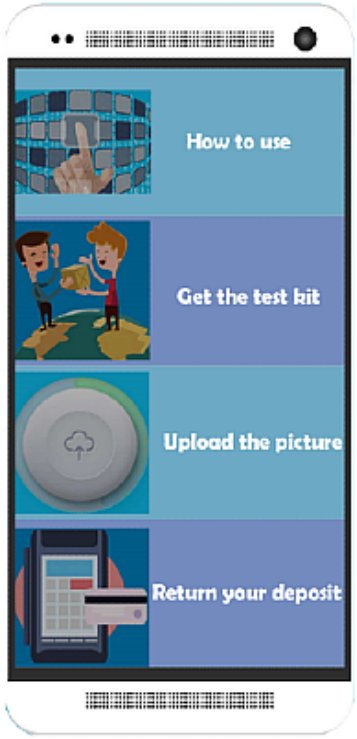
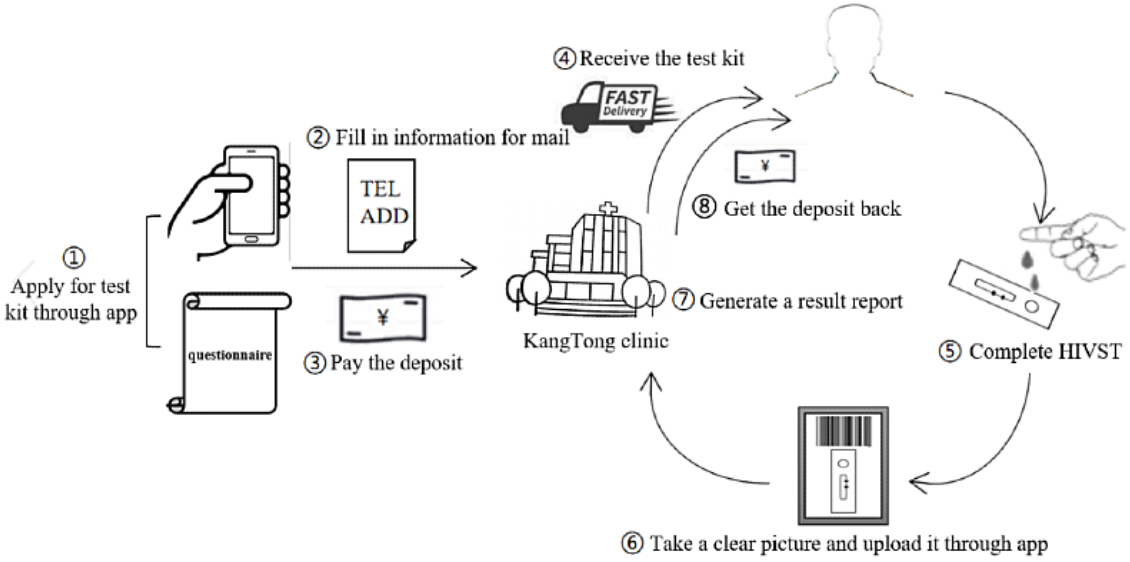


Figure 2. Flow chart of the "Mailing rapid test reagent kit" HIV self-testing (HIVST) service.



Statistical Analysis

We calculated the proportion of MSM who adhered to HIV testing, and then compared characteristics between the adherence and nonadherence groups using the chi-square test. We used multiple imputations to deal with missing data in the variables of age, marital status, and age of first anal sex with a man because there was a less than 10% missing rate for these three variables. Univariate and multivariate logistic regression analyses were used to explore the impact factors of HIV testing adherence. For sensitivity analysis, different multivariate logistic regression models were constructed by adjusting using different variables to verify the stability of impact factors, and different multivariate logistic regression models were also constructed using original data with missing values. As another method to evaluate participants' adherence to HIV testing, we also calculated the proportion of adhering tests. If the interval between a given HIV test and the previous test was shorter than 100 days (including 100 days), the given HIV test was defined as an "adhering test." The proportion of adhering tests was calculated as the total number of adhering tests among participants divided by the total number of tests, excluding their first HIV tests.

Participants with at least two HIV test results were also included in the analysis to calculate the rates of HIV infection in the two

groups. Because the enrollment time of each user was different, the follow-up period for each user was calculated individually. The date of HIV seroconversion was defined as the midway point between the last negative test date and the first positive test date. Univariate and multivariate Cox proportional hazards regression analyses were used to estimate the hazard ratio, reported with 95% CIs and *P* values, to compare the ability to detect HIV seroconversions according to adherence to HIV testing. The multivariate Cox proportional hazards regression model was adjusted for age, condom use behavior, and other sexually transmitted diseases (STDs).

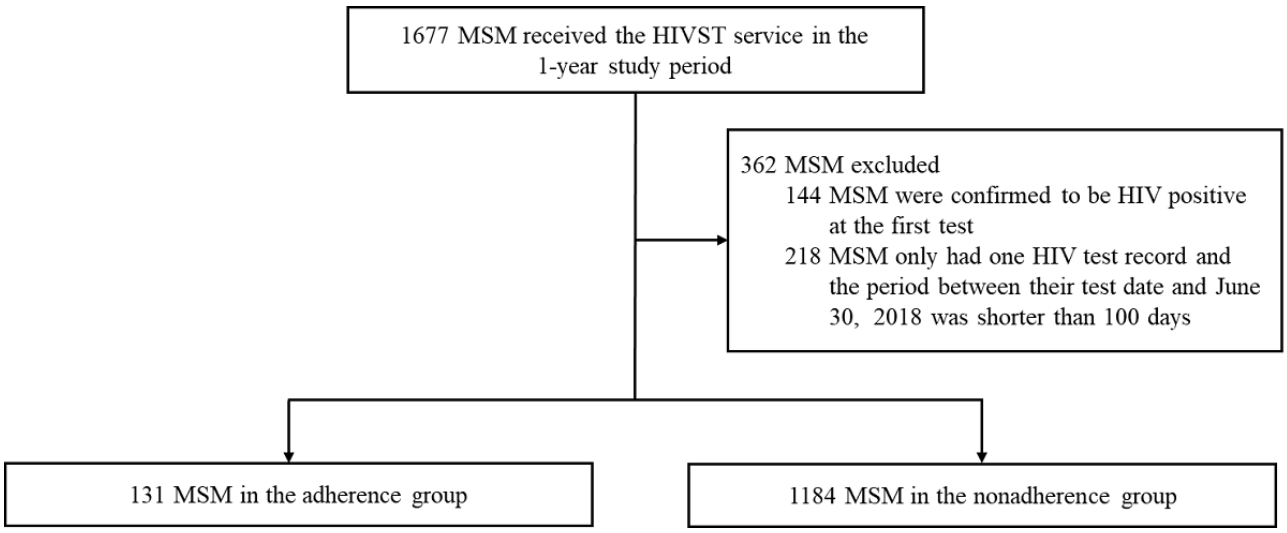
The condom use behavior data were analyzed as a categorical variable based on the questionnaire response about condom use: "Never," "Ever," "Consistent," and "Uncertain." We considered the consistent use of condoms as a protective behavior. We divided the change of condom use behaviors into three categories: "Better," "Worse," and "Other." "Better" means that the users changed their behavior to a more protective category (such as changing from "Never" to "Ever," "Never" to "Consistent," or "Ever" to "Consistent"), or maintained as "Consistent" (including individuals whose baseline information of condom use was uncertain, but consistent use of condom was reported in the follow up). "Worse" means that the users changed the behavior to a more unsafe category (such as changing from "Ever" to "Never," "Consistent" to "Ever," or "Consistent" to "Never"). "Other" means that the users' condom use behaviors were maintained in the less protective categories ("Never" and "Ever") and other conditions with uncertain condom use behavior. Participants with at least two HIV test results were included in the analysis to reflect the change of condom use behaviors. We compared the changes of condom use behaviors between the adherence and nonadherence groups using the chi-square test.

A two-sided *P* value of .05 or less was regarded as significant. The data were verified and statistical analyses were performed with SAS version 9.4 (SAS Institute, Inc., Cary, NC, USA) and SPSS version 21.0 (IBM Corp).

Results

From July 1, 2017 to June 30, 2018, a total of 2702 MSM were recorded as registered users of the app. Among them, 1677 (62.07%) MSM received the HIVST service and uploaded the test results. Data of 1315 app users were included in the final analysis and divided into the adherence group (n=131, 9.96%) and nonadherence group (n=1184, 90.04%) based on the intervals of their HIVSTs. Those who were confirmed to be HIV-positive at the first test (n=144) were excluded, and 218 MSM who only had one HIV test record and the period between their test date and endpoint of this study was shorter than 100 days were also excluded, because we could not confirm whether or not they would adhere to HIV testing (see Figure 3). A total of 2634 HIVSTs were received by the 1315 MSM. Among them, 920 HIVSTs (34.93%) were determined to be adhering tests.

Figure 3. Study profile flowchart. HIVST: HIV self-testing: MSM: men who have sex with men.



As shown in Table 1, the majority of the 1315 MSM were unmarried and had attained higher education of university/college or above. With respect to sexual behaviors, although the majority of the participants reported first having anal sex with a man between 19 and 25 years of age, nearly a third of the participants reported their first such encounter when they were underage (≤18 years). Most of the MSM had always used the internet and some apps to seek sexual partners instead of traditional offline venues. Nearly 10% of the participants reported having had 3 or more casual man sex partners (9.1%) in the past 6 months, and approximately 15% also reported having had sex with women. With respect to drug use behaviors, approximately a quarter of the MSM used drugs (including methamphetamine, Rush Poppers, and other hallucinogens) to enhance sex. Approximately 10% reported a diagnosis of syphilis and other STDs.

Table 1. Basic characteristics of participants.

Characteristic	Total (N=1315), n (%)	Adherence to HIV testing (n=131), n (%)	Nonadherence to HIV testing (n=1184), n (%)	χ^2	df	*P* value
Demographic and social characteristics						
Age (years)				3.451	2	.18
≤25	362 (27.53)	33 (25.2)	329 (27.79)			
26-40	525 (39.92)	46 (35.1)	479 (40.46)			
≥41	428 (32.55)	52 (39.7)	376 (31.76)			
Marital status				3.209	2	.20
Married	222 (16.88)	15 (11.5)	207 (17.48)			
Unmarried	1038 (78.94)	111 (84.7)	927 (78.29)			
Divorced or widowed	55 (4.18)	5 (3.8)	50 (4.22)			
Education				1.816	2	.40
Junior high school or below	188 (14.30)	16 (12.2)	172 (14.53)			
High school	284 (21.60)	24 (18.3)	260 (21.96)			
University/college or above	843 (64.11)	91 (69.5)	752 (63.51)			
Sexual behaviors						
Age of first anal sex with a man (years)				1.309	3	.73
≤18	373 (28.37)	32 (24.4)	341 (28.80)			
19-25	689 (52.40)	71 (54.2)	618 (52.20)			
26-40	203 (15.44)	23 (17.6)	180 (15.20)			
≥41	50 (3.80)	5 (3.8)	45 (3.80)			
Sex role				2.856	2	.24
Insertive only	475 (36.12)	56 (42.7)	419 (35.39)			
Receptive only	322 (24.49)	30 (22.9)	292 (24.66)			
Both	518 (39.39)	45 (34.4)	473 (39.95)			
Place for seeking sexual partners				1.751	1	.19
Offline venues (eg, bars, parks)	269 (20.46)	21 (16.0)	248 (20.95)			
Internet/software/app	1046 (79.54)	110 (84.0)	936 (79.05)			
Number of regular man sexual partners in the past 6 months				5.499	3	.14
0	543 (41.29)	51 (38.9)	492 (41.55)			
1	514 (39.09)	50 (38.2)	464 (39.19)			
2	174 (13.23)	25 (19.1)	149 (12.58)			
3 or above	84 (6.39)	5 (3.8)	79 (6.67)			
Number of casual man sexual partners (with no money transaction) in the past 6 months				3.397	3	.33
0	735 (55.89)	68 (51.9)	667 (56.33)			
1	295 (22.43)	28 (21.4)	267 (22.55)			
2	165 (12.55)	23 (17.5)	142 (11.99)			
3 or above	120 (9.13)	12 (9.2)	108 (9.12)			
Sex with male sex workers in the past 6 months				0.891	1	.35
No	1256 (95.51)	123 (93.9)	1133 (95.69)			
Yes	59 (4.49)	8 (6.1)	51 (4.31)			

Characteristic	Total (N=1315), n (%)	Adherence to HIV testing (n=131), n (%)	Nonadherence to HIV testing (n=1184), n (%)	χ^2	df	*P* value
Sex with women in the past 6 months				1.957	1	.16
No	1109 (84.33)	116 (88.5)	993 (83.87)			
Yes	206 (15.67)	15 (11.5)	191 (16.13)			
Drug use for enhancing sex in the past 6 months				2.313	1	.13
No	1004 (76.35)	93 (71.0)	911 (76.94)			
Yes	311 (23.65)	38 (29.0)	273 (23.06)			
HIV and STDs[a]						
HIV seroconversion				20.429	1	<.001
No	1295 (98.48)	123 (93.9)	1172 (98.99)			
Yes	20 (1.52)	8 (6.1)	12 (1.01)			
Other STDs				0.028	1	.87
No	1180 (89.73)	117 (89.3)	1063 (89.78)			
Yes	135 (10.27)	14 (10.7)	121 (10.22)			
Number of HIVSTs[b]				103.031	2	<.001
2 times or below	992 (75.44)	55 (42.0)	937 (79.14)			
3 times	155 (11.79)	26 (19.8)	129 (10.90)			
4 times or above	168 (12.78)	50 (38.2)	118 (9.97)			

[a]STD: sexually transmitted disease.

[b]HIVST: HIV self-testing.

The baseline characteristics of MSM from the adherence and nonadherence groups were similar, with no significant differences observed in terms of sociodemographic characteristics and sexual behaviors. However, some notable differences were found between the groups, including a larger proportion of HIV seroconversions in the adherence group than in the nonadherence group, and MSM in the adherence group tended to have more HIVSTs compared with those in the nonadherence group (Table 1).

Regression analysis showed that MSM who were unmarried were more likely than those who were married to women to adhere to HIVST. In comparison with MSM who had 2 or less HIVSTs, those who had 3 or more tests were more likely to adhere to HIVST (Table 2). No other significant factors contributing to HIVST adherence were found among the sexual behaviors and STD variables. The results remained stable in sensitivity analysis when variables were adjusted differently in different multivariate logistic regression models (see Multimedia Appendix 1).

At 1-year follow-up, a total of 20 seroconversions were found, including 8 in the adherence group and 12 in the nonadherence group. The rate of HIV infection in the adherence group was significantly higher than that in the nonadherence group (Table 3).

With respect to condom use behaviors, almost half of the MSM used condoms consistently in their recent anal sex with men at baseline at the time of their first HIV tests, and only 1.5% never used a condom. The differences in condom use behaviors between the two groups were not significant at baseline (χ^2_3=3.096, *P*=.38; Table 4). However, most of the MSM had changed the behaviors to a better level or maintained the consistent use of condoms by the 1-year follow-up. In addition, the proportion of MSM whose behaviors became “Better” was higher in the adherence group than that in the nonadherence group when comparing the behaviors at their second HIV tests to the baseline, and the result was similar when comparing the behaviors at their fourth HIV tests to the baseline; although none of the differences between the two groups was statistically significant, the implications are meaningful (Table 5, Multimedia Appendix 2). The proportions of behaviors becoming “Better” in the two groups were similar when comparing the third HIV testing to the baseline (see Table 5, Multimedia Appendix 2).

Table 2. Impact factors of HIV self-testing (HIVST) adherence.

Characteristic	Proportion, % (n/N)[a]	OR[b] (95%CI)	*P* value	AOR[c] (95%CI)	*P* value
Demographic and social characteristics					
Age (years)					
≤25	9.1 (33/362)	1.00	N/A[d]	1.00	N/A
26-40	8.8 (46/525)	0.99 (0.61-1.59)	.96	0.99 (0.58-1.70)	.98
≥41	12.1 (52/428)	1.44 (0.90-2.30)	.13	1.66 (0.96-2.90)	.07
Marital status					
Married	6.8 (15/222)	1.00	N/A	1.00	N/A
Unmarried	10.7 (111/1038)	1.65(0.94-2.88)	.08	2.31 (1.13-4.71)	.02
Divorced or widowed	9.1 (5/55)	1.36 (0.47-3.92)	.57	1.27 (0.41-3.94)	.68
Education					
Junior high school or below	8.5 (16/188)	1.00	N/A	1.00	N/A
High school	8.5 (24/284)	0.99 (0.51-1.92)	.98	0.89 (0.44-1.80)	.74
University/college or above	10.8 (91/843)	1.30 (0.75-2.27)	.35	0.96 (0.52-1.79)	.91
Sexual behaviors					
Age of first anal sex with a man (years)					
≤18	8.6 (32/373)	1.00	N/A	1.00	N/A
19-25	10.3 (71/689)	1.24 (0.80-1.93)	.33	1.23 (0.76-1.98)	.41
26-40	11.3 (23/203)	1.35 (0.77-2.38)	.30	1.69 (0.86-3.31)	.13
≥41	10.0 (5/50)	1.05 (0.36-3.08)	.93	1.55 (0.46-5.20)	.48
Sex role					
Insertive only	11.8 (56/475)	1.00	N/A	1.00	N/A
Receptive only	9.3 (30/322)	0.77 (0.48-1.23)	.27	0.88 (0.52-1.48)	.63
Both	8.7 (45/518)	0.71 (0.47-1.08)	.11	0.79 (0.51-1.24)	.31
Place for seeking sexual partners					
Offline venues (eg, bars, parks)	7.8 (21/269)	1.00	N/A	1.00	N/A
Internet/software/app	10.5 (110/1046)	1.39 (0.85-2.26)	.19	1.50 (0.89-2.54)	.13
Number of regular man sexual partners in the past 6 months					
0	9.4 (51/543)	1.00	N/A	1.00	N/A
1	9.7 (50/514)	1.04 (0.69-1.57)	.85	0.87 (0.54-1.40)	.57
2	14.4 (25/174)	1.62 (0.97-2.70)	.07	1.35 (0.72-2.54)	.36
3 or above	6.0 (5/84)	0.61 (0.24-1.58)	.31	0.50 (0.18-1.39)	.19
Number of casual man sexual partners (with no money transaction) in the past 6 months					
0	9.3 (68/735)	1.00	N/A	1.00	N/A
1	9.5 (28/295)	1.03 (0.65-1.63)	.91	0.92 (0.54-1.55)	.75
2	13.9 (23/165)	1.59 (0.96-2.64)	.07	1.12 (0.62-2.05)	.70
3 or above	10.0 (12/120)	1.09 (0.57-2.08)	.79	1.12 (0.53-2.33)	.77
Had sex with male sex workers in the past 6 months					
No	9.8 (123/1256)	1.00	N/A	1.00	N/A
Yes	13.6 (8/59)	1.45 (0.67-3.11)	.35	1.56 (0.66-3.71)	.31
Had sex with women in the past 6 months					
No	10.5 (116/1109)	1.00	N/A	1.00	N/A

Characteristic	Proportion, % (n/N)[a]	OR[b] (95%CI)	*P* value	AOR[c] (95%CI)	*P* value
Yes	7.3 (15/206)	0.67 (0.38-1.18)	.16	0.84 (0.45-1.58)	.59
Drug use for enhancing sex in the past 6 months					
No	9.3 (93/1004)	1.00	N/A	1.00	N/A
Yes	12.2 (38/311)	1.36 (0.91-2.04)	.13	1.20 (0.75-1.92)	.44
STDs[e]					
Have other STDs					
No	9.9 (117/1180)	1.00	N/A	1.00	N/A
Yes	10.4 (14/135)	1.05 (0.58-1.89)	.87	1.03 (0.54-1.94)	.94
Number of HIVSTs					
2 times or below	5.5 (55/992)	1.00	N/A	1.00	N/A
3 times	16.8 (26/155)	3.43 (2.08-5.67)	<.001	3.36 (2.01-5.63)	<.001
4 times or above	29.8 (50/168)	7.22 (4.70-11.08)	<.001	7.30 (4.67-11.42)	<.001

[a]Proportion=number of participants in adherence group/total number of participants in the category.

[b]OR: odds ratio.

[c]AOR: adjusted odds ratio; all characteristics were adjusted in the multivariate logistic model.

[d]N/A: not applicable.

[e]STD: sexually transmitted disease.

Table 3. Comparison of HIV infection rates of the adherence and nonadherence groups.

Group	Seroconversions	Person years	Rate[a] (95% CI)	HR[b] (95%CI)	*P* value	AHR[c] (95%CI)	*P* value
Total[d]	20	296.69	6.74 (4.38-10.00)	N/A[e]	N/A	N/A	N/A
Nonadherence to HIV testing	12	249.91	4.80 (2.77-7.88)	1.00	N/A	1.00	N/A
Adherence to HIV testing	8	46.77	17.10 (8.80-30.84)	3.48 (1.42-8.54)	.006	3.33 (1.35-8.20)	.009

[a]Rate is calculated as the number of seroconversions per 100 person years.

[b]HR: hazard ratio.

[c]AHR: adjusted hazard ratio; multivariate Cox proportional hazards model adjusted for age, condom use behaviors, and other sexually transmitted diseases.

[d]A total of 680 men who have sex with men with 2 or more HIV tests were included in this analysis.

[e]N/A: not applicable.

Table 4. Comparison of baseline condom use behaviors in the adherence and nonadherence groups.

Condom usage	Total (N=680), n (%)	Adherence to HIV testing (n=131), n (%)	Nonadherence to HIV testing (n=549), n (%)
Never	10 (1.5)	4 (3.1)	6 (1.1)
Ever	162 (23.8)	30 (22.9)	132 (24.0)
Consistent	305 (44.9)	56 (42.7)	249 (45.4)
Uncertain	203 (29.8)	41 (31.3)	162 (29.5)

Table 5. Comparison of condom use behavior changes in the adherence and nonadherence groups.

Testing interval	Adherence to HIV testing, n (%)	Nonadherence to HIV testing, n (%)	χ^2 (df=2)	*P* value
Second test vs first test (n=680)			1.202	.55
Better	65 (49.6)	244 (44.5)		
Worse	15 (11.5)	65 (11.8)		
Other	51 (38.9)	240 (43.7)		
Third test vs first test (n=323)			0.539	.76
Better	31 (40.8)	103 (41.7)		
Worse	12 (15.8)	31 (12.6)		
Other	33 (43.4)	113 (45.7)		
Fourth test vs first test (n=168)			1.109	.57
Better	29 (58.0)	58 (49.1)		
Worse	4 (8.0)	12 (10.2)		
Other	17 (34.0)	48 (40.7)		

Discussion

Principal Findings

In this study, we estimated the HIV adherence among MSM using first-hand data recorded by the app of HIVST, demonstrating that about 90% of MSM do not adhere to HIV testing as recommended by the Chinese CDC without targeted intervention. The proportion was much higher than that lost to follow up (23%-54%) reported by cohort studies and trials that provided HIV testing services [20-22]. In addition, the proportion of adhering tests was only 34.9%. These results indicate the need to focus on the poor HIV testing adherence among MSM.

The negative attitude of MSM toward regular HIVST might be related to certain social and cultural factors. First, MSM may not have sufficient awareness of the need for regular HIV testing because of the lack of related publicity and education [23,24]. The education of HIV testing for MSM has mainly focused on the action to persuade them to be tested and to know their HIV status; however, only focusing on whether they have had HIV test results in their lifetime and the number of tests they received is inadequate, as promoting regular HIV testing is also important [24,25]. Unfortunately, the mobilization for "regular testing" of sexually active people is inadequate all over the world [24-28]. Thus, greater awareness and habit of adhering to HIV testing need to be cultivated. In this study, we found that MSM who received more HIV testing had better HIV testing adherence (adjusted odds $\text{ratio}_{3\text{ times}}$ 3.36, 95% CI 2.01-5.63; adjusted odds $\text{ratio}_{4\text{ times or above}}$ 7.30, 95% CI 4.67-11.42), which showed that improving HIV testing adherence is possible once the habit of regular testing is cultivated, and targeted interventions are needed to focus on MSM who are in the habit-formation period.

Second, MSM in China still face prejudice, discrimination, and stigma associated with their sexual orientation and same-sex behavior [27]. This phenomenon is related to the social and cultural environment of China, although it has improved in recent years, which might still hinder MSMs' adherence to HIV testing. Our study showed that MSM who were unmarried were more likely to adhere to HIV testing than those married to women (adjusted odds ratio 2.31, 95% CI 1.13-4.71). The majority of MSM who were married to women did not want to disclose their male sex behaviors and risk of HIV infection to their families. The stigma and psychological distress from privacy disclosure might prevent them from receiving HIV testing regularly [29].

Third, MSM, especially those with persistent high-risk behaviors, are prone to denial of the risk of HIV infection; thus, frequent testing may introduce greater psychological stress [27]. In addition, several studies have shown that economic factors and testing costs are the main factors restricting the use of testing services [23,27,30]. However, in our study, the testing reagent kits were provided free of charge, and only a small express fee was required, which could effectively avoid the influence of economic and cost factors on testing intentions.

We found that adhering to HIV testing regularly had two potential beneficial effects. First, it could promote case identification of HIV-infected MSM. The rate of HIV infection in the adherence group was 3.3 times higher than that in the nonadherence group (95% CI 1.35-8.20). Given the similar characteristics of sexual behaviors and STDs between the two groups, this effect of case identification was mainly attributable to the timely and regular testing, which could contribute to realization of the first "90%" in the WHO's 90-90-90 target and improve the effectiveness of HIV testing services. Second, adhering to HIV testing might have a potential positive impact on condom use behaviors. There was a tendency for those adhering to HIV testing to be more likely to improve their condom use behavior than those who did not adhere to testing (5% better at their second HIVST, 9% better at their fourth HIVST). Although the difference was not statistically significant, this beneficial effect might be meaningful because

the goal of high-risk behavior reduction among MSM was only 10% in the "Thirteenth Five-year Plan" (2017-2022) for HIV prevention and control in China, and adhering to HIVST within the 1-year period of this study could achieve a 5% reduction in high-risk sexual behavior [9]. Tang and colleagues [31] showed that HIVST was associated with subsequent consistent condom use, which also supports our finding.

Therefore, it is necessary to improve the low level of HIV testing adherence. An internet-based HIVST service such as "Mailing rapid test reagent kit" could be a feasible method to promote the adherence and regularity of HIV testing. The main benefits of such a service include dependable privacy protection, cost-effectiveness, and acceptability [32-34]. In this study, 62.07% (1677/2702) of the registered users of the app received the HIVST service and uploaded the test results, which is much higher than the proportion of 21.4% reported by De Boni et al [35]. In addition, approximately a quarter of registered app users (680/2702) received more than one test. Moreover, the internet-based method could be used as a feasible and utilizable tool to solve the following challenges brought by use of the internet in HIV/AIDS control for MSM and to promote case identification. On the one hand, the internet has gradually become the main source of seeking sexual partners for MSM [36]. Indeed, in our study, about 80% of MSM sought sexual partners using internet software or apps, which are more convenient but also bring about more challenges for HIV transmission [37]. On the other hand, more than a quarter of the MSM in our study (28.4%) first reported having anal sex when they were underage. There is a trend that internet users tend to be younger in China [38]. Quayle and Newman [39] indicated that some images and messages with sexual inducement on the internet offered sexual motivation and interest to users, especially for minors. Therefore, it is necessary to pay attention to the internet-based HIV prevention and testing services for minors. In addition, 23.7% of the MSM in this study reported using drugs for enhancing sex, and recent studies have shown that the internet is facilitating drug trafficking [40]. Drug use could result in the loss of self-control and is often accompanied by high-risk sexual behaviors that increase the risk of HIV transmission [41].

Given that the HIV testing adherence was only 10% in this study, we should consider whether the HIVST service provided needs to be improved by considering aspects of empowerment and feasibility. Although the HIVST package is free and convenient to use, HIVST users may feel greater distress, anxiety, and less social support when testing at home, especially for the first test. However, with the increase of the number of tests, this psychological pressure will be gradually diluted [42], which further highlights the importance of regular testing. Enabling MSM to perform self-testing at home may increase their sense of empowerment or self-assurance by taking charge of their own health [43], and then eventually form a benign positive feedback mechanism. Therefore, targeted improvements of interventions along with internet-based HIVST are essential. More online health education needs to be provided to promote MSMs' awareness of not only HIV testing but also the importance of adhering to HIV testing as far as possible to reduce the threshold of inner stigmas. In addition, improving the participants' empowerment might be a gradual process. To cultivate their habit of adhering to HIV testing, a module can be added to internet-based HIVST tools to automatically remind MSM to receive HIV testing regularly based on their previous test records.

Limitations

Our study has several limitations. First, the 1-year observational period was relatively short, and only 20 seroconversions were found during this time, which might lead to a biased estimation of the rate of HIV infection. Second, our study was only conducted at a single site in an urban area of Harbin and overlooked the HIV testing adherence in rural areas, which will need further investigations. Finally, some people might have purchased an HIV testing kit on their own or received HIV testing in other cities, which might have resulted in underestimating the HIV testing adherence in our study. Nevertheless, our study clearly showed the adherence to the "Mailing rapid test reagent kit" HIVST service in Harbin.

To our knowledge, this is the first study that aimed to focus on adherence to HIV testing in a real-world situation. The characteristics of MSM who adhered to HIV testing regularly in this study could serve as an important external control for future studies and provide good evidence for guideline development in HIV testing services. In future studies, researchers should recruit a larger number of MSM participants at multiple centers with a much longer follow-up period to further estimate HIV testing adherence among MSM.

Conclusions

In this longitudinal study, only 10% of participants adhered to HIVST using the app of "Mailing rapid test reagent kit." Regular HIV testing was shown to be necessary for the early detection of HIV infection and might potentially promote condom use behaviors to some extent. Our study has major practical implications for public health decisions and policymakers. In particular, more attention should be paid to encourage regular HIV testing rather than only recommending HIV testing among sexually active subpopulation at high risk of HIV exposure. HIV testing adherence could become an indicator recommended in guidelines to measure the implementation quality of HIV testing. More feasible and affordable public health management paradigms combined with internet-based methods could be implemented to promote the adherence and regularity of HIV testing.

Acknowledgments

We thank the staff of the KangTong clinic for their contribution in participant recruitment, app daily management, and HIV testing service. This work was supported by the National Natural Science Foundation of China (91546203, 91846302).

Authors' Contributions

XY, HS, and ZJ were responsible for the study design; XY, HS, LZ, and ZJ contributed to writing the report; XY, HS, BZ, and YL contributed to data analysis. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Different models to explore impact factors of HIV self-testing (HIVST) adherence.
[DOCX File , 39 KB - jmir_v22i9e19627_app1.docx]

Multimedia Appendix 2
Follow-up condom use behaviors changes compared with the baseline. AHT: adherence to HIV testing; MSM: men who have sex with men.
[PNG File , 473 KB - jmir_v22i9e19627_app2.png]

References

1. 90-90-90 An ambitious treatment target to help end the AIDS epidemic. UNAIDS. 2014. URL: https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf [accessed 2020-09-09]
2. Miles to go—closing gaps, breaking barriers, righting injustices. UNAIDS. 2018. URL: https://www.unaids.org/sites/default/files/media_asset/miles-to-go_en.pdf [accessed 2020-01-04]
3. Press conference on the progress of HIV/AIDS prevention and treatment in China. National Health Commission of the People's Republic of China. 2018. URL: http://www.scio.gov.cn/xwfbh/gbwxwfbh/xwfbh/wsb/Document/1642083/1642083.htm [accessed 2020-01-04]
4. Zou H, Hu N, Xin Q, Beck J. HIV testing among men who have sex with men in China: a systematic review and meta-analysis. AIDS Behav 2012 Oct;16(7):1717-1728. [doi: 10.1007/s10461-012-0225-y] [Medline: 22677975]
5. World AIDS Day 2018. World Health Organization. 2018. URL: https://www.who.int/campaigns/world-aids-day/2018 [accessed 2020-09-09]
6. Xu J, An M, Han X, Jia M, Ma Y, Zhang M, et al. Prospective cohort study of HIV incidence and molecular characteristics of HIV among men who have sex with men(MSM) in Yunnan Province, China. BMC Infect Dis 2013 Jan 04;13:3 [FREE Full text] [doi: 10.1186/1471-2334-13-3] [Medline: 23286213]
7. Lau JTF, Li D, Wang Z, Lai CHY. Repeated HIV Voluntary Counseling and Testing Increased Risk Behaviors Among Men Who Have Sex with Men in China: A Prospective Cohort Study. AIDS Behav 2015 Nov;19(11):1966-1977. [doi: 10.1007/s10461-015-1034-x] [Medline: 25863465]
8. Guidance on provider-initiated HIV testing and counselling in health facilities. World Health Organization. 2007. URL: https://apps.who.int/iris/bitstream/handle/10665/43688/9789241595568_eng.pdf;jsessionid=6D35E3149B5F5EBB5564985B8CC25115?sequence=1 [accessed 2020-09-09]
9. Guidelines on HIV/AIDS prevention interventions for men who have sex with men. NCAIDS, NCSTD, China CDC. 2016. URL: http://www.chinaaids.cn/jszn/201609/t20160913_134118.htm [accessed 2020-09-09]
10. Centers for Disease Control and Prevention (CDC). HIV testing among men who have sex with men--21 cities, United States, 2008. MMWR Morb Mortal Wkly Rep 2011 Jun 03;60(21):694-699 [FREE Full text] [Medline: 21637183]
11. The AIDS epidemic in China remains at a low epidemic level. National Health Commission of the People's Republic of China. 2019. URL: http://www.gov.cn/xinwen/2019-12/01/content_5457304.htm [accessed 2020-04-04]
12. Qin Y, Tang W, Nowacki A, Mollan K, Reifeis SA, Hudgens MG, et al. Benefits and Potential Harms of Human Immunodeficiency Virus Self-Testing Among Men Who Have Sex With Men in China: An Implementation Perspective. Sex Transm Dis 2017 Apr;44(4):233-238 [FREE Full text] [doi: 10.1097/OLQ.0000000000000581] [Medline: 28282650]
13. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social Media Interventions to Promote HIV Testing, Linkage, Adherence, and Retention: Systematic Review and Meta-Analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
14. Guidelines on HIV self-testing and partner notification- Supplement to consolidated guidelines on HIV testing services. World Health Organization. 2016. URL: https://www.who.int/hiv/pub/self-testing/hiv-self-testing-guidelines/en/ [accessed 2020-09-09]
15. Tang W, Wu D. Opportunities and challenges for HIV self-testing in China. Lancet HIV 2018 Nov;5(11):e611-e612 [FREE Full text] [doi: 10.1016/S2352-3018(18)30244-3] [Medline: 30213725]
16. Thirteenth Five-Year Plan (2017-22) for HIV prevention and control. The State Council of the People's Republic of China. 2017. URL: http://www.gov.cn/zhengce/content/2017-02/05/content_5165514.htm [accessed 2020-01-04]

17. Tao J, Li M, Qian H, Wang L, Zhang Z, Ding H, et al. Home-based HIV testing for men who have sex with men in China: a novel community-based partnership to complement government programs. PLoS One 2014;9(7):e102812 [FREE Full text] [doi: 10.1371/journal.pone.0102812] [Medline: 25051160]
18. Zhong F, Tang W, Cheng W, Lin P, Wu Q, Cai Y, et al. Acceptability and feasibility of a social entrepreneurship testing model to promote HIV self-testing and linkage to care among men who have sex with men. HIV Med 2017 May;18(5):376-382. [doi: 10.1111/hiv.12437] [Medline: 27601301]
19. Wang Z, Lau JTF, Ip M, Ho SPY, Mo PKH, Latkin C, et al. A Randomized Controlled Trial Evaluating Efficacy of Promoting a Home-Based HIV Self-Testing with Online Counseling on Increasing HIV Testing Among Men Who Have Sex with Men. AIDS Behav 2018 Jan;22(1):190-201. [doi: 10.1007/s10461-017-1887-2] [Medline: 28831616]
20. Jia Z, Huang X, Wu H, Zhang T, Li N, Ding P, et al. HIV burden in men who have sex with men: a prospective cohort study 2007-2012. Sci Rep 2015 Jul 02;5:11205. [doi: 10.1038/srep11205] [Medline: 26135810]
21. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: A closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug 28;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
22. Tang W, Huan X, Zhang Y, Mahapatra T, Li J, Liu X, et al. Factors associated with loss-to-follow-up during behavioral interventions and HIV testing cohort among men who have sex with men in Nanjing, China. PLoS One 2015;10(1):e115691 [FREE Full text] [doi: 10.1371/journal.pone.0115691] [Medline: 25559678]
23. Guadamuz TE, Cheung DH, Wei C, Koe S, Lim SH. Young, Online and in the Dark: Scaling Up HIV Testing among MSM in ASEAN. PLoS One 2015 May 14;10(5):e0126658 [FREE Full text] [doi: 10.1371/journal.pone.0126658] [Medline: 25973907]
24. Li R, Pan X, Ma Q, Wang H, He L, Jiang T, et al. Prevalence of prior HIV testing and associated factors among MSM in Zhejiang Province, China: a cross-sectional study. BMC Public Health 2016 Nov 10;16(1):1152 [FREE Full text] [doi: 10.1186/s12889-016-3806-2] [Medline: 27832761]
25. Li B, Bi P, Ward A, Bell C, Fairley CK. Trends and predictors of recent HIV testing over 22 years among a clinic sample of men who have sex with men in South Australia. Sex Health 2017 Apr;14(2):164-169. [doi: 10.1071/SH16091] [Medline: 27832580]
26. Nelson KM, Pantalone DW, Gamarel KE, Carey MP, Simoni JM. Correlates of Never Testing for HIV Among Sexually Active Internet-Recruited Gay, Bisexual, and Other Men Who Have Sex with Men in the United States. AIDS Patient Care STDS 2018 Jan;32(1):9-15 [FREE Full text] [doi: 10.1089/apc.2017.0244] [Medline: 29232170]
27. Ong JJ, Wu D, Huang W, Fu H, Desmond N, Ma W, et al. Pressured HIV testing "in the name of love": a mixed methods analysis of pressured HIV testing among men who have sex with men in China. J Int AIDS Soc 2018 Mar;21(3):e25098 [FREE Full text] [doi: 10.1002/jia2.25098] [Medline: 29577598]
28. Li R, Wang H, Pan X, Ma Q, Chen L, Zhou X, et al. Prevalence of condomless anal intercourse and recent HIV testing and their associated factors among men who have sex with men in Hangzhou, China: A respondent-driven sampling survey. PLoS One 2017 Mar 8;12(3):e0167730 [FREE Full text] [doi: 10.1371/journal.pone.0167730] [Medline: 28273077]
29. Mayer KH, Gangakhedkar R, Sivasubramanian M, Biello KB, Abuelezam N, Mane S, et al. Differing Identities but Comparably High HIV and Bacterial Sexually Transmitted Disease Burdens Among Married and Unmarried Men Who Have Sex With Men in Mumbai, India. Sex Transm Dis 2015 Nov;42(11):629-633 [FREE Full text] [doi: 10.1097/OLQ.0000000000000344] [Medline: 26462187]
30. Creasy SL, Henderson ER, Bukowski LA, Matthews DD, Stall RD, Hawk ME. HIV Testing and ART Adherence Among Unstably Housed Black Men Who Have Sex with Men in the United States. AIDS Behav 2019 Nov 27;23(11):3044-3051. [doi: 10.1007/s10461-019-02647-w] [Medline: 31456200]
31. Tang W, Huang W, Lu H, Cao B, Wu D, Ong J, et al. What happens after HIV self-testing? Results from a longitudinal cohort of Chinese men who have sex with men. BMC Infect Dis 2019 Sep 14;19(1):807 [FREE Full text] [doi: 10.1186/s12879-019-4455-8] [Medline: 31521123]
32. Zhang C, Li X, Brecht M, Koniak-Griffin D. Can self-testing increase HIV testing among men who have sex with men: A systematic review and meta-analysis. PLoS One 2017;12(11):e0188890 [FREE Full text] [doi: 10.1371/journal.pone.0188890] [Medline: 29190791]
33. Lippman SA, Lane T, Rabede O, Gilmore H, Chen Y, Mlotshwa N, et al. High Acceptability and Increased HIV-Testing Frequency After Introduction of HIV Self-Testing and Network Distribution Among South African MSM. J Acquir Immune Defic Syndr 2018 Mar 01;77(3):279-287 [FREE Full text] [doi: 10.1097/QAI.0000000000001601] [Medline: 29210826]
34. Frye V, Wilton L, Hirshfied S, Chiasson MA, Usher D, Lucy D, All About Me Study Team. "Just Because It's Out There, People Aren't Going to Use It." HIV Self-Testing Among Young, Black MSM, and Transgender Women. AIDS Patient Care STDS 2015 Nov;29(11):617-624 [FREE Full text] [doi: 10.1089/apc.2015.0100] [Medline: 26376029]
35. De Boni RB, Veloso VG, Fernandes NM, Lessa F, Corrêa RG, Lima RDS, et al. An Internet-Based HIV Self-Testing Program to Increase HIV Testing Uptake Among Men Who Have Sex With Men in Brazil: Descriptive Cross-Sectional Analysis. J Med Internet Res 2019 Aug 01;21(8):e14145 [FREE Full text] [doi: 10.2196/14145] [Medline: 31373276]

36. Burnham KE, Cruess DG, Kalichman MO, Grebler T, Cherry C, Kalichman SC. Trauma symptoms, internalized stigma, social support, and sexual risk behavior among HIV-positive gay and bisexual MSM who have sought sex partners online. AIDS Care 2016 Oct 13;28(3):347-353 [FREE Full text] [doi: 10.1080/09540121.2015.1096894] [Medline: 26461452]
37. Cruess DG, Burnham KE, Finitsis DJ, Cherry C, Grebler T, Goshe BM, et al. Online Partner Seeking and Sexual Risk Among HIV+ Gay and Bisexual Men: A Dialectical Perspective. Arch Sex Behav 2017 May 26;46(4):1079-1087. [doi: 10.1007/s10508-016-0833-0] [Medline: 27671781]
38. The 43rd China statistical report on internet development. China Internet Network Information Center (CNNIC). 2019. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201902/P020190318523029756345.pdf [accessed 2020-01-04]
39. Quayle E, Newman E. The Role of Sexual Images in Online and Offline Sexual Behaviour With Minors. Curr Psychiatry Rep 2015 Jun 21;17(6):43. [doi: 10.1007/s11920-015-0579-8] [Medline: 25894354]
40. Li B, Bi P, Ward A, Bell C, Fairley CK. Trends and predictors of recent HIV testing over 22 years among a clinic sample of men who have sex with men in South Australia. Sex Health 2017;14(2):164-169. [doi: 10.1071/sh16091]
41. Sun H, Bao Y, Zhou S, Meng S, Lu L. The new pattern of drug abuse in China. Curr Opin Psych 2014;27(4):251-255. [doi: 10.1097/yco.0000000000000073]
42. Liu F, Qin Y, Meng S, Zhang W, Tang W, Han L, et al. HIV self-testing among men who have sex with men in China: a qualitative implementation research study. J Virus Erad 2019 Nov 04;5(4):220-224 [FREE Full text] [Medline: 31754445]
43. Wray TB, Chan PA, Simpanen EM. Longitudinal effects of home-based HIV self-testing on well-being and health empowerment among men who have sex with men (MSM) in the United States. AIDS Care 2020 Feb;32(2):148-154. [doi: 10.1080/09540121.2019.1622636] [Medline: 31146548]

Abbreviations

CDC: Center for Disease Control and Prevention
HIVST: HIV self-testing
MSM: men who have sex with men
STD: sexually transmitted disease
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.04.20; peer-reviewed by Y Fang, F Ferrucci; comments to author 12.06.20; revised version received 06.07.20; accepted 26.07.20; published 17.09.20.

Please cite as:
Yan X, Su H, Zhang B, Li Y, Zhang L, Jia Z
Adherence of HIV Self-Testing Among Men Who Have Sex With Men in China: Longitudinal Study
J Med Internet Res 2020;22(9):e19627
URL: http://www.jmir.org/2020/9/e19627/
doi:10.2196/19627
PMID:32940619

Original Paper

New Path to Recovery and Well-Being: Cross-Sectional Study on WeChat Use and Endorsement of WeChat-Based mHealth Among People Living With Schizophrenia in China

Yu Yu[1,2], PhD; Yilu Li[1], MSc; Tongxin Li[1], MSc; Shijun Xi[1], MSc; Xi Xiao[3], MD; Shuiyuan Xiao[1], MD; Jacob Kraemer Tebes[2], PhD

[1]Department of Social Medicine and Health Management, Xiangya School of Public Health, Central South University, Changsha, China
[2]Division of Prevention and Community Research, Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States
[3]Department of Psychiatry, Changsha Psychiatric Hospital, Changsha, China

Corresponding Author:
Shuiyuan Xiao, MD
Department of Social Medicine and Health Management
Xiangya School of Public Health, Central South University
Upper Mayuanling Road No. 238
Changsha, 410008
China
Phone: 86 0731 84805459
Email: xiaosy@csu.edu.cn

Abstract

Background: The past few decades have seen an exponential increase in using mobile phones to support medical care (mobile health [mHealth]) among people living with psychosis worldwide, yet little is known about WeChat use and WeChat-based mHealth among people living with schizophrenia (PLS) in China.

Objective: This study aims to assess WeChat use, endorsement of WeChat-based mHealth programs, and health related to WeChat use among PLS.

Methods: We recruited a random sample of 400 PLS from 12 communities in Changsha City of Hunan Province, China. WeChat use was assessed using the adapted WeChat Use Intensity Questionnaire (WUIQ). We also compared psychiatric symptoms, functioning, disability, recovery, quality of life, and general well-being between WeChat users and nonusers using one-to-one propensity-score matching.

Results: The WeChat use rate was 40.8% in this sample (163/400); 30.7% (50/163) had more than 50 WeChat friends and nearly half (81/163, 49.7%) spent more than half an hour on WeChat, a pattern similar to college students and the elderly. PLS also showed higher emotional connectedness to WeChat use than college students. About 80.4% (131/163) of PLS were willing to participate in a WeChat-based mHealth program, including psychoeducation (91/163, 55.8%), professional support (82/163, 50.3%), and peer support (67/163, 41.1%). Compared with nonusers, WeChat users were younger, better educated, and more likely to be employed. WeChat use was associated with improved health outcomes, including lower psychiatric symptoms, lower depression, higher functioning, better recovery, and higher quality of life.

Conclusions: WeChat-based mHealth programs hold promise as an empowering tool to provide cost-effective interventions, to foster global recovery, and to improve both physical and mental well-being among PLS. WeChat and WeChat-based mHealth programs have the potential to offer a new path to recovery and well-being for PLS in China.

(J Med Internet Res 2020;22(9):e18663) doi:10.2196/18663

KEYWORDS

WeChat; mHealth; schizophrenia; China; symptoms; functioning; recovery; quality of life; well-being

Introduction

People living with schizophrenia (PLS) are individuals with a diagnosis of schizophrenia, which is a disturbance of thought, perception, and a blunting of affect, which can be characterized by 3 major symptoms: psychosis, cognitive dysfunction, and negative symptoms [1,2]. Recent decades have seen a growing prevalence of mobile phone technology and an increasing interest in using mobile phones to support medical care (mobile health [mHealth]) among PLS worldwide. Mounting evidence shows that PLS own mobile phones and are highly engaged with mHealth [3-5]. A recent meta-analysis shows that 81.4% of PLS owned a mobile phone at some time in the last 2 years, a rate similar to the 90% observed in the general population [3]. The use of mobile phones and the internet among PLS is also similar to that in the general population [5]. The wide accessibility of mobile phones makes mHealth a valuable intervention approach for PLS as it can provide cost-effective, nonstigmatizing, and ongoing support, while overcoming geographic and temporal constraints [3-5]. PLS generally express great interest in mHealth with most in favor of using mobile phones to enhance contact with services and support self-management [3].

A burgeoning literature has consistently demonstrated strong evidence for the feasibility, effectiveness, and efficacy of mHealth [4-9]. For instance, a recent literature review shows an overall retention rate of 92% (95% CI 82%-98%) in mHealth among PLS, as well as a range of potential benefits and satisfaction reported by users [4]. Numerous studies have also shown benefits of mHealth for PLS, such as increased access to health care, remote monitoring and tracking of functioning, the capacity for supplementing and augmenting traditional therapy, enhanced self-monitoring and self-management, increased self-esteem and empowerment, enhanced social interactions and social support, improved medication adherence, decreased stigma due to their condition, relapse prevention, and improved social functioning [6-8].

Given these benefits, mHealth presents new opportunities to promote recovery and well-being among PLS. Recovery is a multifaceted concept that involves the development of new meaning and purpose in one's life as one grew beyond the catastrophic effects of mental illness [10]. For PLS, recovery is not just about symptom reduction and functional improvement, but more an appreciation and satisfaction with life, self-esteem, and improved social functioning [11]. A large number of mHealth programs have been found to promote recovery for people with mental illness, including PLS [12-17].

In China, the most prevalent mobile app is WeChat (literally: micro message) owned by Chinese tech giant Tencent [18]. First released in January 2011, WeChat quickly gathered momentum and popularity, and has become the most widely and extensively used mobile social networking app in China [19]. According to a recent 2019 quarterly report, WeChat boasts over 1.13 billion monthly active users across a wide range of age groups [20], with 93% of urban users logging into WeChat on a daily basis [19]. WeChat has similar characteristics to WhatsApp for message release in various formats (eg, texts, videos, voices, and images) [18]. Additionally, it is similar to Facebook's Newsfeed by enabling members to post text messages, pictures, emojis, web pages, and even small videos to Moments and to give and get comments [18]. Furthermore, WeChat provides additional functions such as entertainment, shopping, payments, banking, and city services, such as paying traffic fines and booking transportation [18]. Collectively, WeChat has seamlessly infiltrated every aspect of daily life for people in China, including PLS, and has generated innovative ways for connection, communication, and interactions as an all-purpose multifunctional app. Because of its high popularity and multiple functions, WeChat has been increasingly utilized as a medium for health interventions, with acceptability, feasibility, and efficacy well-established in people with various health conditions [21-24]. However, to date, most studies have focused on WeChat users from the general population or people with physical illnesses. Although WeChat is being used by PLS, research examining WeChat use rate and endorsement of WeChat-based mHealth among PLS is lacking. Such information may inform the development of WeChat-based mHealth interventions for PLS in China.

The wide recognition of the extensive benefits of mHealth for PLS worldwide, and the lack of research on WeChat use and WeChat-based mHealth among PLS in China represent a significant knowledge gap relevant to mental health services research in China. As the growth in demand for mental health care exceeds the resources available to China's mental health system, it is critical to develop more innovative and cost-effective methods of health care delivery. WeChat and WeChat-based mHealth hold great promise for improving mental health care delivery through extending the reach of services and supplementing existing models of care [3]. However, the potential benefits of WeChat-based mHealth among PLS and their families depend on information on levels of WeChat access and engagement for this population [3]. It is thus essential to learn more about whether and how PLS use WeChat and WeChat-based mHealth if relevant WeChat-based mHealth programs are developed.

This study was conducted to fill this research gap by examining WeChat use, endorsement of WeChat-based mHealth programs, and health outcomes of WeChat users in an urban community sample of PLS in China. Specifically, we first examined participants' WeChat use rate and patterns and compared these with 2 college samples and 1 elderly sample. We then examined participants' endorsement of WeChat-based mHealth programs by assessing their interest in joining various potential WeChat-based programs. Finally, we examined how WeChat use is related to clinical outcomes, personal recovery, and well-being for this population

Methods

Participants and Procedures

This was a cross-sectional study conducted in 12 community health centers from May 2019 to September 2019. All participants were recruited from China's largest demonstration project in mental health services—the "686 Program". The "686 Program" is aimed at integrating hospital and community

services for serious mental illness, with a series of services provided including a monthly free medicine distribution to registered patients [25,26]. Our target population was adult people aged 18 or older with a diagnosis of schizophrenia by the Chinese Classification of Mental Disorders-3 (CCMD-3) or the International Classification of Diseases-10 (ICD-10), living with at least one family member, and able to read and communicate. Those who were younger than 18 years of age with diagnosis other than schizophrenia, living alone, or being unable to read or communicate effectively due to illiteracy or disability were excluded. Using prevalence of WeChat use as our primary outcome variable, sample size was calculated based on the following formula for cross-sectional study: $N=Z^2 \times (P \times [1–P])/E^2$. Assuming $\alpha=.05$ (accordingly, $Z=1.96$), P (prevalence of WeChat use)=0.4, and E (error)=5%, we came with a sample size of 369. In this study, a total of 400 PLS completed the interviews, satisfying the sample size requirement. The sample had a mean age of 46.87 (SD 10.99; range 18-77) and was comparably distributed by gender. Most were unemployed (358/400, 89.5%) and with an education level of middle and high school (271/400, 67.8%). The largest proportion of participants were married or living with partners (172/400, 43.0%), followed by single status (150/400, 37.5%). Most PLS had a diagnosis of schizophrenia for over 10 years (347/400, 86.8%), with a mean duration of 21.42 (SD 10.62) years. Most PLS started treatment and took medication within 1 year of their diagnosis (357/400, 89.3%). All PLS received free standard and unified medication through the 686 Program, which mainly included risperidone, clozapine, and aripiprazole.

The study was approved by the Institutional Review Board of the Xiangya School of Public Health of Central South University. During the monthly free medicine distribution day, a research team of 3-5 psychiatrists went to each health center, where registered people with mental illness receive medication refills. A poster with detailed information about the study was posted in each health center to promote study participation. Individuals who expressed interest in participating in the study completed a clinical assessment about their current symptoms and functioning by 3 psychiatrists and a brief survey by the research team. All participants had the study explained to them and provided written informed consent before participating. Their responses were then checked by a quality control member of the team to ensure there were no inconsistencies or missing items. All participants were reimbursed with RMB 10 (US $1.4) in return for their time for participating.

Measures

WeChat Use

WeChat use was assessed with the WeChat Use Intensity Questionnaire (WUIQ), as adapted by 2 studies (Wen et al [27] and Pang et al [28]) with minor modifications (as in Ellison et al's [29]) from the original assessment tool on Facebook use intensity. It is a 7-item questionnaire measuring the infiltration of WeChat into everyday life and users' emotional attachment to WeChat [27-29]. The first question asks about the number of total WeChat friends one has, with optional answers ranging from 1 (0-50 friends) to 5 (more than 200 friends). The second question asks about the amount of time a person spends on WeChat in a typical day, with options ranging from 1 (0-30 minutes) to 5 (more than 3 hours). The rest of the survey includes a 5-item scale of emotional attachment to WeChat that asks about attitudes toward WeChat use. Each item is rated on a 5-point Likert scale from 1 (totally disagree) to 5 (strongly agree). A WeChat intensity score is then calculated by averaging the total set of question scores. The WUIQ has been increasingly used in China due to the wide popularity of WeChat-based studies, with satisfactory psychometric properties reported [27-29]. In this study, the WUIQ showed good internal consistency, with a Cronbach α of .80.

Psychiatric Symptoms

Psychiatric symptoms were measured with the 18-item Brief Psychiatric Rating Scale (BPRS-18) to assess a set of common symptom characteristics in patients with psychiatric disorders [30]. It covers 5 domains of clinical symptoms: affect, positive symptoms, negative symptoms, resistance, and activation as proposed by Shafer [31]. Each item is rated on an 8-point Likert scale from 0 (not assessed), 1 (not at all) to 7 (extremely severe). The total score ranges from 0 to 126, with higher scores representing greater severity of symptoms. The BPRS-18 has been frequently used in schizophrenia research and has well-established psychometric properties [30,31]. In this study, the BPRS-18 showed good internal consistency, with a Cronbach α of .85.

Functioning

Participant functioning was assessed using the Global Assessment of Functioning (GAF) scale to measure a person's psychological, social, and occupational functioning on a hypothetical continuum of mental health illness ranging from 1 to 100 [32], with examples provided for each 10-level interval. This 1-item scale, with higher scores indicating better functioning, has been widely used in clinical assessments and has satisfactory psychometric properties [33,34]. In this study we assessed the functional level of PLS over the past month.

Disability

The 12-item World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) [35] was used to measure participant's disability and functional impairment. It covers 6 domains of functioning: cognition, mobility, self-care, getting along with people, life activities, and participation in society [35]. Each item is rated on a 5-point Likert scale from 0 (no difficulty) to 4 (extreme difficulty) to assess the level of difficulty experienced while performing the activities. The total score ranges from 0 to 48, with higher score representing higher level of disability [36]. The WHODAS 2.0 has been widely used in China with good psychometric performance established [37,38]. In this study, the WHODAS 2.0 showed good internal consistency, with a Cronbach α of .89.

Depression

Depression was assessed with the Patient Health Questionnaire-9 (PHQ-9) [39] to screen for depressive symptoms by asking whether respondents have experienced a series of depressive symptoms in the past 2 weeks. Each item is rated on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day) to assess the severity degree of depression symptoms. The total score

ranges from 0 to 27, with higher score indicating more depression. The Chinese version of PHQ-9 has also been widely shown to be both culturally acceptable and psychometrically valid [40,41]. In this study, the PHQ-9 showed good internal consistency, with a Cronbach α of .92.

Anxiety

Anxiety was assessed with the Generalized Anxiety Disorder Scale-7 (GAD-7) [42] by asking whether respondents have experienced a series of anxiety symptoms in the past 2 weeks. Each item is rated on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day) to assess the severity degree of anxiety symptoms. The total score ranges from 0 to 27, with higher scores indicating more anxiety symptoms. The Chinese version of GAD-7 has also been widely shown to be both culturally acceptable and psychometrically valid [43,44]. In this study, the GAD-7 showed good internal consistency, with a Cronbach α of .96.

Recovery

Recovery was assessed with the Recovery Assessment Scale (RAS). The RAS is the most widely used scale for measuring a personal perspective on recovery globally, and originally included 5 factors. In this study, we used an 8-item short form of RAS (RAS-8) composed of 2 factors: (1) personal confidence and hope, and (2) no domination by symptoms [45]. Each item is rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The total score ranges from 8 to 40, with higher score indicating better perceived recovery. In this study, the RAS-8 showed good internal consistency, with a Cronbach α of .91.

Quality of Life and General Well-Being

Quality of life and general well-being were measured using the first 2 general questions from the World Health Organization Quality of Life Brief Scale (WHOQOL-BREF) [46]. WHOQOL-BREF is a generic cross-cultural instrument to measure quality of life and is available in more than 40 countries [47]. Quality of life was assessed by asking "How do you evaluate your quality of life in the past two weeks?" on a 5-point scale from 1 (very bad) to 5 (very good). General well-being was assessed by asking "Are you satisfied with your health status?" on a 5-point scale from 1 (very unsatisfied) to 5 (very satisfied).

Statistical Analysis

Scales and indices were tested for reliability, and exploratory and summary statistics were computed for all variables. Data were examined for the presence of missing and influential values, as well as for outliers, skew, and kurtosis. Continuous variables were described using mean and standard deviation, and categorical variables were described using frequency and percentage. WeChat use patterns in this study were compared with those of other populations in other studies using the same or similar scale [27,48,49]. WeChat users and nonusers were compared for sociodemographics by a two-tailed unpaired *t* test for age and χ^2 test for gender, marriage, education, and employment. We also compared a series of health outcomes (symptoms, functioning, disability, depression, anxiety, recovery, quality of life, and general well-being) between WeChat users and nonusers using propensity score matching. Propensity score matching is a widely used method for robust comparison of outcomes between groups by controlling for observed group differences [50]. To control for selection bias, the WeChat use group and nonuse group were 1:1 matched using an optimal matching algorithm on the propensity score. Optimal matching is chosen to retain the maximal number of matched pairs by minimizing the global propensity score distance with replacement, thus minimize sample size loss and improve study external validity [51]. For match tolerance, we used a default value of 1×10^{-5} to check the overlap assumption. Propensity scores were generated using logistic regression with the model variables of age, gender, marital status, education, and employment. Differences in health outcomes can be estimated by conditioning on the estimated propensity score, assuming no unmeasured confounders [50]. All data were analyzed using STATA version 16 (StataCorp).

Results

WeChat Use and Patterns

WeChat use rate was 40.8% in this sample (163/400). As shown in Table 1, among the 163 WeChat users, 50 (30.7%) had more than 50 WeChat friends and nearly half spent more than half an hour on WeChat. Among the 5 items of emotional attachment to WeChat subscale, the top-ranking item was "WeChat is part of my everyday activity" with a mean score of 3.61 (SD 0.94; range 0-5) and for which 58.9% (96/163) of participants indicated agreement; by contrast, the lowest-ranking item was "I feel out of touch when I haven't logged onto WeChat for a day" with a mean score of 3.06 (SD 1.17; range 0-5) and for which 41.7% (68/163) of participants indicated agreement. These results indicate that WeChat has become an important part of daily life for a sizable portion of WeChat users with schizophrenia.

We also compared the WUIQ score of the current PLS sample with 3 other available samples—2 college student samples and 1 elderly sample [27,48,49] in China. PLS had fewer WeChat friends than college students (mean 1.52 vs 2.34/2.44), and the percentage of PLS with more than 50 WeChat friends was similar to that of the elderly (50/163, 30.7% vs 12/35, 34.3%). However, the time PLS spent on WeChat daily was comparable to college students (mean 2.08 vs 2.00/2.77), but the percentage of PLS spending more than 30 minutes daily on WeChat was lower than that of the elderly (81/163, 49.7% vs 27/35, 77.1%). Comparisons on emotional attachment to WeChat were only available for college students, but these showed that PLS generally reported higher emotional attachment to WeChat than college students, with higher mean scores in all 5 items of the emotional attachment to WeChat subscale. Overall, these results indicate that PLS had WeChat friend numbers similar to the elderly, spent time on WeChat similar to college students, and yet had higher emotional attachment to WeChat use than college students.

Table 1. Summary statistics for WeChat intensity score and comparison with other studies.

Variables	People living with schizophrenia (N=163) (this study)	Undergraduates and graduates (N=339) [27]	College students (N=508) [48]	Old persons aged ≥50 (N=35) [49]
Number of WeChat friends[a]				
Continuous, mean (SD)	1.52 (0.99)	2.34 (1.71)	2.44 (1.31)	—[b]
≤50, n (%)	113 (69.33)	125 (36.87)	—	23 (65.71)
>50, n (%)	50 (30.67)	214 (63.13)	—	12 (34.29)
Time spent on WeChat daily[c]				
Continuous, mean (SD)	2.08 (1.38)	2.00 (1.47)	2.77 (1.51)	—
≤30 minutes, n (%)	82 (50.31)	165 (48.67)	—	8 (22.86)
>30 minutes, n (%)	81 (49.69)	174 (51.33)	—	27 (77.14)
WeChat is part of my everyday activity[d], mean (SD)	3.61 (0.94)	3.61 (1.03)	3.48 (0.99)	—
I am proud to tell people I am on WeChat[d], mean (SD)	3.28 (1.18)	2.77 (1.16)	2.74 (0.86)	—
I feel out of touch when I haven't logged onto WeChat for a day[d], mean (SD)	3.06 (1.17)	2.27 (1.13)	2.76 (1.11)	—
I feel I am part of the WeChat community[d], mean (SD)	3.31 (1.01)	3.10 (1.08)	3.04 (1.00)	—
I would be sorry if WeChat shut down[d], mean (SD)	3.28 (1.04)	2.93 (1.25)	3.09 (1.08)	—
Total mean score for all 7 items, mean (SD)	2.83 (0.77)	2.72 (1.26)	2.98 (0.78)	—

[a]Optional answers include: 1=50 or less; 2=51-100; 3=101-150; 4=151-200; 5=more than 200; in the table we reclassified the answers into 2 classes with 50 as cutoff.

[b]Not available.

[c]Optional answers include: 1=less than 30 minutes; 2=30-60 minutes; 3=1-2 hours; 4=2-3 hours; 5=more than 3 hours; in the table we reclassified the answers into 2 classes with 30 minutes as cutoff.

[d]Optional answers include: 1=strongly disagree; 2=disagree; 3=neutral; 4=agree; 5=strongly agree.

Endorsement of WeChat-Based mHealth Programs

Table 2 presents participants' endorsement of WeChat-based mHealth programs. Among all 163 participants, 131 (80.4%) indicated a willingness to participate in any kind of WeChat-based mHealth program. Among the 3 proposed WeChat-based mHealth programs, the most commonly endorsed was psychoeducation (91/163, 55.8%), followed by professional support (82/163, 50.3%) and peer support (67/163, 41.1%). As for the number of WeChat-based mHealth programs participants were willing to participate, the majority were endorsing 1 kind of program (60/163, 36.8%), followed by all 3 kinds (38/163, 23.3%), 2 kinds (33/163, 20.2%), and no program at all (32/163, 19.6%).

Table 2. Endorsement of WeChat-based mHealth programs (N=163).

Variables	n (%)
Are you willing to participate in WeChat-based mHealth programs?	
No	32 (19.6)
Yes	131 (80.4)
Are you willing to participate in WeChat-based psychoeducation?	
No	72 (44.2)
Yes	91 (55.8)
Are you willing to join WeChat-based peer support group?	
No	96 (58.9)
Yes	67 (41.1)
Are you willing to receive WeChat-based professional support?	
No	81 (49.7)
Yes	82 (50.3)
Number of WeChat-based mHealth programs that respondents were willing to participate	
0	32 (19.6)
1	60 (36.8)
2	33 (20.2)
3	38 (23.3)

Sociodemographics Comparison Between WeChat Users and Nonusers

Table 3 compares sociodemographics between WeChat users and nonusers by two-tailed unpaired *t* test for age and χ^2 test for gender, marital status, education, and employment. Significant differences were found in age, education, and employment between WeChat users and nonusers. Compared with nonusers, WeChat users were younger (41.70 vs 50.42, *P*<.001), more educated (college and above: 41/163, 25.2% vs 13/237, 5.5%, *P*<.001), and more likely to be employed (27/163, 16.6% vs 15/237, 6.3%, *P*=.001).

Table 3. Sociodemographic comparison between WeChat users and nonusers.

Characteristic	All respondents (N=400)	WeChat users		*P* value[a]
		No (N=237)	Yes (N=163)	
Age, mean (SD)	46.87 (10.99)	50.42 (0.67)	41.70 (0.77)	<.001[b]
Gender				.611
Male, n (%)	200 (50.00)	121 (51.05)	79 (48.47)	
Female, n (%)	200 (50.00)	116 (48.95)	84 (51.53)	
Marriage				.403
Single, n (%)	150 (37.50)	83 (35.02)	67 (41.10)	
Married/cohabited, n (%)	172 (43.00)	104 (43.88)	68 (41.72)	
Else[c], n (%)	78 (19.50)	50 (21.10)	28 (17.18)	
Education				<.001[b]
Primary and below, n (%)	75 (18.75)	66 (27.85)	9 (5.52)	
Middle and high, n (%)	271 (67.75)	158 (66.67)	113 (69.33)	
College and above, n (%)	54 (13.50)	13 (5.49)	41 (25.15)	
Employment				.001[b]
Unemployed, n (%)	358 (89.50)	222 (93.67)	136 (83.44)	
Employed, n (%)	42 (10.50)	15 (6.33)	27 (16.56)	

[a]Descriptive statistics were compared with chi-square tests for categorical variables (gender, marriage, education, and employment) and unpaired *t* test for continuous variable (age).

[b]Significance at $P<.05$ or $P<.01$.

[c]Else include divorced, separated, and widowed.

Health Outcome Comparison Between WeChat Users and Nonusers

Table 4 compares health outcomes between WeChat users and nonusers by propensity score matching modeled on age, gender, marriage, education, and employment. For health outcomes, significant differences were found in psychiatric symptoms, functioning, depression, recovery, and quality of life. Compared with nonusers, WeChat users had lower scores in psychiatric symptoms (30.47 vs 34.40, $P=.030$) and depression (8.58 vs 9.19, $P=.024$), as well as higher scores in functioning (66.13 vs 59.26, $P<.001$), recovery (23.92 vs 17.30, $P<.001$), and quality of life (3.28 vs 2.96, $P=.002$). In addition, WeChat users had lower disability and higher general well-being scores than nonusers, but the differences only reached a trend level effect after matching. No significant difference was found in anxiety scores between WeChat users and nonusers.

In order to know if health outcomes of WeChat users are further affected by the number of WeChat friends and the time PLS spent on WeChat, we conducted similar comparisons of health outcomes by number of friends (with 50 as cutoff) and time spent on WeChat (with 30 minutes as cutoff) using propensity score matching. Our results showed the greater than 50 WeChat friend group had significantly better functioning than the less than or equal to 50 WeChat friend group (difference=4.61, $P=.024$); they also reported higher a recovery (difference=3.20, $P=.078$), but with only a trend effect. No significant differences on other health outcomes were observed between these WeChat friend groups and the daily WeChat use groups (>30 and ≤30 minutes daily; Results are shown in Appendix 1).

Table 4. Health outcome comparison between the WeChat use group and nonuse group.[a]

Characteristic	All respondents (N=400), mean (SD)	WeChat use group (A), (N=163), mean (SD)	Non-use full group (B) (N=237), mean (SD)	(C) = A–B, *P* value (before matching)	Non-use matched group (D) (N=163), mean (SD)	(E) = A–D, *P* value (after matching)
BPRS-18[b]	32.90 (11.43)	30.47 (10.41)	34.53 (11.81)	<.001[c]	34.40 (10.72)	.030[c]
WHODAS 2.0[d]	26.02 (10.22)	23.49 (9.03)	27.74 (10.63)	<.001[c]	27.10 (9.31)	.080[e]
GAF[f]	61.83 (13.58)	66.13 (13.09)	58.97 (13.18)	<.001[c]	59.26 (11.61)	<.001[c]
PHQ-9[g]	9.01 (7.54)	8.58 (7.48)	9.30 (7.58)	.352	9.19 (6.52)	.024[c]
GAD-7[h]	6.67 (6.43)	6.08 (6.17)	7.08 (6.58)	.133	6.20 (5.67)	.101
RAS-8[i]	20.29 (9.31)	23.92 (9.07)	17.80 (8.65)	<.001[c]	17.30 (8.17)	<.001[c]
QOL-1[j]	3.05 (0.90)	3.28 (0.89)	2.89 (0.88)	<.001[c]	2.96 (0.82)	.002[c]
QOL-2[j,e]	3.02 (0.95)	3.16 (0.96)	2.92 (0.94)	.014[c]	3.00 (0.87)	.065[e]

[a]Continuous variables were compared using independent 2-sample unpaired *t* test for unmatched samples and paired *t* test for matched sample.

[b]BPRS-18: the 18-item Brief Psychiatric Rating Scale

[c]Significant at *P*<.05 or *P*<.01.

[d]WHODAS 2.0: the 12-item World Health Organization Disability Assessment Schedule 2.0

[e]Trend effect at *P*<.10.

[f]GAF: Global Assessment of Functioning

[g]PHQ-9: Patient Health Questionnaire-9

[h]GAD-7: Generalized Anxiety Disorder Scale-7

[i]RAS: Recovery Assessment Scale

[j]QOL: quality of life

Discussion

Summary

This study provides a first examination of WeChat use patterns, endorsement of WeChat-based mHealth programs, and health outcomes related to WeChat use among an urban community sample of PLS in China. We also compared use patterns and intensity of use with comparable data available for other groups, such as for college students or an elderly sample. Our findings show a promising WeChat use rate of 40.8% (163/400) among this sample, whose WeChat friend number is similar to that of an elderly sample. Although PLS spent comparable time on WeChat to the college student samples, they showed a higher level of emotional connectedness to WeChat use than college students. About 80.4% (131/163) of PLS were willing to participate in any kind of WeChat-based mHealth programs, with psychoeducation being the most commonly endorsed program (91/163, 55.8%). Compared with nonusers, WeChat users were younger, better educated, and more likely to be employed. WeChat use was also associated with improved health outcomes, including lower psychiatric symptoms, lower depression, higher functioning, better recovery, and higher quality of life. This suggests that WeChat users among PLS may represent a higher functioning group that are particularly amenable to mHealth programs using WeChat.

WeChat Use and Patterns

In this study, WeChat use rate was 40.8% (163/400) among PLS, a rate within the range of 27%-71% for social media use reported among PLS in various studies across the world [52-57]. Thus, WeChat use among PLS in China is consistent with other popular social media use rates among PLS reported globally. When compared with 2 college student samples and 1 elderly sample, PLS had similar number of friends to the elderly, and spent similar time on WeChat to college students, but reported stronger emotional connectedness to WeChat use. The results extend the growing body of research showing that people with serious mental illness, including schizophrenia, are heavy consumers of internet and social media content [5,58,59]. PLS's high emotional connectedness to WeChat may be related to negative stereotypes attached to mental illness that disrupt direct social contacts but makes the internet and social media a potentially safer alternative to meet social needs [60]. As a result, the WeChat use rate and emotional connectedness to WeChat use among PLS suggests that WeChat may be a widely accepted and appealing platform for health interventions in China.

Endorsement of WeChat-Based mHealth Programs

The finding that 80.4% (131/163) of PLS were willing to participate in any kind of WeChat-based mHealth programs is consistent with past literature showing high endorsement of mHealth intervention among PLS and other persons with other psychoses [3]. This finding indicates a wide acceptability of WeChat-based mHealth programs for people with mental illness in China. Among the 3 proposed WeChat-based mHealth programs, the most commonly endorsed was psychoeducation (91/163, 55.8%), followed by professional support (82/163,

50.3%), and peer support (67/163, 41.1%). These results are consistent with the literature showing psychoeducation, peer support, and professional support as the top 3 promising and feasible interventions to effectively improve prognosis and well-being among PLS [61-63]. These findings have implications for future WeChat-based mHealth programs to first assess participants' preferences and then provide targeted interventions based on their preferences.

WeChat Use and Sociodemographic

Consistent with most of previous studies, our study finds that WeChat users are generally of younger age, have higher education, and being employed than nonusers. That younger people are more likely to use WeChat in China is constituent with worldwide data that younger people are more likely to get online and use social media, such as Instagram, Snapchat, and Facebook [64-66]. That people with more education were more likely to use WeChat is also consistent with other research showing positive associations between higher education and social media use [5,66]. For instance, the most recent research by the Pew Research Center showed an internet use rate of 79% among people with a college degree, which is much higher than the rate of 64% among people with a high-school education or less [64]. In addition, employed people were more likely to use WeChat which may be indicative of their higher economic status, thus making mobile phones and WeChat more accessible to them in general, as evidence shows that availability and access to mHealth are accounted for by economic rather than disease factors [3].

WeChat Use and Health Outcomes

We found generally better health outcomes in WeChat users than nonusers, including lower psychiatric symptoms, lower depression, higher functioning, better recovery, and higher quality of life. While higher number of WeChat friends was associated with higher functioning and better recovery, daily time spent on WeChat use showed no significant relationship to any of the health outcomes in this study. It seems like WeChat use alone, regardless of WeChat use intensity, was associated with better health outcomes. This finding aligns with the growing literature showing that social media use is associated with improved clinical and psychosocial outcomes among PLS and other persons with other psychoses, with reports of fewer symptoms, better functioning, better recovery, and better quality of life [52,58,67]. One potential mechanism of better health outcomes among WeChat users may be through a larger social network and more social support [18,27]. This hypothesis is partially supported by our findings showing higher functioning and better recovery among the over 50 WeChat friends group. Further research is still needed to directly measure and compare social networks and social support between WeChat users and nonusers. It is also likely that the association between WeChat use and health outcomes may influence each other in a bidirectional way. For instance, the positive association between social functioning and WeChat use corresponds to an underdeveloped yet important conceptualization of social media participation as a dimension of social functioning, as raised by Bjornestad et al [58]. In this recent review, Bjornestad et al [58] found that only 1 out of the total 58 identified social functioning measures included social media as a social activity. They thus proposed that social media use be included into social functioning measurement in the future to avoid inadequate clinical assessment of social functioning.

Limitations

This study has several limitations. First, our sample was drawn from 12 urban communities and may not be generalizable to other locales such as rural communities, where mobile phone and WeChat access may be lower. As a result, the findings in this study may not capture the whole picture of social media use among PLS in China. Future research may benefit from conducting a nationally representative sample of WeChat users who are PLS. Second, when comparing WeChat use intensity, we used 2 college student samples and 1 elderly sample as comparison groups, instead of the general population or similar PLS samples. This was due to the lack of similar WeChat use intensity data on these populations. We expect more studies on WeChat use among various populations, including PLS using the WUIS in the future to allow for more cross-comparisons. Third, the cross-sectional study design did not make it possible to establish causality between WeChat use and health outcomes. Future research should examine WeChat use longitudinally and examine the relationship between use and well-being. Fourth, WeChat, like any other social media, may carry the potential risk of violating personal privacy and confidentiality, which should be taken into consideration when designing WeChat-based mHealth research and interventions. All participants in research or interventions should be made aware of the potential privacy issues pertaining to their WeChat data and provide informed consent prior to participation.

Conclusions and Implications

This initial study provides new data on the relationship between social media use using WeChat, mHealth program interest, and characteristics of WeChat users (PLS) in China. Our findings show a promising WeChat use rate and wide acceptability of WeChat-based mHealth programs among this population. This finding has implications for enhancing the current community-based treatment of PLS in China to augment existing treatment programs with WeChat-based mHealth interventions. Such interventions hold promise for reaching a larger population with schizophrenia, especially in regions of the world where traditional resources are scarce [3]. We also found that WeChat use was related to a series of positive health outcomes including decreased symptoms and depression, as well as improved functioning, recovery, and quality of life. This finding offers early validation of the interest and opportunities for leveraging popular social media platforms, such as WeChat, for supporting the health and well-being of PLS. WeChat-based mHealth programs can be an empowering tool to provide cost-effective interventions, to foster recovery and to improve both physical and mental well-being among PLS. WeChat and WeChat-based mHealth programs have the potential to lead to a new path to recovery and well-being for PLS in China. Finally, future research should build on the findings from this study to develop a WeChat-based integrative family intervention program that includes 3 key components surveyed in this study: psychoeducation, peer support, and professional support. Such

an intervention could provide support and training for PLS and family members to facilitate recovery and improve the well-being of PLS directly as well as indirectly through improved quality of care provided by family caregivers [68].

Acknowledgments

This work was funded by a grant from the National Natural Science Foundation of China (Grant Number 71804197) and the China Scholarship Council (CSC) in support of the first author. The authors thank all PLS and their families for their participation in the study. We are also grateful to Changsha Psychiatric Hospital as well as the 12 community health centers for their collaboration and support.

Authors' Contributions

All authors have made substantial contributions to the study conception and design, data collection and analysis, and to the development and editing of the manuscript. The principal investigator (YY) led the initial study design, while S Xiao and JT substantially revised and updated the research question and study design prior to initiating the project. YY, YL, TL, S Xi, and XX contributed to the research conduction and data collection; YY and YL contributed to data analyses; TL, S Xi, XX, S Xiao, and JT contributed to data interpretation; YY drafted the article while TL, S Xi, XX, S Xiao, and JT critically appraised it and revised it. All authors approved the final version of manuscript for submission and publication.

Conflicts of Interest

None declared.

References

1. Austin J. Schizophrenia: an update and review. J Genet Couns 2005 Oct;14(5):329-340. [doi: 10.1007/s10897-005-1622-4] [Medline: 16195940]
2. Coffey M. Schizophrenia: a review of current research and thinking. J Clin Nurs 1998 Nov;7(6):489-498. [doi: 10.1046/j.1365-2702.1998.00204.x] [Medline: 10222943]
3. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile Phone Ownership and Endorsement of "mHealth" Among People With Psychosis: A Meta-analysis of Cross-sectional Studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
4. Firth J, Torous J. Smartphone Apps for Schizophrenia: A Systematic Review. JMIR Mhealth Uhealth 2015;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
5. Villani M, Kovess-Masfety V. How Do People Experiencing Schizophrenia Spectrum Disorders or Other Psychotic Disorders Use the Internet to Get Information on Their Mental Health? Literature Review and Recommendations. JMIR Ment Health 2017 Jan 03;4(1):e1 [FREE Full text] [doi: 10.2196/mental.5946] [Medline: 28049620]
6. Välimäki M, Athanasopoulou C, Lahti M, Adams CE. Effectiveness of Social Media Interventions for People With Schizophrenia: A Systematic Review and Meta-Analysis. J Med Internet Res 2016;18(4):e92 [FREE Full text] [doi: 10.2196/jmir.5385] [Medline: 27105939]
7. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
8. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804. [doi: 10.1177/0145445515595198] [Medline: 26187164]
9. Hategeka C, Ruton H, Law MR. Effect of a community health worker mHealth monitoring system on uptake of maternal and newborn health services in Rwanda. Glob Health Res Policy 2019;4:8 [FREE Full text] [doi: 10.1186/s41256-019-0098-y] [Medline: 30949586]
10. Anthony WA. Recovery from mental illness: The guiding vision of the mental health service system in the 1990s. Psychosocial Rehabilitation Journal 1993 Apr;16(4):11-23. [doi: 10.1037/h0095655]
11. Liberman RP. Review of Recovery from disability: Manual of psychiatric rehabilitation. In: Recovery From Disability: Manual of Psychiatric Rehabilitation. Washington, DC: American Psychiatric Association; 2009:68-69.
12. Ben-Zeev D, Brian RM, Jonathan G, Razzano L, Pashka N, Carpenter-Song E, et al. Mobile Health (mHealth) Versus Clinic-Based Group Intervention for People With Serious Mental Illness: A Randomized Controlled Trial. Psychiatr Serv 2018 May 25:appips201800063. [doi: 10.1176/appi.ps.201800063] [Medline: 29793397]
13. Curtis B, Bergman B, Brown A, McDaniel J, Harper K, Eisenhart E, et al. Characterizing Participation and Perceived Engagement Benefits in an Integrated Digital Behavioral Health Recovery Community for Women: A Cross-Sectional Survey. JMIR Ment Health 2019 Aug 26;6(8):e13352 [FREE Full text] [doi: 10.2196/13352] [Medline: 31452520]

14. Jonathan G, Carpenter-Song EA, Brian RM, Ben-Zeev D. Life with FOCUS: A qualitative evaluation of the impact of a smartphone intervention on people with serious mental illness. Psychiatr Rehabil J 2019 Jun;42(2):182-189. [doi: 10.1037/prj0000337] [Medline: 30589278]
15. Noel VA, Acquilano SC, Carpenter-Song E, Drake RE. Use of Mobile and Computer Devices to Support Recovery in People With Serious Mental Illness: Survey Study. JMIR Ment Health 2019 Feb 20;6(2):e12255 [FREE Full text] [doi: 10.2196/12255] [Medline: 30785401]
16. Varnfield M, Rajesh K, Redd C, Gibson S, Gwillim L, Polkinghorne S. Health-e Minds: a Participatory Personalised and Gamified mHealth Platform to Support Healthy Living Behaviours for People with Mental Illness. Conf Proc IEEE Eng Med Biol Soc 2019:6943-6947. [doi: 10.1109/embc.2019.8857286] [Medline: 31947436]
17. Vitger T, Austin SF, Petersen L, Tønder ES, Nordentoft M, Korsbek L. The Momentum trial: the efficacy of using a smartphone application to promote patient activation and support shared decision making in people with a diagnosis of schizophrenia in outpatient treatment settings: a randomized controlled single-blind trial. BMC Psychiatry 2019 Jun 17;19(1):185 [FREE Full text] [doi: 10.1186/s12888-019-2143-2] [Medline: 31208376]
18. Montag C, Becker B, Gan C. The Multipurpose Application WeChat: A Review on Recent Research. Front Psychol 2018;9:2247 [FREE Full text] [doi: 10.3389/fpsyg.2018.02247] [Medline: 30618894]
19. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
20. WeChat: users 2019 statistic. URL: http://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/
21. Zhu R, Xu X, Zhao Y, Sharma M, Shi Z. Decreasing the use of edible oils in China using WeChat and theories of behavior change: study protocol for a randomized controlled trial. Trials 2018 Nov 16;19(1):631 [FREE Full text] [doi: 10.1186/s13063-018-3015-7] [Medline: 30445989]
22. Zhang X, Xiao H. Development and evaluation of a WeChat-based life review programme for patients with cancer: protocol for a randomised controlled trial. BMJ Open 2018 Dec 06;8(12):e020239 [FREE Full text] [doi: 10.1136/bmjopen-2017-020239] [Medline: 30530470]
23. Ma D, Cheng K, Ding P, Li H, Wang P. Self-management of peripherally inserted central catheters after patient discharge via the WeChat smartphone application: A systematic review and meta-analysis. PLoS One 2018 Aug 28;13(8):e0202326 [FREE Full text] [doi: 10.1371/journal.pone.0202326] [Medline: 30153253]
24. Liu J, Zheng X, Chai S, Lei M, Feng Z, Zhang X, et al. Effects of using WeChat-assisted perioperative care instructions for parents of pediatric patients undergoing day surgery for herniorrhaphy. Patient Educ Couns 2018 Aug;101(8):1433-1438. [doi: 10.1016/j.pec.2018.02.010] [Medline: 29499997]
25. Good BJ, Good MD. Significance of the 686 Program for China and for global mental health. Shanghai Arch Psychiatry 2012 Jun;24(3):175-177 [FREE Full text] [doi: 10.3969/j.issn.1002-0829.2012.03.008] [Medline: 25324623]
26. Ma H. Integration of hospital and community services-the '686 Project'-is a crucial component in the reform of China's mental health services. Shanghai Arch Psychiatry 2012 Jun;24(3):172-174 [FREE Full text] [doi: 10.3969/j.issn.1002-0829.2012.03.007] [Medline: 25324622]
27. Wen Z, Geng X, Ye Y. Does the Use of WeChat Lead to Subjective Well-Being?: The Effect of Use Intensity and Motivations. Cyberpsychol Behav Soc Netw 2016 Oct;19(10):587-592. [doi: 10.1089/cyber.2016.0154] [Medline: 27732075]
28. Pang H. Is mobile app a new political discussion platform? An empirical study of the effect of WeChat use on college students' political discussion and political efficacy. PLoS One 2018;13(8):e0202244 [FREE Full text] [doi: 10.1371/journal.pone.0202244] [Medline: 30092076]
29. Ellison NS, Lampe C. The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Site. Journal of Computer-Mediated Communication 2007 Jul 1;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
30. Dazzi F, Shafer A, Lauriola M. Meta-analysis of the Brief Psychiatric Rating Scale - Expanded (BPRS-E) structure and arguments for a new version. J Psychiatr Res 2016 Oct;81:140-151. [doi: 10.1016/j.jpsychires.2016.07.001] [Medline: 27451107]
31. Shafer A. Meta-analysis of the brief psychiatric rating scale factor structure. Psychol Assess 2005 Sep;17(3):324-335. [doi: 10.1037/1040-3590.17.3.324] [Medline: 16262458]
32. Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. Am J Psychiatry 1992 Sep;149(9):1148-1156. [Medline: 1386964]
33. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV). Washington DC: American Psychiatric Association; 1994.
34. Endicott J, Spitzer RL, Fleiss JL, Cohen J. The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. Arch Gen Psychiatry 1976 Jun;33(6):766-771. [doi: 10.1001/archpsyc.1976.01770060086012] [Medline: 938196]
35. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol 1996 Jun;11 Suppl 3:89-95. [doi: 10.1097/00004850-199606003-00015] [Medline: 8923116]

36. World Health Organization. WHO Disability Assessment Schedule 2.0 (WHODAS2.0). Geneva, Switzerland: World Health Organization; 2017. URL: http://www.who.int/classifications/icf/whodasii/en/ [accessed 2020-09-12]
37. Chiu T, Yen C, Chou C, Lin J, Hwang A, Liao H, et al. Development of traditional Chinese version of World Health Organization disability assessment schedule 2.0 36--item (WHODAS 2.0) in Taiwan: validity and reliability analyses. Res Dev Disabil 2014 Nov;35(11):2812-2820. [doi: 10.1016/j.ridd.2014.07.009] [Medline: 25094056]
38. Yen C, Hwang A, Liou T, Chiu T, Hsu H, Chi W, et al. Validity and reliability of the Functioning Disability Evaluation Scale-Adult Version based on the WHODAS 2.0--36 items. J Formos Med Assoc 2014 Nov;113(11):839-849 [FREE Full text] [doi: 10.1016/j.jfma.2014.08.008] [Medline: 25294100]
39. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
40. Zhang L, Fritzsche K, Liu Y, Wang J, Huang M, Wang Y, et al. Validation of the Chinese version of the PHQ-15 in a tertiary hospital. BMC Psychiatry 2016 Apr 05;16:89 [FREE Full text] [doi: 10.1186/s12888-016-0798-5] [Medline: 27044309]
41. Liu Z, Yu Y, Hu M, Liu H, Zhou L, Xiao S. PHQ-9 and PHQ-2 for Screening Depression in Chinese Rural Elderly. PLoS One 2016;11(3):e0151042 [FREE Full text] [doi: 10.1371/journal.pone.0151042] [Medline: 26978264]
42. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
43. Zhou Y, Cao Z, Yang M, Xi X, Guo Y, Fang M, et al. Comorbid generalized anxiety disorder and its association with quality of life in patients with major depressive disorder. Sci Rep 2017 Jan 18;7:40511 [FREE Full text] [doi: 10.1038/srep40511] [Medline: 28098176]
44. Tong X, An D, McGonigal A, Park S, Zhou D. Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. Epilepsy Res 2016 Feb;120:31-36. [doi: 10.1016/j.eplepsyres.2015.11.019] [Medline: 26709880]
45. He SF, Huang Z, Li T, Tebes J, Xiao S, Yu Y. Validation of an 8-item Recovery Assessment Scale (RAS-8) for persons living with schizophrenia in China. Journal of Psychopathology and Behavioral Assessment 2020 (forthcoming).
46. World Health Organization. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
47. Lodhi FS, Montazeri A, Nedjat S, Mahmoodi M, Farooq U, Yaseri M, et al. Assessing the quality of life among Pakistani general population and their associated factors by using the World Health Organization's quality of life instrument (WHOQOL-BREF): a population based cross-sectional study. Health Qual Life Outcomes 2019 Jan 14;17(1):9 [FREE Full text] [doi: 10.1186/s12955-018-1065-x] [Medline: 30642360]
48. Pang H. WeChat use is significantly correlated with college students' quality of friendships but not with perceived well-being. Heliyon 2018 Nov;4(11):e00967 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e00967] [Medline: 30519660]
49. Guo L. WeChat as a Semipublic Alternative Spherexploring the Use of WeChat Among Chinese Older Adults. International Journal of Communication 2017;11:408-428.
50. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009 Nov 10;28(25):3083-3107 [FREE Full text] [doi: 10.1002/sim.3697] [Medline: 19757444]
51. Bergstralh EK, Jacobsen S. Software for optimal matching in observational studies. Epidemiology 1996;7(3):331-332. [doi: 10.4135/9781412953948.n327]
52. Brusilovskiy E, Townley G, Snethen G, Salzer MS. Social media use, community participation and psychological well-being among individuals with serious mental illnesses. Computers in Human Behavior 2016 Dec;65:232-240. [doi: 10.1016/j.chb.2016.08.036]
53. Rekhi G, Ang MS, Lee J. Clinical determinants of social media use in individuals with schizophrenia. PLoS One 2019;14(11):e0225370 [FREE Full text] [doi: 10.1371/journal.pone.0225370] [Medline: 31747434]
54. Bonet LL, Hernandez M, Arce D, Blanquer I, Cañete C. S249. Is internet harmful for psychotic patients? Schizophrenia Bulletin 2018;(suppl_1):44.
55. Naslund JA, Aschbrenner KA, Bartels SJ. How people with serious mental illness use smartphones, mobile apps, and social media. Psychiatr Rehabil J 2016 Dec;39(4):364-367 [FREE Full text] [doi: 10.1037/prj0000207] [Medline: 27845533]
56. Trefflich F, Kalckreuth S, Mergl R, Rummel-Kluge C. Psychiatric patients' internet use corresponds to the internet use of the general public. Psychiatry Res 2015 Mar 30;226(1):136-141. [doi: 10.1016/j.psychres.2014.12.037] [Medline: 25623020]
57. Miller BJ, Stewart A, Schrimsher J, Peeples D, Buckley PF. How connected are people with schizophrenia? Cell phone, computer, email, and social media use. Psychiatry Res 2015 Feb 28;225(3):458-463. [doi: 10.1016/j.psychres.2014.11.067] [Medline: 25563669]
58. Bjornestad J, Hegelstad WTV, Berg H, Davidson L, Joa I, Johannessen JO, et al. Social Media and Social Functioning in Psychosis: A Systematic Review. J Med Internet Res 2019 Jun 28;21(6):e13957 [FREE Full text] [doi: 10.2196/13957] [Medline: 31254338]

59. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for Future Mobile-Health (mHealth): A Systematic Review of Mobile Phone and Web-Based Text Messaging in Mental Health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
60. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
61. Zapata Ospina JP, Rangel Martínez-Villalba AM, García Valencia J. [Psychoeducation in schizophrenia]. Rev Colomb Psiquiatr 2015;44(3):143-149. [doi: 10.1016/j.rcp.2015.03.005] [Medline: 26578414]
62. Parker Oliver D, Patil S, Benson JJ, Gage A, Washington K, Kruse RL, et al. The Effect of Internet Group Support for Caregivers on Social Support, Self-Efficacy, and Caregiver Burden: A Meta-Analysis. Telemed J E Health 2017 Aug;23(8):621-629. [doi: 10.1089/tmj.2016.0183] [Medline: 28328392]
63. Sorrell JM. Moving beyond caregiver burden: identifying helpful interventions for family caregivers. J Psychosoc Nurs Ment Health Serv 2014 Mar;52(3):15-18. [doi: 10.3928/02793695-20140128-05] [Medline: 24496059]
64. Pew Research Center. Internet, Science & Tech. Social Networking Fact Sheet. 2013. URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [accessed 2019-12-12]
65. Naslund JA, Aschbrenner KA, McHugo GJ, Unützer J, Marsch LA, Bartels SJ. Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. Early Interv Psychiatry 2019 Jun;13(3):405-413 [FREE Full text] [doi: 10.1111/eip.12496] [Medline: 29052947]
66. Borzekowski DLG, Leith J, Medoff DR, Potts W, Dixon LB, Balis T, et al. Use of the internet and other media for health information among clinic outpatients with serious mental illness. Psychiatr Serv 2009 Sep;60(9):1265-1268. [doi: 10.1176/ps.2009.60.9.1265] [Medline: 19723745]
67. Torous J, Keshavan M. The role of social media in schizophrenia: evaluating risks, benefits, and potential. Curr Opin Psychiatry 2016 May;29(3):190-195. [doi: 10.1097/YCO.0000000000000246] [Medline: 26967314]
68. Yu Y, Li T, Xi S, Li Y, Xiao X, Yang M, et al. Assessing a WeChat-Based Integrative Family Intervention (WIFI) for Schizophrenia: Protocol for a Stepped-Wedge Cluster Randomized Trial. JMIR Res Protoc 2020 Aug 25;9(8):e18538 [FREE Full text] [doi: 10.2196/18538] [Medline: 32687478]

Abbreviations

BPRS-18: the 18-item Brief Psychiatric Rating Scale
CCMD-3: Chinese Classification of Mental Disorders-3
GAD-7: Generalized Anxiety Disorder Scale-7
GAF: Global Assessment of Functioning
ICD-10: International Classification of Diseases-10
PHQ-9: Patient Health Questionnaire-9
PLS: people living with schizophrenia
RAS: Recovery Assessment Scale
WHODAS 2.0: the 12-item World Health Organization Disability Assessment Schedule 2.0
WHOQOL-BREF: World Health Organization Quality of Life Brief Scale
WUIQ: WeChat Use Intensity Questionnaire

Edited by G Eysenbach; submitted 11.03.20; peer-reviewed by G Grispos, J Coquet, F Dong; comments to author 12.06.20; revised version received 02.07.20; accepted 03.08.20; published 18.09.20.

Please cite as:
Yu Y, Li Y, Li T, Xi S, Xiao X, Xiao S, Tebes JK
New Path to Recovery and Well-Being: Cross-Sectional Study on WeChat Use and Endorsement of WeChat-Based mHealth Among People Living With Schizophrenia in China
J Med Internet Res 2020;22(9):e18663
URL: http://www.jmir.org/2020/9/e18663/
doi:10.2196/18663
PMID:32945774

Original Paper

Determinants of Scale-up From a Small Pilot to a National Electronic Immunization Registry in Vietnam: Qualitative Evaluation

Huyen Dang[1], MD, PhD; Sang Dao[2], MPH; Emily Carnahan[3], MPH; Nami Kawakyu[3,4], RN, MPH; Hong Duong[1], MD, PhD; Trung Nguyen[1], MPH; Doan Nguyen[1], MD; Linh Nguyen[2], MPH; Maya Rivera[2], MPH; Tuan Ngo[2], MScIT; Laurie Werner[3], MPA; Nga Nguyen[2], MD, PhD

[1]National Expanded Program on Immunization, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

[2]PATH, Hanoi, Vietnam

[3]PATH, Seattle, WA, United States

[4]Department of Global Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Nga Nguyen, MD, PhD
PATH
#1101, 11th Floor, Hanoi Towers, 49 Hai Bai Trung Street
Hoan Kiem District
Hanoi,
Vietnam
Phone: 84 24 3936 2216 ext 105
Email: ntnguyen@path.org

Abstract

Background: Digital health innovations can improve health system performance, yet previous experience has shown that many innovations do not advance beyond the pilot stage to achieve scale. Vietnam's National Immunization Information System (NIIS) began as a series of digital health pilots, first initiated in 2010, and was officially launched nationwide in 2017. The NIIS is one of the few examples of an electronic immunization registry (EIR) at national scale in low- and middle-income countries.

Objective: The aim of this study was to understand the determinants of scale-up of the national EIR in Vietnam.

Methods: This qualitative study explored the facilitators and barriers to national scale-up of the EIR in Vietnam. Qualitative data were collected from October to December 2019 through in-depth key informant interviews and desk review. The mHealth Assessment and Planning for Scale (MAPS) Toolkit guided the development of the study design, interview guides, and analytic framework. MAPS defines the key determinants of success, or the "axes of scale," to be groundwork, partnerships, financial health, technology and architecture, operations, and monitoring and evaluation.

Results: The partnership and operations axes were critical to the successful scale-up of the EIR in Vietnam, while the groundwork and monitoring and the evaluation axes were considered to be strong contributors in the success of all the other axes. The partnership model leveraged complementary strengths of the technical working group partners: the Ministry of Health General Department of Preventive Medicine, the National Expanded Program on Immunization, Viettel (the mobile network operator), and PATH. The operational approach to introducing the NIIS with lean, iterative, and integrated training and supervision was also a key facilitator to successful scale-up. The financial health, technology and architecture, and operations axes were identified as barriers to successful deployment and scale-up. Key barriers to scale-up included insufficient estimates of operational costs, unanticipated volume of data storage and transmission, lack of a national ID to support interoperability, and operational challenges among end users. Overall, the multiple phases of EIR deployment and scale-up from 2010 to 2017 allowed for continuous learning and improvement that strengthened all the axes and contributed to successful scale-up.

Conclusions: The results highlight the importance of the measured, iterative approach that was taken to gradually expand a series of small pilots to nationwide scale. The findings from this study can be used to inform other countries considering, introducing, or in the process of scaling an EIR or other digital health innovations.

(J Med Internet Res 2020;22(9):e19923) doi:10.2196/19923

KEYWORDS

immunization; immunization information system; electronic immunization registry; scale-up; digital health intervention; mHealth; eHealth

Introduction

Digital health innovations are changing the way health is delivered worldwide. The World Health Organization (WHO) defines digital health as "the field of knowledge and practice associated with the development and use of digital technologies to improve health" [1]. Digital health innovations can play an important role in improving health system performance and can advance progress toward achieving universal health coverage and sustainable development goals [2,3]. However, previous experience has shown that many digital health innovations do not advance beyond the pilot stage to become institutionalized within health systems [4-6].

Vietnam's National Immunization Information System (NIIS) began as a series of digital health pilots, first initiated in 2010 [7], and was officially launched nationwide in 2017. By 2020, the NIIS included 20 million client records. The NIIS is an example of an electronic immunization registry (EIR), a confidential, computerized, population-based routine system to capture, store, access, and share individual-level, longitudinal health information on vaccine doses administered [8,9]. Immunization is among the most cost-effective child health interventions and saves 2 to 3 million lives per year; however, 1 in 5 infants do not receive all their required vaccine doses [10]. EIRs aim to improve the immunization delivery system to reach every child by supporting the delivery of more effective, efficient, data-driven care. EIRs can capture other individual-level demographic or health data and can link to other systems that manage vaccine stock and logistics, human resources, or other individual or population health data [11]. In Vietnam, the EIR, which includes SMS text message reminders, has been shown to improve immunization coverage and timeliness of vaccination [12] and is one of the few examples of an EIR at national scale in low- and middle-income countries.

Scale-up refers to "deliberate efforts to increase the impact of innovations successfully tested in pilot or experimental projects so as to benefit more people and to foster policy and program development on a lasting basis" [13]. Scaling goes beyond expanding an innovation to more users or geographies and in fact often results in new organizational or technological complexities [14]. Others have defined scale more expansively, including integration with the health system, sustainable funding and government support, and the ability to replicate, refine, and improve over time [4,15].

The digital health community has increasingly recognized the importance of scale and has initiated efforts to support scaling digital health interventions [15]. The Principles for Digital Development, collaboratively developed by 500+ implementers to capture best practices for integrating technology in development projects, highlight "design for scale" as one of 9 guiding principles [16]. Initiatives like the Health Data Collaborative [17], Digital Impact Alliance [18], and Digital Square [19] have been launched to support scale-up through alignment and coordination, strategic resourcing, operational guidance, and development of "global goods" that can be adapted and scaled in new contexts [15].

The evidence base to understand factors that influence scaling digital health interventions is limited [20,21]. The WHO and others have identified the need for implementation research to understand the complexities of implementing and scaling digital interventions [2,6]. National EIRs are far from being universal, even in high-income countries [22,23], and to our knowledge, there are limited studies of the facilitators and barriers influencing their scale-up, particularly in low-resource settings.

The aim of this study was to understand the determinants of scale-up of the national EIR in Vietnam, thereby contributing to the evidence on how and why digital interventions can successfully achieve scale.

Methods

Study Design

This study explored the facilitators and barriers to national scale-up of the EIR in Vietnam through qualitative methods, using the mHealth Assessment and Planning for Scale (MAPS) Toolkit as a conceptual framework [13]. The MAPS Toolkit, codeveloped in 2015 by the WHO, the United Nations Foundation, and Johns Hopkins University, outlines 6 axes to measure digital health project maturity. Qualitative data were collected through in-depth key informant interviews and document review and were analyzed according to the MAPS Toolkit axes.

Conceptual Framework

The MAPS Toolkit served as the conceptual framework for this study to ground Vietnam's experience in scaling up an EIR within known factors that influence successful scale-up and sustainability of digital health products. Although the toolkit was designed to prospectively guide iterative program implementation, it has also been used for retrospective program evaluation [24]. MAPS defines the key determinants of success, also known as the "axes of scale," to be groundwork, partnerships, financial health, technology and architecture, operations, and monitoring and evaluation (M&E) (Figure 1), each of which are divided into more specific drivers of success (Figure 2) [13].

Figure 1. The mHealth Assessment and Planning for Scale Toolkit axes of scale [13]. mHealth: mobile health.

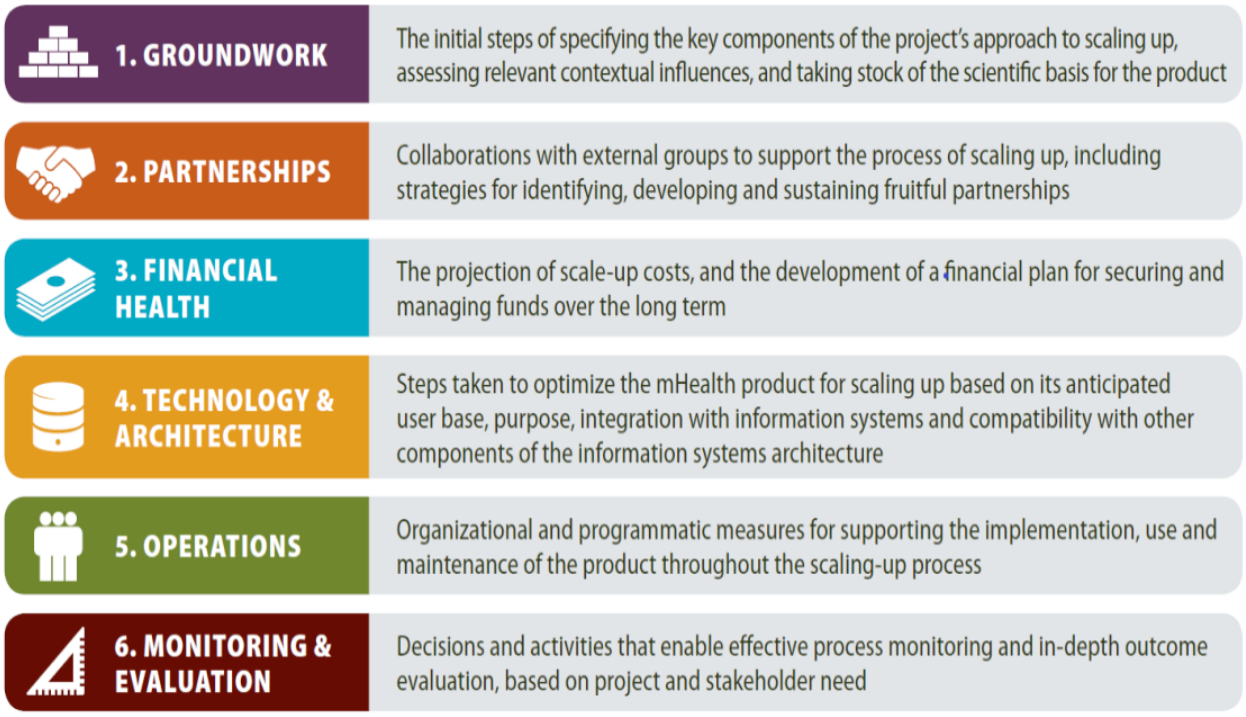
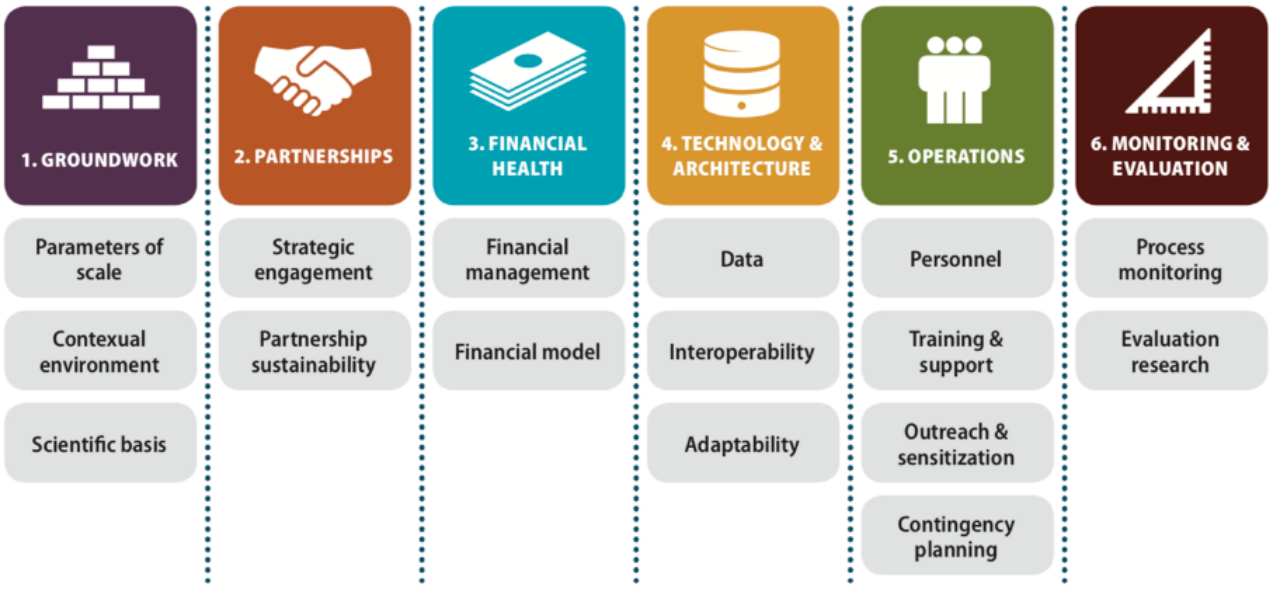


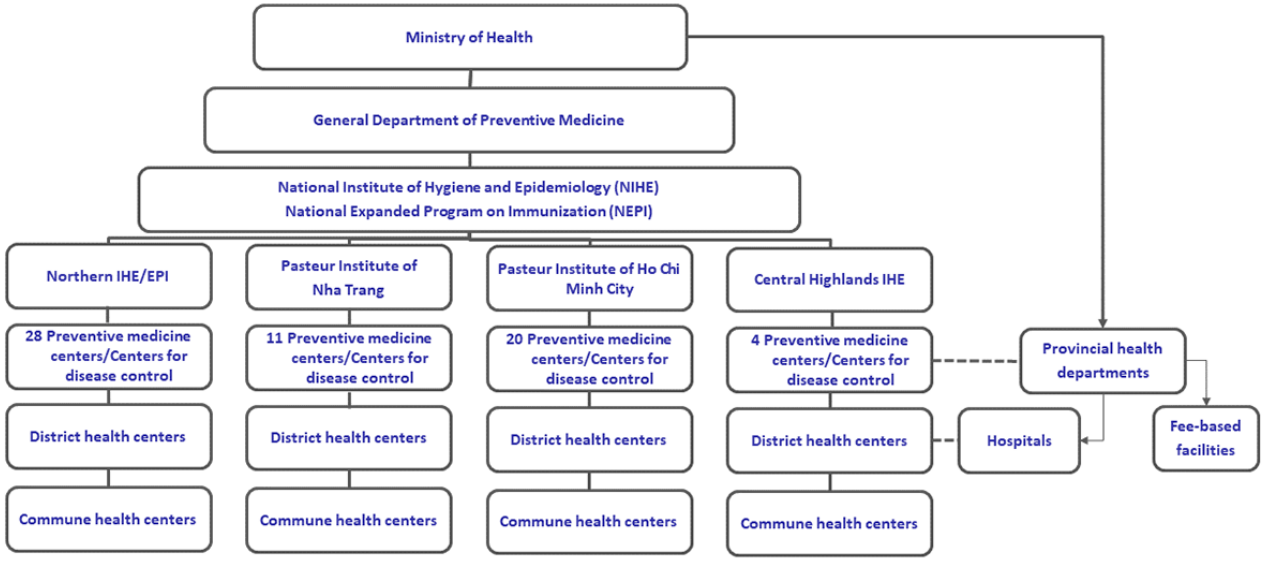
Figure 2. The mHealth Assessment and Planning for Scale Toolkit drivers of success within each axis of scale [13].



Setting

Vietnam has a 5-level health system, with administrative levels at the national, regional, provincial, district, and commune levels [25]. At the national level, the Ministry of Health (MOH) sets national policies and programs and manages the national and regional-level hospitals and institutes [26]. The provincial level oversees provincial health departments and provincial health centers and hospitals that follow national MOH policies [26]. The Vietnam National Expanded Program on Immunization (NEPI) was first introduced in Vietnam in 1981 with the primary goal of protecting children from the most common infectious diseases by providing free immunization services to children [27]. At the national level, the MOH's General Department of Preventive Medicine (GDPM) oversees NEPI activities in 4 regions, 63 provinces and cities, 696 districts, and 11,138 communes (Figure 3) [25]. Public immunization services are mainly delivered at the commune level, the hepatitis B vaccine is provided at hospitals, and vaccines that are not supported by NEPI are administered at fee-based immunization facilities [25]. Until 2009, Vietnam's immunization records and vaccine supply tracking were paper based. Health centers began to be equipped with computers in 2005; internet connection at the commune level became available in 2012.

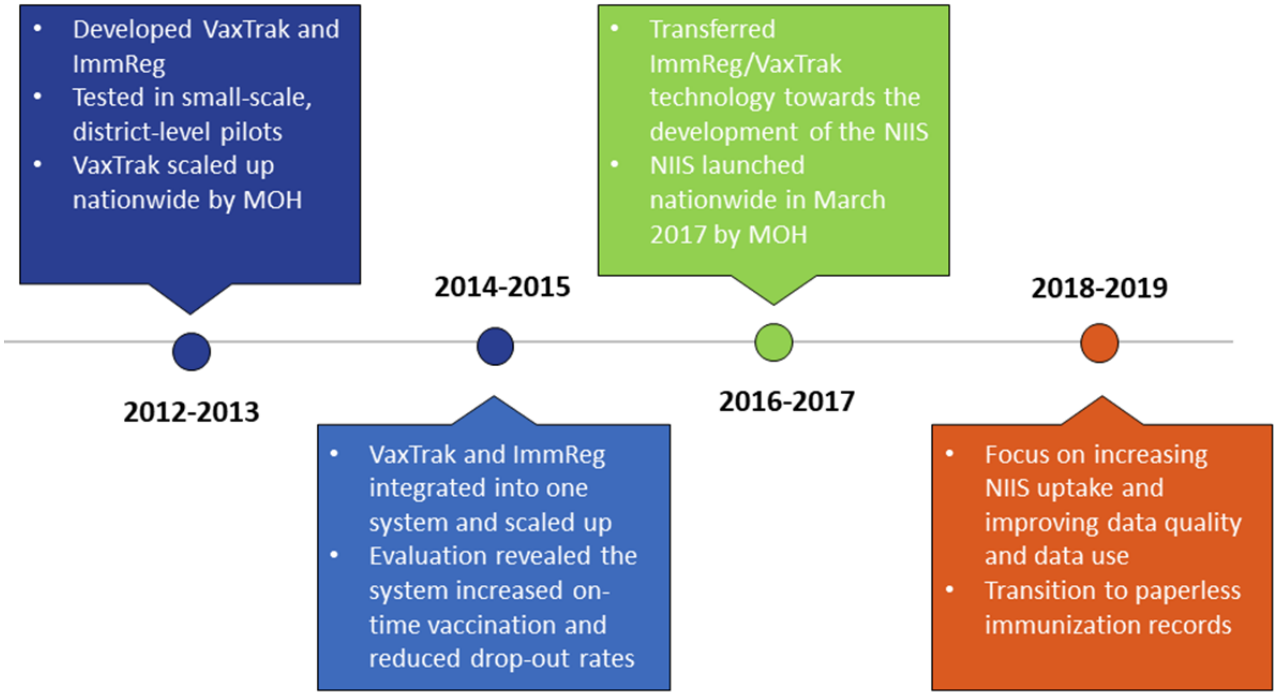
Figure 3. Structure of the health system in Vietnam. EPI: Expanded Program on Immunization; IHE: Institute of Hygiene and Epidemiology.



Intervention

The evolution of the current national EIR in Vietnam began with a small pilot (Figure 4). From 2010 to 2012, NEPI and PATH, an international nongovernmental organization serving as a technical partner, collaborated with the WHO to develop and pilot an electronic vaccine stock management system (VaxTrak) in 3 provinces and an immunization registry software (ImmReg) in one district of Ben Tre province. The goal of the software was to improve the ability to track babies who were due for vaccination and reduce the time for immunization recording and reporting compared with a paper-based system. Both systems were successful in reducing the time burden of reporting among health workers and most users found the systems to be acceptable and feasible for scale-up [7].

Figure 4. Timeline of the electronic immunization registry introduction and scale-up in Vietnam. MOH: Ministry of Health; NIIS: National Immunization Information System.



These two systems were combined, upgraded, and then deployed by NEPI with support from PATH in all districts of Ben Tre province from 2014 to 2015. This resulted in increased full immunization coverage and improved on-time vaccination rates [12].

In 2016, this combined system was integrated into the NIIS, which was being developed by the MOH in partnership with Viettel, the largest mobile network operator (MNO) in Vietnam. The NIIS was to be an EIR system in which health workers could register and track the immunization records of pregnant women and newborns, as well as inform vaccine stock

management. The NIIS technical working group includes GDPM, NEPI, Viettel, and PATH, who partnered to pilot the NIIS in 5 provinces. The NIIS was further upgraded based on learnings from this experience and in preparation for national scale-up. In June 2017, the NIIS was officially deployed nationally. Leading up to nationwide scale-up, the MOH requested all communes to back enter the full vaccination history for all children born from January 2015 through June 2017.

From 2018 to 2019, the NIIS technical working group focused on increasing NIIS uptake, improving data quality and data use, and transitioning completely to paperless immunization records. As of January 2020, over 20 million records have been registered into the system (including back-entered data, immunization data for pregnant women, and childhood immunizations delivered among the annual birth cohort of approximately 1.7 million [28]).

Participants

Project documents were selected for inclusion in the document review based on project and study relevance. A total of 1 project proposal and 8 project evaluation reports detailing the project objectives, processes, successes, and challenges were included in the document review. Documents included end users' perspectives on the acceptability and feasibility of the system through surveys of Expanded Program on Immunization (EPI) staff at provincial, district, and commune health centers conducted in 2012 and 2015.

Critical case purposive sampling [29] was used to select key informants to interview. A case was identified as critical based on their essential and extensive involvement with the project, with consideration for diversity of perspectives based on each participant's role on the project. A total of 6 key informants were selected: 2 Ministry of Health staff (1 from GDPM and 1 from NEPI), 2 representatives from Viettel, and 2 PATH staff members. These key informants represented each of the organizations in the national-level NIIS technical working group (TWG), had been involved in every phase of the NIIS implementation, and were selected for their ability to speak to all the axes of scale in the MAPS framework.

Data Collection

PATH staff conducted 6 in-depth semistructured interviews in person with key informants from October to December 2019. The semistructured interview guide was developed based on the evaluation questions and the MAPS framework. Interviews were audiorecorded with permission from the informant, and the interviewer documented summary notes during the interview. Audiorecordings were transcribed into Microsoft Word and translated into English.

Data Analysis

An a priori codebook was developed, guided by the MAPS axes. The codebook was then refined by coding data from project documents and in-depth interviews in Microsoft Excel and Atlas.ti (version 8; ATLAS.ti Scientific Software Development GmbH). Coding was completed independently and concurrently by 2 PATH staff and an external evaluator. Codes were reviewed by the evaluation team and any changes or disagreements were discussed until resolved. Once coding was finalized, codes were grouped into key themes guided by the evaluation questions and MAPS axes. Exemplary quotations were identified to present the themes and were validated with key stakeholders.

Ethics

The study procedures were reviewed and received nonresearch determination by PATH. Prior to the interviews, participants were informed of the study's objectives, advantages and disadvantages of participating, and rights of participants. Importance of maintaining confidentiality was emphasized during training of data collectors and the start of interviews. Written consent was obtained from all participants. Study data were stored in an access-restricted server, only available to study staff for the purpose of data analysis.

Results

Overview

Table 1 summarizes the key facilitators and barriers experienced during scale-up of the NIIS, organized by the MAPS axes of scale.

Table 1. Key facilitators and barriers to scale-up.

Axes of scale	Facilitators	Barriers
Groundwork and monitoring and evaluation	- Learnings from multiple pilots and phases contributed to optimization of all other axes - Learnings established scientific basis for scale-up	N/A[a]
Partnerships	- Each member of the partnership plays a critical role - MOH[b] is owner and leader - MOH contracts MNO[c] to provide a service - Expansive technological, network, human, and financial resource capacity of MNO - MNO well established and trusted by government and public	N/A
Financial health	- Estimated time and cost of system - MNO provides services free of charge - Government commitment to allocate budget	- Unanticipated operational costs, particularly for refresher trainings, supportive supervision, and regular system upgrades
Technology and architecture	- User-friendly system - System adaptability - Strong data security and privacy - Existing infrastructure	- Enormity of data affects system capacity and network connectivity - Lack of national ID
Operations	- Availability of pretested and enhanced training materials and standard operating procedures - Cascading training-of-trainer training for nationwide training and technical support network - Integration of supervision within existing structures	- Dual reporting systems (paper and digital) - Low computer literacy of end users - High turnover of health care workers

[a]N/A: not applicable.

[b]MOH: Ministry of Health.

[c]MNO: mobile network operator.

Key Facilitators to Scale-up

The partnership and operations axes were most commonly perceived to be critical to the successful deployment and scale-up of the EIR in Vietnam. Though key informants identified elements in the financial health axis and technology and architecture axis that were facilitators to successful scale-up, these axes were not as commonly or strongly mentioned in comparison to the partnership and operations axes. The groundwork and monitoring and evaluation axes were considered to be strong contributors in the success of all the other axes.

Partnership

The NIIS is a product of a hybrid model, with the public, private, and civil society sectors contributing to the successful scale-up and sustainability of the NIIS. The NIIS partnership has enabled the government to own and run the system alongside supportive partners working together as a team.

The NIIS TWG was formed in 2016 with the aim to contribute to the technical implementation and sustainability of the NIIS. Membership comprises MOH's GDPM as the management authority, NEPI as the immunization expert, Viettel as the technology expert, and PATH as the liaison, serving as the connection across stakeholders. Initially the TWG met weekly, then over time it shifted to once or twice per month, depending on whether NIIS operationalization challenges emerged that required a timely response.

The most commonly expressed theme among key informants was the importance of the partnership in the success of the EIR implementation and scale-up. Key informants repeatedly noted the importance of the role that each member held in the TWG. One informant commented:

> *Each member of the TWG plays an important role. It could not be as it is without each member.* [Key Informant A]

Key informants described the importance of the MOH's role, both to sustainably lead and to oversee the national system. As key informants observed:

> *Political commitment is key. First you need to have political commitment, to invest in resources and provide guidance from the government for implementation.* [Key Informant B]

> *Only the MOH can manage such a large system that includes so much of the population.* [Key Informant C]

NEPI provided immunization expertise and acted as the technical lead to define and develop immunization workflows,

user requirements, and reporting systems and mechanisms. NEPI also oversaw and strengthened the implementation of the NIIS.

Given the expansive technological requirements of a national EIR, the MOH sought to partner with an MNO that already had an extensive network and was capable of high-capacity data storage. In addition to these technical specifications, key informants noted the importance of selecting a well-established MNO with a history of success:

> *If you want to implement a system at this scale, it is imperative to work with a big company like Viettel. The MNO of choice must have large capacity and a large presence in the country, be financially secure, and be able to provide necessary human resources in both quality and quantity. Choosing an MNO, it would be best to select the largest MNO with the longest history of success in the country of implementation.* [Key Informant D]

Additionally, some informants observed that the nature of the partnership between MOH and Viettel was a driver for successful scale-up and sustainability. That is, Viettel provided key development, storage, and maintenance services rather than only providing software developer services that would eventually be transitioned to the government to maintain and upgrade. As one informant explained:

> *Viettel provides a service, not just the software, so there is always a partnership and not just a handover. So, whenever the system needs to be updated, the MOH and Viettel are working together. NIIS belongs to the MOH and Viettel provides a service. This is a good model for other countries: outsource the service to a mobile network operator.* [Key Informant E]

PATH has played a critical role throughout the pilot–to–national scale-up process. Having both local and global experience in piloting, scaling up, and evaluating digital health products, PATH was able to transfer all technology, such as data flow and database structure design, to Viettel and share lessons learned with the TWG. PATH could also act as a liaison between the MOH, a health system–focused entity, and Viettel, an information technology entity, given its experience in global health and digital technologies. One informant explained:

> *PATH complements and supplements the gaps in MOH and Viettel. PATH's experience in the field, piloting and demonstrating and disseminating that information globally, has been critical.* [Key Informant B]

This partnership, in which each partner contributed a critical role, was built on a foundation of trust, which key informants identified as an important driver of successful scale-up and sustainability:

> *Viettel is government owned, so they trust the government, and the government trusts them. So when MOH trusts PATH, then there is a big circle of trust… An MNO that is trusted by the government is very important. Because the health information is very sensitive. So the partnership can only be established with an MNO that is an old and trusted one, not a new or just established MNO. This helps with scaling-up.* [Key Informant B]

Operations

Another recurrent theme across the interviews was the strength of the end user training and support approach that was developed for successful scalability and sustainability.

For NIIS national deployment, training materials and standard operating procedures, which had been developed, tested, and enhanced based on learnings from a landscape assessment and monitoring and evaluation of multiple phases of the EIR, were readily available for use. Additionally, the cascade approach to conduct training-of-trainer trainings had been developed to quickly establish and expand a network to train end users across the country. As one informant explained:

> *You want to conduct a cascade training because you don't have the resources to go everywhere. Instead, conduct the training-of-trainer training to provincial and district health workers, then they can provide training and technical support to end-users in hospitals, fee-based facilities, and commune health centers. This is the best method due to limited resources and the trainers will be the supervisors, so this informs them about who is the strongest and who is weak so they can provide feedback to them to improve.* [Key Informant E]

Through these trainings, a multilevel technical support network was established across the country. The trained provincial and district health staff acted as mentors who served as focal points for technical support at each health level and provided direct support to end users. Supportive supervision was integrated into existing structures of immunization supervision. In addition, Viettel trained district-level staff on the NIIS, who are now able to provide technical support to end users. This approach is working to sustain country ownership and sustainability of the NIIS in Vietnam.

Financial Health

Estimating the cost of the system and its implementation to appropriately allocate resources to scale up was critical to the success of the NIIS. After the first pilot from 2010 to 2012, a business model framework was developed to provide an overview of the proposed key partners, resources, and activities for sustaining and scaling an EIR in Vietnam. A costing model was also created to determine the financial resources needed to implement the system across multiple provinces in Vietnam.

Identified costs included software development and maintenance, training, supportive supervision and monitoring, and the cost for technical meetings after the first months of implementation. In the case of the NIIS, the MOH has not incurred the cost of software development and maintenance, as the MNO has been providing these services in kind, currently with the expectation that the pricing model may shift in the long-term.

In addition to estimating costs of the system, as a necessary component to successful EIR scale-up, it was essential to have

the government commitment to allocate a budget to this work. As one informant explained:

> *The government needs to commit and support the implementation and [provide] direction and leadership to the local government to allocate the budget for implementation and maintenance.* [Key Informant E]

Technology and Architecture

Key informants noted 3 main technology and architecture facilitators to successful scale-up of the EIR related to the technology itself: its user-centered design, adaptability, and strong data security and privacy. Initial development of the EIR and subsequent upgrades prioritized end user needs. As such,

> *The system is designed with high-rate of acceptability and is user-friendly. It meets the basic requirements of end-users and it operates smoothly. These are factors to excite provinces to implement and use the system.* [Key Informant F]

The NIIS supports facility health workers to document and track immunization records and can easily print electronic records for storage in compliance with current regulations. Individual immunization records can be updated with vaccine doses delivered in different facilities, which is particularly important in large cities where many children receive vaccines from public and private facilities and there is a large migrant population moving between facilities.

To maintain system acceptability by ensuring smooth operation of the system, the NIIS has a mechanism to monitor system operations so that issues that appear can quickly be resolved. The system is updated regularly, and adaptations can be made to fit evolving needs. Relatedly, the system has strong data security and privacy measures to maintain public trust of the NIIS.

In addition to the NIIS design, the importance of an already existing technology infrastructure at health facilities cannot be overstated as a facilitator to the successful deployment and scale of the EIR. That is, because almost all health facilities in Vietnam already had computers and internet, and health care workers had access to computer systems, the NIIS could be developed with this starting point in mind. Had computers and internet not already been available at health facilities, the NIIS would likely have been designed differently or would have required additional resources at the outset for equipment and connectivity.

Groundwork and Monitoring and Evaluation

The steps taken in the early phases of the EIR and the M&E learnings from multiple EIR implementation phases set the groundwork for successful national scale-up, especially informing the planning, implementation, and optimization of the other axes of scale. Given this interconnected relationship between groundwork and M&E throughout the many implementation phases, we merged these two axes for our analysis.

Key groundwork activities included a landscape analysis and the development of a business model framework. The main aim of the landscape analysis was to evaluate how a digital registry might improve the ability to track children due for vaccinations and how it might shorten the time required for recording and reporting immunizations compared to a paper-based registry. This analysis provided a snapshot of the current policies, technical capacities, and health information systems already in place or under development in Vietnam. On the importance of conducting a landscape analysis, one informant emphasized:

> *It is a necessary step to look at two sides: the policy and technical sides. Before you scale-up, you must look at the policy environment and see what are the advantages, what is available, and what are the gaps. To understand what is financially necessary and what are some foreseen challenges.* [Key Informant B]

The analysis was particularly critical in informing the development of the pilot EIR system in 2010 but also continued to guide decisions about the operations and the technology and architecture of the EIR phases. One major takeaway from the analysis was the importance of understanding user needs to build a system that supports end users at all levels. The analysis also informed the development of the end user training plan, as it identified potential issues such as high turnover and low computer literacy of health care workers that could affect successful implementation and scale-up of the EIR.

Learnings gathered from both groundwork activities and M&E facilitated effective and efficient national scale-up:

> *We only had 8 months to work on the NIIS together [in 2016], but at that time we already had the business model and understood well about the business and the technical, so we could just transfer the knowledge and the technology to develop the NIIS. This shortened the time for NIIS development.* [Key Informant E]

The multiple phases of the EIR that preceded NIIS national scale-up created many opportunities for learning and optimization. In total, the Vietnam EIR system phased through 4 pilots and deployments from 2010 to 2017, starting with a pilot at the district level, then scaled and piloted at the provincial level, scaled again and piloted in multiple provinces, and then finally, scaled nationally. As one informant noted:

> *The piloting of the system is so important to get feedback from the end-users, before scale-up. And it is important to collect lessons learned as you go, to determine when you can and how you can scale-up.* [Key Informant E]

This continuous test-learn-optimize and scale-learn-optimize improvement cycle allowed for improvements in the operations and technology and architecture of the EIR, such that by the time it was scaled nationally, many things had been pretested and improved.

Furthermore, evaluation activities from these phases contributed to the growing evidence of the benefits of an EIR, as well as local validation that such a system was both feasible and effective in the Vietnam context. It was these quantifiable successes that catalyzed the support for nationwide scale-up. As one key informant recommended:

> *The pilot should be conducted in two years, to get the feedback and upgrade the system, and have enough time to evaluate the system. Two years is enough to evaluate the impact of the system rather than just evaluate the acceptability and feasibility of the system. In the short-term, end-users just want to say the good things.* [Key Informant E]

Key Barriers to Scale-up

The financial health, technology and architecture, and operations axes were most commonly perceived to be barriers to the successful deployment and scale-up of the EIR in Vietnam.

Financial Health

Although the cost of EIR deployment was estimated during various EIR phases, there were some costs to national scale-up that were underestimated or not estimated. The majority of the initial financial model was based on costs associated with software development and maintenance. Budgets for operational costs of training, supportive supervision, and monitoring and evaluation were sometimes insufficient. For example, there was insufficient funding and human resources for EIR's supportive supervision visits:

> *A low budget for supportive supervision visits made it difficult for higher level supervisors to conduct timely supervision of end-users.* [Key Informant F]

> *Local governments do not allocate budget for research and some don't have supervision budgets, so they had to integrate [NIIS supervision] with supervision of other health areas.* [Key Informant E]

Another unanticipated cost to scale-up was that, because of the enormity of the data, the server has needed to be frequently updated and expanded. Through this experience, the TWG has identified the importance of accurately estimating the operational aspects of scaling up an EIR.

Technology and Architecture

Two key barriers to successful scale-up and sustainability were identified within the technology and architecture axis, namely issues with data transmission and storage and interoperability.

Although a sizing and infrastructure assessment was conducted to inform the server capacity and connectivity bandwidth, the substantial increase in the demand for data storage and transmission for national scale-up were not appropriately anticipated. There are nearly 20 million clients in the NIIS, and this number will continue to increase as the population grows. Informants explained:

> *We did not foresee just how large the data would be, how large of a number of people would be registered into the system, and how to store this data without slowing down the system. The number of people registered is growing rapidly… During times of high influx, the NIIS system is overloaded. The server needs to be updated and expanded frequently to accommodate the enormity of the data.* [Key Informant F]

> *We did not estimate the increase of data, therefore could not optimize the algorithm as well as the database infrastructure from the beginning.* [Key Informant A]

This, along with a large number of people using the system, has meant that the system can sometimes run slowly. The TWG is working to find a solution, which includes upgrading the server, separating the data storage and data query servers, and developing standard operating procedures in cases where the system is down. The MNO has also been working to address this issue by identifying and indexing frequently queried data files to optimize the speed of queries.

Another barrier to successful scale-up has been the lack of a national ID to develop an integrated, interoperable system. Viettel is currently working to develop a health information system that integrates all data from existing systems, including immunization, infectious diseases, and noncommunicable diseases, all of which have their own software without a common unique identifier for each client. This has been an issue even within the NIIS, with duplication of individual records. (Based on a January 2019 GDPM report, 1.2% of NIIS records were duplicates [30].) Similarly, facilities that use their own systems, such as hospitals and private facilities, are not interoperable with the NIIS. The NIIS includes barcode technology, and it has been suggested that each child's unique barcode could serve as a national ID; however, with limited resources, there are few public facilities equipped with barcode readers and printers. Through this experience, the TWG has identified that having a national unique ID would be a facilitator to building an interoperable system.

Operations

There were 3 key barriers to successful scale-up and sustainability identified within the operations axis: dual reporting, low computer literacy of end users, and high turnover of health care workers.

From the beginning, there have been persistent concerns among health sector staff about whether an EIR would increase the workload of health care workers. Evaluations of the early phases of the system found that it successfully reduced the time burden of reporting among health workers; however, the current dual-reporting scenario is burdening the health workforce workload. Because a completely digital system is not mandated by the government, end users are tasked with using both the paper-based system and the NIIS to ensure data accuracy in both systems. During planning for national scale-up, a clear transition plan from paper to digital would reduce the workload for end users.

During the landscape analysis conducted in 2010, low computer literacy of end users was identified as a possible barrier to effective implementation and scale-up of an EIR. Some health care workers were not well versed in computer usage, including basic computer functions, such as typing and Microsoft Excel. Low computer literacy also meant that staff would not know how to use the data to calculate indicators and interpret results for decision making. As one key informant described:

> *One of the difficulties in the pilot phase was the computer skills of the health care workers. Some needed a lot of support, even in typing. We trained on everything. This was a big challenge for implementing the system.* [Key Informant E]

Given that the issue was identified during the landscape analysis, the project team was able to use various strategies in the development of the EIR and training approach to increase the usability of the EIR for end users:

> *Due to different levels of digital literacy of commune health workers, each commune must assign two people to take part in the training sessions. After the training, this peer support is very important to ensure success in implementation.* [Key Informant E]

A lot has changed since these early years of EIR implementation, with an increasingly computer-literate health workforce and increased computer and internet infrastructure at the commune level. This change has helped increase the accessibility and usability of the NIIS.

Another barrier to scale-up and sustainability identified during the landscape analysis was the frequent staff rotation and high staff turnover rate. Again, because this was identified during the landscape analysis, it was possible to address this during the development of the training approach. Multiple staff members at every level of the health system, including the health facility, district, provincial, and national levels, are trained. The cascading training-of-trainers approach (described in "Operations" under the "Key Facilitators to Scale-up" section) also ensures a sustainable peer-training system is in place to train new staff. Vietnam is also exploring e-learning options to train new staff.

Discussion

Principal Findings

While there were facilitators across all axes within the MAPS framework, the partnership and operations axes were most commonly perceived to be critical to the successful scale-up of the EIR in Vietnam. The EIR scale-up has been facilitated by a partnership comprised of public, private, and civil society actors working together toward shared goals and leveraging each partner's expertise. It was operationalized through a cascading training-of-trainer approach, which was integrated within existing supervision structures and used training materials that were iteratively refined during each phase of implementation.

The most commonly perceived barriers to the scale-up of the EIR in Vietnam were related to the financial health, technology and architecture, and operations axes. Related to financial health, there were challenges in sufficiently estimating operational costs associated with training, supportive supervision, and monitoring and evaluation. Technology and architecture barriers included issues with the unanticipated volume of data storage and transmission demands and the lack of a national ID to support interoperability with other health information systems. Key operational barriers were mainly among end users and included challenges of dual-reporting systems, low computer literacy, and high staff turnover.

Overall, the phased approach of multiple small pilots allowed for iterative assessment and planning that strengthened all the axes and contributed to the successful scale-up.

Implications for Scale-up and Future Research

Though the MAPS Toolkit was not designed as an evaluation framework, overall, the axes of scale and their specified definitions helped ground the analysis for this evaluation. The MAPS conceptual model acknowledges that scaling up is an "iterative cyclical process of thorough assessment, careful planning and targeted improvements" [13]. However, there is not an axis or a cross-cutting axis to represent the dynamic, continuous learning-and-improvement process that was so critical to the success of the full national-scale deployment of the NIIS. In Vietnam, each of the 4 phases of the EIR deployment from 2010 to 2017 allowed for opportunities to learn and make improvements that strengthened multiple axes. The technology itself improved with feedback from end users to improve its acceptability and functionality, operations benefitted by fine-tuning training materials and standard operating procedures, costing models were refined at various points, and with time and experience, trust between the partners continued to grow. Furthermore, the evaluation activities from each phase built the evidence base, which strengthened partners' support to continue to scale the system. Thus, the MAPS framework was adapted during analysis by merging the M&E and groundwork axes as a way to acknowledge this interconnectedness of the two axes. When applying the MAPS framework, implementors and evaluators should strive to understand the interconnectness between and dynamic influences across axes.

Using the MAPS Toolkit allowed for comparison between the experience in Vietnam with that of Tanzania and Zambia, where a recent study explored the factors influencing the introduction and adoption of EIRs in low-resource settings using the MAPS Toolkit [24]. Tanzania and Zambia took a similar approach by starting implementation in a single pilot district or province before expanding the EIR further, which was also perceived as a facilitator of scale-up, as it allowed for opportunities to learn and iterate [24]. The value of pilot testing an EIR has been documented in the United States as well [9] and is consistent with the scale-up literature for other nondigital interventions [31].

The central role of the partnership model in Vietnam cannot be overstated. There were initial challenges as partners developed shared language and understanding across their areas of expertise, which included public health and immunization, financial health, and information technology. It was important to allow time for this learning curve and to formalize roles as part of the NIIS TWG. Lessons learned from implementing subnational EIRs in the United States also highlighted the importance of a steering committee or coalition to guide EIR development and garner support [9]. Similar to Vietnam, in both Tanzania and Zambia, the MOH leadership's support of the EIR was an important facilitator, and the MOH relied on PATH as a technical partner. However, the partnership model

in these African countries differed from Vietnam's in that nongovernmental partners, including PATH, John Snow Inc, and the Catholic Medical Mission Board were more directly involved with implementing the EIR at subnational levels in Tanzania and Zambia [24]. Other countries may also not be able to replicate the MNO partnership that was a key facilitator to scale-up in Vietnam, as Viettel is uniquely positioned as a state-owned MNO with large market share across the country.

In terms of operations, Tanzania initially used an on-the-job training approach that relied on PATH staff to visit each health facility, but later Tanzania and Zambia implemented a training-of-trainers approach similar to Vietnam's model. As mentioned, nongovernmental partners were more heavily involved in the initial subnational deployment and trainings in Tanzania and Zambia, which was also a large cost driver in these countries [32]. In all three countries, EIR supervision was integrated within existing structures, which was a facilitator for scale and sustainability, but also meant there was limited time spent specifically on EIR issues during supervision visits. Similar to Vietnam, Tanzania and Zambia also faced barriers due to dual-reporting systems, low computer literacy of end users, and high health care worker turnover. Other studies echo that human resources are a challenge to the Vietnam EPI as a result of limited resources for supervision in hard-to-reach areas, underpaid and unmotivated frontline workers, and low levels of knowledge and computer literacy [33].

Whereas in Vietnam, the existing technology infrastructure was in place at health facilities, this was not the case in Tanzania and Zambia, where lack of hardware and electricity was a barrier to initial introduction of the EIR [24]. Other technology and architecture barriers to scale of the EIRs in Tanzania and Zambia included synchronization delays, discrepancies in data across systems, and challenges due to separate immunization service delivery and stock management systems [24]. In Vietnam, this last challenge was avoided by integrating ImmReg and VaxTrak into a single system (NIIS) before scaling nationwide. The iterative piloting of the technology in Vietnam may have avoided the heavy costs that can be associated with EIR redesign or reconfiguration [9,32]. Both Tanzania and Zambia had to modify their EIR requirements and design a second system for scale [34].

As more countries aim to introduce EIRs, these findings can inform efforts to plan for scale-up. Seth Berkley, CEO of Gavi, the Vaccine Alliance, wrote about the importance of harnessing digital innovations to support immunization, noting that "one of the biggest needs is for affordable, secure digital identification systems that can store a child's medical history, and that can be accessed even in places without reliable electricity" [35]. Previous studies have recommended using digital health innovations for immunization, based on systematic reviews [36], but there is recognition that current research on the role of digital health in immunization initiatives is limited [22,36]. This is true for the field of digital health overall, with repeated calls to continue to strengthen the evidence base [6]. The findings presented here touch on priority evidence gaps related to evaluating economics, enabling ecosystems, financial and programmatic sustainability, and data use pathways related to digital health interventions [6].

Limitations

This study was based on the experience scaling the EIR in Vietnam, and care should be taken when generalizing these findings beyond Vietnam. This limitation was in part addressed by identifying determinants guided by the MAPS Toolkit, a conceptual framework that can be applied to digital health implementations in varied contexts.

This study purposively selected key informants to represent the range of partners on the NIIS TWG; each was very involved in the EIR scale-up and could speak to the details of the implementation. The study focused on national-level stakeholders and did not include key informants representing subnational or community perspectives. Although end user perspectives were represented in the documents reviewed, for a more comprehensive understanding of the factors influencing the EIR scale-up in Vietnam, future studies could also include perceptions of the barriers and facilitators from EIR end users at the district and health facility levels, as well as caregiver or community perceptions of immunization service delivery. Different determinants may emerge at each level. For example, a systemic literature review (with the majority of studies from the United States) identified health care providers' perceptions of usefulness and ease of use of digital health innovations as influential factors to their acceptance of the innovation [37]. Dolan et al [24] highlighted some of the barriers influencing EIR scale-up at subnational levels in Tanzania and Zambia, including inadequate data bundles, increased workload due to dual systems, and lack of EIR integration with the health management information system (Zambia only). In Vietnam, subnational differences in urbanization, socioeconomic status, behaviors, and other factors that may have affected EIR scale-up were not the focus of this study.

Conclusion

This study described the key facilitators and barriers that influenced the scale-up of the NIIS in Vietnam using a comprehensive digital health framework. The results highlight the importance of the measured, iterative approach that was taken to gradually expand a series of small pilots to nationwide scale. Key facilitators included the partnership model, which leveraged complementary strengths of the MOH and GDPM, NEPI, Viettel, and PATH, and the operational approach to introducing the NIIS with lean, iterative, and integrated training and supervision. Key barriers to scale-up included insufficient estimates of operational costs, unanticipated volume of data storage and transmission, lack of a national ID to support interoperability, and operational challenges among end users. The findings from this study can be used to inform other countries considering, introducing, or in the process of scaling an EIR or other digital health innovations.

Acknowledgments

The authors thank the Bill and Melinda Gates Foundation (BMGF) for providing support for this study. The authors also thank the technical working group team members involved in implementing the EIR scale-up in Vietnam, including representatives from GDPM, NEPI, Northern EPI, Hanoi Center for Disease Control, and PATH. Thank you to the PATH Vietnam team: Hung Pham, Thi Hoang, and Hai Le. The authors thank Tove Ryman (BMGF) for her support and guidance. This study was funded by BMGF (Investment ID OPP1198325).

Conflicts of Interest

None declared.

References

1. Draft global strategy on digital health 2020-2024. Geneva, Switzerland: World Health Organization URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2020-09-02]
2. WHO Guideline: recommendations on digital interventions for health system strengthening. Geneva, Switzerland: World Health Organization; 2019. URL: https://www.who.int/publications-detail/who-guideline-recommendations-on-digital-interventions-for-health-system-strengthening [accessed 2020-09-02]
3. Mehl G, Labrique A. Prioritizing integrated mHealth strategies for universal health coverage. Science 2014 Sep 12;345(6202):1284-1287. [doi: 10.1126/science.1258926] [Medline: 25214614]
4. Wilson K, Gertz B, Arenth B, Salisbury N. The Journey to Scale: Moving Together Past Digital Health Pilots. Seattle, WA: PATH; 2014. URL: https://www.path.org/resources/the-journey-to-scale-moving-together-past-digital-health-pilots/ [accessed 2020-09-04]
5. Lemaire J. Scaling up mobile health: Elements necessary for the successful scale up of mHealth in developing countries. Geneva, Switzerland: Advanced Development for Africa; 2011 Dec. URL: http://www.adaorganization.net/uploads/2/3/7/1/23713723/scaling_up_mobile_health_elements_necessary_for_the_successful_scale_up_of_mhealth_in_developing_countries.pdf [accessed 2020-09-04]
6. Labrique A, Vasudevan L, Weiss W, Wilson K. Establishing Standards to Evaluate the Impact of Integrating Digital Health into Health Systems. Glob Health Sci Pract 2018 Dec 10;6(Suppl 1):S5-S17 [FREE Full text] [doi: 10.9745/GHSP-D-18-00230] [Medline: 30305335]
7. PATH, National Expanded Programme on Immunization. Summary of Optimize Activities Conducted With Vietnam's National Expanded Programme on Immunization. Hanoi, Vietnam: PATH; 2012. URL: https://www.path.org/resources/summary-of-optimize-activities-conducted-with-vietnams-national-expanded-programme-on-immunization/ [accessed 2020-09-04]
8. Classification of Digital Health Interventions v1.0. Geneva, Switzerland: World Health Organization; 2018. URL: http://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf?sequence=1 [accessed 2020-09-04]
9. Freeman VA, DeFriese GH. The challenge and potential of childhood immunization registries. Annu Rev Public Health 2003;24:227-246. [doi: 10.1146/annurev.publhealth.24.100901.140831] [Medline: 12668757]
10. UNICEF. Vaccination and Immunization Statistics. UNICEF Data: Monitoring the situation of children and women. 2019. URL: https://data.unicef.org/topic/child-health/immunization/ [accessed 2020-03-20]
11. Hinman AR, Ross DA. Immunization registries can be building blocks for national health information systems. Health Aff (Millwood) 2010 Apr;29(4):676-682. [doi: 10.1377/hlthaff.2007.0594] [Medline: 20368598]
12. Nguyen NT, Vu HM, Dao SD, Tran HT, Nguyen TXC. Digital immunization registry: evidence for the impact of mHealth on enhancing the immunization system and improving immunization coverage for children under one year old in Vietnam. Mhealth 2017;3:26 [FREE Full text] [doi: 10.21037/mhealth.2017.06.03] [Medline: 28828373]
13. World Health Organization. The MAPS toolkit: mHealth assessment and planning for scale. Geneva, Switzerland: World Health Organization; 2015. URL: http://www.who.int/entity/reproductivehealth/publications/mhealth/maps/en/index.html [accessed 2020-09-04]
14. Sanner TA, Roland LK, Braa K. From pilot to scale: Towards an mHealth typology for low-resource contexts. 2012 Sep Presented at: ECIS 2012 Proceedings; May 15, 2012; Barcelona, Spain URL: https://aisel.aisnet.org/ecis2012/112 [doi: 10.1016/j.hlpt.2012.07.009]
15. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
16. Principles for Digital Development. URL: https://digitalprinciples.org/principles/ [accessed 2019-01-30]
17. Health Data Collaborative. URL: https://www.healthdatacollaborative.org/ [accessed 2020-04-24]
18. Digital Impact Alliance. URL: https://digitalimpactalliance.org/ [accessed 2020-04-24]
19. Digital Square. URL: https://digitalsquare.org [accessed 2020-04-24]

20. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
21. Agarwal S, Lefevre AE, Labrique AB. A Call to Digital Health Practitioners: New Guidelines Can Help Improve the Quality of Digital Health Evidence. JMIR Mhealth Uhealth 2017 Oct 06;5(10):e136 [FREE Full text] [doi: 10.2196/mhealth.6640] [Medline: 28986340]
22. Tozzi AE, Gesualdo F, D'Ambrosio A, Pandolfi E, Agricola E, Lopalco P. Can Digital Tools Be Used for Improving Immunization Programs? Front Public Health 2016;4:36 [FREE Full text] [doi: 10.3389/fpubh.2016.00036] [Medline: 27014673]
23. Derrough T, Olsson K, Gianfredi V, Simondon F, Heijbel H, Danielsson N, et al. Immunisation Information Systems – useful tools for monitoring vaccination programmes in EU/EEA countries, 2016. Euro Surveill 2017 Apr 27;22(17):30519. [doi: 10.2807/1560-7917.es.2017.22.17.30519]
24. Dolan SB, Alao ME, Mwansa FD, Lymo DC, Bulula N, Carnahan E, et al. Perceptions of factors influencing the introduction and adoption of electronic immunization registries in Tanzania and Zambia: a mixed methods study. Implement Sci Commun 2020;1:38 [FREE Full text] [doi: 10.1186/s43058-020-00022-8] [Medline: 32885195]
25. Organizational system.: Vietnam Expanded Program on Immunization URL: http://www.tiemchungmorong.vn/vi/content/he-thong-chuc.html [accessed 2020-04-24]
26. Nguyen TD, Dang AD, Van Damme P, Nguyen CV, Duong HT, Goossens H, et al. Coverage of the expanded program on immunization in Vietnam: Results from 2 cluster surveys and routine reports. Hum Vaccin Immunother 2015;11(6):1526-1533 [FREE Full text] [doi: 10.1080/21645515.2015.1032487] [Medline: 25970593]
27. TCMR History.: Vietnam Expanded Program on Immunization URL: http://www.tiemchungmorong.vn/vi/content/lich-su-tcmr.html [accessed 2020-04-24]
28. National Expanded Program on Immunization. Vietnam Expanded Program on Immunization (EPI) Report 2018. Hanoi, Vietnam: Vietnam Ministry of Health; 2019.
29. Cohen D, Crabtree B. Qualitative Research Guidelines Project. 2006. URL: http://www.qualres.org/ [accessed 2020-04-24]
30. General Department of Preventive Medicine. Report on duplications in 9 provinces. Quang Ninh, Vietnam: Vietnam Ministry of Health; 2019.
31. Barker PM, Reid A, Schall MW. A framework for scaling up health interventions: lessons from large-scale improvement initiatives in Africa. Implement Sci 2016 Jan 29;11:12 [FREE Full text] [doi: 10.1186/s13012-016-0374-x] [Medline: 26821910]
32. Mvundura M, Di Giorgio L, Lymo D, Mwansa FD, Ngwegwe B, Werner L. The costs of developing, deploying and maintaining electronic immunisation registries in Tanzania and Zambia. BMJ Glob Health 2019;4(6):e001904 [FREE Full text] [doi: 10.1136/bmjgh-2019-001904] [Medline: 31803511]
33. Nguyen C, Grappasonni I, Scuri S, Nguyen B, Nguyen T, Petrelli F. Immunization in Vietnam. Ann Ig 2019;31(3):291-305 [FREE Full text] [doi: 10.7416/ai.2019.2291] [Medline: 31069373]
34. Seymour D, Werner L, Mwansa F, Bulula N, Mwanyika H, Dube M, et al. Electronic Immunization Registries in Tanzania and Zambia: Shaping a Minimum Viable Product for Scaled Solutions. Front Public Health 2019;7:218 [FREE Full text] [doi: 10.3389/fpubh.2019.00218] [Medline: 31440494]
35. Berkley S. Immunization needs a technology boost. Nature 2017 Nov 16;551(7680):273. [doi: 10.1038/d41586-017-05923-8] [Medline: 29144479]
36. Dumit EM, Novillo-Ortiz D, Contreras M, Velandia M, Danovaro-Holliday MC. The use of eHealth with immunizations: An overview of systematic reviews. Vaccine 2018 Dec 18;36(52):7923-7928 [FREE Full text] [doi: 10.1016/j.vaccine.2018.06.076] [Medline: 29983255]
37. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health Care Provider Adoption of eHealth: Systematic Literature Review. Interact J Med Res 2013;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]

Abbreviations

BMGF: Bill and Melinda Gates Foundation
EIR: electronic immunization registry
GDPM: General Department of Preventive Medicine
MAPS: mHealth Assessment and Planning for Scale
M&E: monitoring and evaluation
MNO: mobile network operator
MOH: Ministry of Health
NEPI: The Vietnam National Expanded Program on Immunization
NIIS: Vietnam's National Immunization Information System
TWG: technical working group
WHO: World Health Organization

Edited by G Eysenbach; submitted 07.05.20; peer-reviewed by T Collins, M Das, N Jarour; comments to author 12.06.20; revised version received 07.08.20; accepted 10.08.20; published 22.09.20.

Please cite as:
Dang H, Dao S, Carnahan E, Kawakyu N, Duong H, Nguyen T, Nguyen D, Nguyen L, Rivera M, Ngo T, Werner L, Nguyen N Determinants of Scale-up From a Small Pilot to a National Electronic Immunization Registry in Vietnam: Qualitative Evaluation J Med Internet Res 2020;22(9):e19923
URL: http://www.jmir.org/2020/9/e19923/
doi:10.2196/19923
PMID:32960184

Original Paper

Health Goal Attainment of Patients With Chronic Diseases in Web-Based Patient Communities: Content and Survival Analysis

Jiahe Song[1], PhD; Pei Xu[2], PhD; David B Paradice[2], PhD

[1]Haworth College of Business, Western Michigan University, Kalamazoo, MI, United States

[2]Harbert College of Business, Auburn University, Auburn, AL, United States

Corresponding Author:
Pei Xu, PhD
Harbert College of Business
Auburn University
420 Lowder Hall
Auburn, AL, 36849
United States
Phone: 1 3348446513
Email: pzx0002@auburn.edu

Abstract

Background: Activities directed at attaining health goals are a major part of the daily lives of those fighting chronic diseases. A proliferating population of patients with chronic diseases are participating in web-based patient communities, wherein they can exchange health information and pursue health goals with others virtually.

Objective: In this study, we aimed to understand the effect of participation in social media–enabled web-based patient communities on health goal attainment. In particular, we studied the antecedents of health goal attainment in terms of social support and self-reflection in web-based patient communities.

Methods: This data set consists of web-based health management activities of 392 patients across 13 health support groups, that is, groups with medical issues such as high blood pressure, diabetes, and breast cancer; the data of the activities were collected from a leading web-based patient community. Content analysis was used to code the social interactions among the patients on the web-based platform. Cox regression for survival analysis was used to model the hazard ratio of health goal attainment.

Results: Our analysis indicated that emotional support from web-based patient communities can increase patients' probability of achieving their goals (hazard ratio 1.957, 95% CI 1.416-2.706; $P<.001$) while informational support does not appear to be effective ($P=.06$). In addition, health-related self-reflection increases the patients' likelihood of goal attainment through web-based patient communities (hazard ratio 1.937, 95% CI 1.318-2.848; $P<.001$), but leisure-oriented self-reflection reduces this likelihood (hazard ratio 0.588, 95% CI 0.442-0.784; $P<.001$).

Conclusions: Social media–enabled web-based platforms assist health goal management via both social interaction and personal discipline. This study extends the understanding of web-based patient communities by investigating the effects of both social and cognitive factors on goal attainment. In particular, our study advocates that health goals relating to chronic conditions can be better managed when patients use the facilities of web-based health communities strategically.

(J Med Internet Res 2020;22(9):e19895) doi:10.2196/19895

KEYWORDS

web-based patient communities; self-reflection; social support; goal attainment; web-based chronic disease management; survival analysis

Introduction

Background

The modern day increase in life expectancy in the global population is coincident with an increase in chronic conditions and diseases. According to a recent report by the Centers for Disease Control and Prevention, 6 in 10 adults in the United States have a chronic disease, which is the leading driver of the nation's $3.5 trillion in annual health care costs [1]. To restrain the growing economic burden, the traditional cost-ineffective medical service model, in which health experts administer complex treatments to patients and guide them in follow-up

self-care, has gradually shifted toward more collaborative, community, and individual empowerment approaches [2,3].

One such approach is community-based goal striving support on social media platforms. Health goal attainment in the face of chronic diseases has long been recognized as an important indicator of patients' health. Based on a report, 90% of the participants used goals to manage their health, and 1 in 3 of them intended to know more about how to eventually attain their health goals [4]. With the increasing popularity of social media–enabled web-based patient communities, an increasing number of patients, especially those with chronic illnesses, are joining their peers in goal-striving activities. However, despite their popularity, the effectiveness of web-based patient communities on the self-managed goal attainment of patients with chronic diseases is largely characterized by a lack of investigation.

From a sociopsychological perspective, the outcome of a person's health management is dependent on both social influence and self-influence [5]. During the goal-striving process in web-based patient communities, social support from peers on these web-based platforms and patients' self-reflection are the 2 essential influencers for goal attainment. Social support functions through peer interaction on social media, which could potentially reduce patients' uncertainty and stress. Self-reflection is a cognitive process, which offers a comparison mechanism [6,7]; therefore, the goal setter can identify what needs to be improved or what needs to be included for goal attainment. In this study, we empirically investigate the impact of social support and self-reflection on goal attainment by conducting a survival analysis on a unique data set collected from a leading web-based patient community.

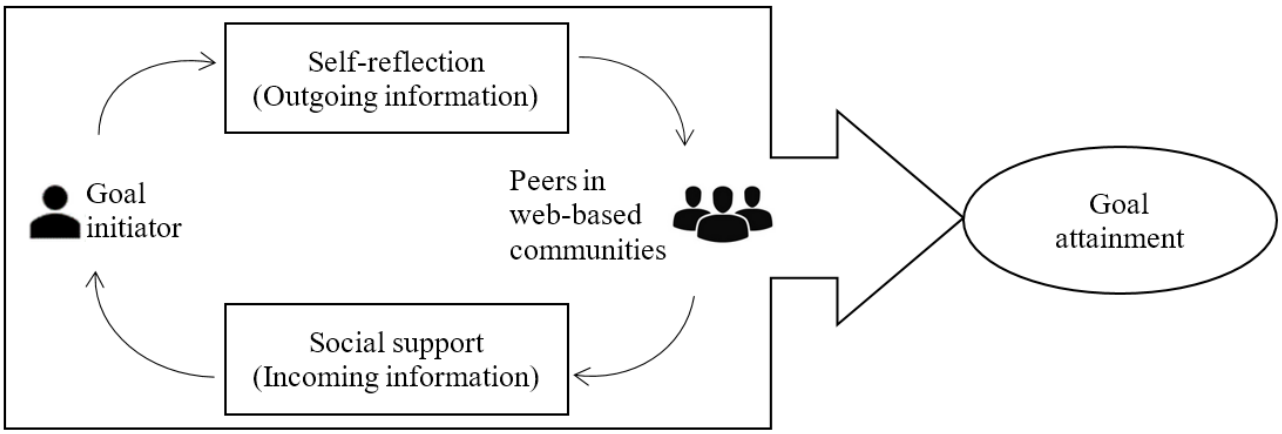
Hypotheses

Disease Self-Management and Web-Based Health Goal Attainment

Disease self-management is a patient-centered approach that reduces health care costs while improving the health status [8]. Such illness management is especially crucial for patients with chronic diseases as they may endure a long period of care and treatment. Most of the past research on disease self-management was related to patient education programs provided by health care professionals [9]. Internet and social media technology have made it easier to observe patients' personal and interpersonal behaviors in web-based disease management. For instance, many web-based patient communities allow patients to monitor their goal-striving activities while developing web-based dialogues with others.

In health care, goal attainment is often viewed as the final completion of goals and purposes for improving one's well-being [10,11]. Individuals can set any goal based on their health concerns, some of which are more intrinsically motivated (eg, walking goals, calorie goals) [11,12] while some are related to more extrinsic medical decisions (eg, treatment-aligned goals) [13,14]. In this study, web-based health goal attainment is the act of achieving health-beneficial goals, either fully or partially intrinsically motivated, through the use of a web-based patient community. Examples of such goals are "finish current treatment" for patients with tumors, "be emotionally stable" for patients with bipolar disorder, and "lose weight" for patients with high blood pressure. Extant studies tend to focus on the antecedent roles of goals such as the effect of a goal's presence or absence [15] and the impact of the distance to goal attainment [15,16]. However, knowledge is lacking on goal attainment as an outcome variable. In this study, we propose a web-based goal attainment model (Figure 1) to understand the impact of social support and self-reflection on goal attainment in web-based patient communities.

Figure 1. Self-reflection and social support driving goal attainment in web-based patient communities.



Social Support and Goal Attainment

Social support refers to "a coping resource from which people may draw when handling stressors" [17]. In the web-based environment, social support can be exchanged between any 2 individuals with social ties. Social support can be viewed as social cognitive means in health practices [5]. In web-based patient communities, informational support and emotional support are frequently generated by patients [18].

Informational support refers to the type of assistance that helps define, comprehend, and cope with stressful problems (ie, health conditions in this research) [19,20]. Web-based informational

support can be in the form of advice, referrals, or teaching [21]. According to the social judgment theory [22], one might expect that people put more weight on opinions closer to their own while discounting more distant ideas [23,24]. With regard to achieving a health goal, the support receiver (ie, the goal initiator) is likely to receive practical, health-oriented information from those who have the same health issue, if not with exactly the same goal. It can be expected that the support receiver will consider such information seriously. Research also shows that conformity in opinions is important in people's decision-making processes. For patients, similarly, the informational support of peers is based on peers' understanding and experience of a given medical condition, which is likely to be helpful for the goal initiator. In addition, prior research has suggested that informational support is given to a patient to reduce the uncertainty he/she is facing and to guide his/her action [25] and should be beneficial for those intending to improve their health [26]. Thus, we anticipate the goal attainment likelihood to be positively influenced by a high volume of the received informational support in web-based patient communities. Therefore, we proposed the following hypothesis.

Hypothesis 1. Web-based informational support is positively related to goal attainment by users of web-based patient communities.

Emotional support is affective and sentimental in nature and communicates love or care [26]. It is usually expressed through understanding/empathy, encouragement, affirmation/validation, sympathy, and caring/concern [21]. Although it does not contain any constructive advice or suggestion, emotional support can help patients reduce their negative feelings due to any issues or reaffirm their self-efficacy due to noticeable progress. This type of social support acts like a stimulant directly affecting the patients' mood state and emotion. Better mood will, in turn, help patients achieve a better state of health [27]. Besides, researchers have found that emotional support has a long-lasting effect of retaining patients in web-based patient communities [18]. Due to such emotional attachments, patients are more likely to keep striving on their goals in their virtual communities. Thus, the received emotional support from web-based patient communities is likely to increase the patients' chances in attaining their health goals, which leads to our next hypothesis.

Hypothesis 2. Web-based emotional support is positively related to goal attainment by users of web-based patient communities.

Self-Reflection and Goal Attainment

Self-reflection is within the umbrella of self-regulation [28] in the goal-related literature. Self-regulation is the "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" [29]. A comparison mechanism exists in a person's cognition to constantly reflect and evaluate his/her goal-striving activities [6,30], helping him/her move closer toward the goal. Such a comparison mechanism is self-reflection [31] through which goal progress is generated. In social cognitive theory, Bandura [32] noted that monitoring performance has little meaning to goal attainment if the comparison mechanism in the human cognitive process is lacking. For patients pursuing health goals in web-based patient communities, their self-evaluated or self-reflected content is a critical asset indicating whether these individuals continue moving their goal-striving endeavor forward.

Because web-based patient communities largely enable interpersonal activities, it is common for patients using web-based platforms to post both health-related content and leisure-oriented content. During the health goal-striving process, the former is the reflection directly related to the person's health or goal, whereas the latter is the cognition process regarding other things in life such as recalling childhood memories, mentioning family events, and commenting on a person. In other words, leisure-oriented self-reflection largely focuses on non–health-related topics that are loosely connected to the health goal. The literature on self-regulation posits that distraction draws one's attention away from monitoring health goals and the related behaviors, thereby causing reduced ability in goal pursuit [33]. Prior studies also indicated that successfully controlling one's attention is beneficial for goal performance [34]. In web-based patient communities, a patient's self-reflection content can indicate the person's attention regulation status. Put simply, if a patient can generate more health-related reflection content, his/her goal regulation process is considered stable and his/her goal-pursuing ability is increased. Otherwise, his/her ability of goal attainment will be hindered. In light of this, we proposed the following 2 hypotheses:

Hypothesis 3. Health-related self-reflection content is positively related to goal attainment.

Hypothesis 4. Leisure-oriented self-reflection content is negatively related to goal attainment.

Methods

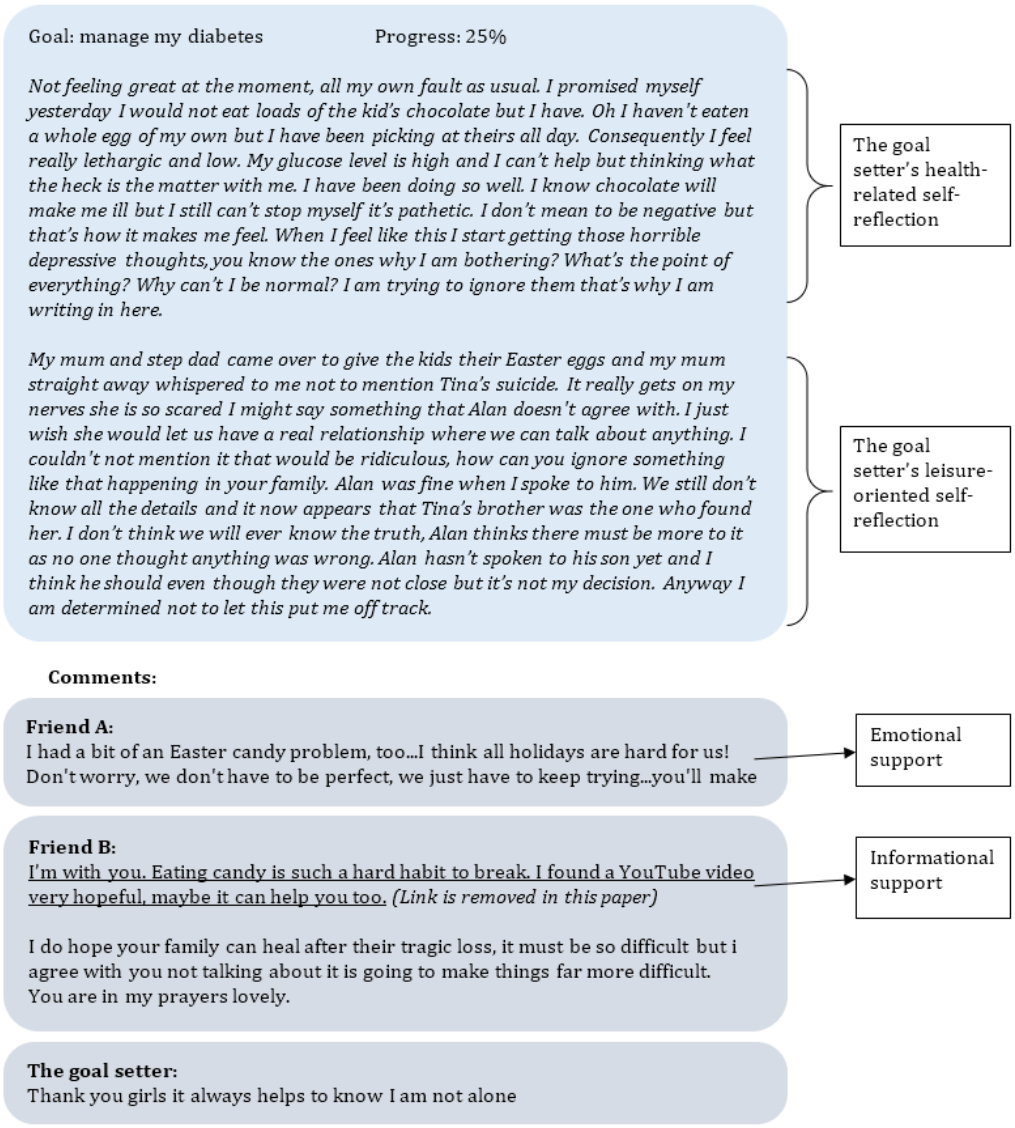
Research Context and Data Collection

We collected data from a large web-based patient community, which launched goal management in December 2007. Prior to the data collection, the website had more than 400,000 members participating in over 500 health support groups. The support groups were generated by the platform, and each group had a unique medical or distress focus, the majority of which were chronic diseases (eg, mental disorder, cancer, diabetes). This web-based patient community allowed its members to choose health goals that were closely related to their distress to manage their health. The available goals were created by a group of health care professionals hired by this web-based platform.

At the beginning of a goal, the progress was set at 0%. Over time, patients using this web-based platform could change the progress in 5% increments. The goals were managed only by these patients. Figure 2 demonstrates a patient's goal page. It includes a progress diagram and goal-related journals. The patient's self-reflection content and the received social support are available to read on each journal page. We employed a web crawler to collect data from this patient community. The details of the data collection can be found in Multimedia Appendix 1. In total, 392 patients with 5 different goal types were included in the panel data set. Among them, 87 patients managed to

complete their goals within the research window, while the remaining 305 left their goals unfinished.

Figure 2. Goal progress diagram and goal update records on a patient's goal page.



Variables and Content Analysis

Dependent variables: The dependent variable, $\lambda(t)$, is the hazard rate measuring the probability that a patient will experience goal attainment at time t. It is derived from goal duration, which represents the time spent from the goal start date to the end date. If a patient experienced health goal attainment before the end of the study period, then the end date is the actual goal end date. Otherwise, the patients are right-censored and the end date is the last day of the research window. Goal duration is assumed to have a continuous probability distribution $f(t)$. Thus, the probability that the duration will be at least t, $S(t)$, is defined as shown in . Since hazard rate is the probability that the duration will end after time t, given that it has lasted until time t, we define $\lambda(t) = f(t)/S(t)$.

Independent variables: (1) Informational represents the number of times peers in web-based patient communities provide informational support to the patient. (2) Emotional represents the number of times peers in web-based patient communities give emotional support to the patient. (3) HealthUpdate is the number of health-related contents posted in a patient's goal update entries to measure health-related self-reflection. (4) LeisureUpdate is the number of leisure-oriented contents posted in the goal update entries to measure the leisure-oriented self-reflection. These variables were derived from a content analysis, which is presented in Multimedia Appendix 2.

Control variables: (1) Tenure indicates the length in time since a patient became a member of the web-based patient community and is measured by days. (2) NumberOfUpdate shows the total number of goal updates a patient posted until the goal attainment was censored or the end of the research window. (3) HealthResponse captures the number of patient's health-related responses to peers in web-based patient communities in the comments section. (4) LeisureResponse captures the number of patient's leisure-oriented responses to peers in web-based patient communities in the comments section. (5) Age in years, gender (0-male/1-female as reported by the website), and goal category dummy variables were also included. The descriptive

statistics, the Pearson correlation matrix, and the variance inflation factors (VIFs) for the variables are presented in Multimedia Appendix 3. The correlations are less than 0.8, indicating that a strong correlation is not a significant concern. VIF results (mean VIF 1.74, individual VIF range 1.03-2.99) indicated that multicollinearity is not an issue in this study.

Survival Analysis

Because the research focus is the final success in achieving health goals (ie, health goal attainment), survival and hazard ratio analysis is the appropriate approach. We used a semiparametric model, Cox proportional hazard regression, to analyze the health goal attainment data, as the predictors in the model are all continuous variables [35]. The model is specified as shown in  . In this model, $\lambda_0(t)$ is a baseline hazard function that describes the risk for patients with $X_i = 0$, who serve as a reference cell.  is the relative risk, a proportionate increase or reduction in risk, associated with the set of characteristics X_i. The model for this research with all the X_i and control variables is shown below, where controls represents all the control variables in the research.

We conducted the Schoenfeld and scaled Schoenfeld residuals plots individually to check the proportionality assumption on each independent variable. *P* values greater than .05 for each term indicate that all of the terms should be kept in the model. We plotted the Cox-Snell residuals to check the goodness of fit. The plot indicated that the model fits well.

We computed the statistical power for a two-sided, .05 significance level test to detect an effect size (log hazard ratio) of 1 for each of the independent variables, given that our sample size is 392. After adjusting the censoring rate and the covariance among variables, we found that the statistical powers for the independent variables were all greater than 98%, which is well above the suggested 80% level. Thus, we conclude that our sample size is sufficient to investigate the effect of the independent variables in a Cox proportional hazard regression.

Results

The Cox proportional hazard regression results are reported in Table 1. The results of the 4 independent variables were consistent across all 3 models, that is, the main effects were significantly related to the dependent variable in this study, except informational support. The probability of obtaining the chi-square statistic, given that there was no effect of the independent variables, taken together, on the dependent variable, was <.001 for all the 3 models.

In Hypothesis 1, we proposed that informational support boosts the chance of goal attainment for web-based patient community users. As shown in Table 1, the coefficient for logInformational is negative and not significant (*P*=.06). Thus, we did not find support for Hypothesis 1. This is probably because practical health information from nonprofessional peers in web-based patient communities does not efficiently guide one's health goal-striving process. As shown in the results, the coefficient of logEmotional is positive and significant (hazard ratio 1.957, *P*<.001), suggesting that emotional support is positively related to goal attainment. Thus, Hypothesis 2 is supported. We used natural log for the logarithmic transformation of the variables. Thus, this result suggests that for every 3 additional emotional support comments received, the probability of goal attainment for the patient increases by 95.7%. The results show that the coefficients for health-related self-reflection and leisure-oriented self-reflection are 0.661 (hazard ratio 1.937, *P*<.001) and –0.530 (hazard ratio 0.588, *P*<.001), respectively. These findings indicate that health-related self-reflection is positively related to goal attainment, while leisure-oriented self-reflection is negatively related to goal attainment. Thus, Hypothesis 3 and Hypothesis 4 are also supported.

Two control variables, Tenure and NumberOfUpdate, are negatively related to goal attainment. For Tenure, this negative effect may be caused by the reduced interest level of older members in web-based patient communities. Over time, patients may become less motivated to use the web-based goal management features. Thus, patients with longer tenure may not frequently interact with the platform and thus delay their goal progress. Next, the negative effect of NumberOfUpdate indicates that given a fixed amount of social support and self-reflection, a patient who updates more frequently on a goal is less likely to attain the goal. In web-based patient communities, since the actual goal progress does not necessarily change in each update, frequent updating may be a result of excessive web-based self-presentation [36], which could be counterproductive to the goal attainment process [37].

The results also show that certain goal categories affect goal attainment. We then checked the Kaplan-Meier survival estimates graph (Figure 3) to obtain a better visualization. A test for equality of survival functions among the groups suggests that the observed differences are significant (*P*=.01). It appears that a goal may be better managed by a patient if it is highly associated with their health providers' treatment plan (eg, beat Hepatitis C virus, complete cancer treatment). This is possibly because these goals are associated with clearer guidance, procedures, and timelines. Although beyond the scope of this study, the level of manageability of a goal due to either intrinsic factors or an extrinsic cause should be taken into account for chronic condition management.

Table 1. Estimates of the Cox proportional hazard model on the rate to goal attainment occurrence.

Variable	Model 1[a] (independent variables), N=392		Model 2[b] (control variables), n=379[c]		Model 3[d] (full model), n=379	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
logInformational	–0.288 (0.211)	.17	—[e]	—	–0.420 (0.220)	.06
logEmotional	0.349 (0.131)	.008	—	—	0.672 (0.165)	<.001
logHealthUpdate	0.540 (0.166)	.001	—	—	0.661 (0.197)	.001
logLeisureUpdate	–0.608 (0.133)	<.001	—	—	–0.530 (0.146)	<.001
logTenure	—	—	–1.847 (0.246)	<.001	–1.861 (0.256)	<.001
logNumberOfUpdate	—	—	–0.500 (0.240)	.04	–1.147 (0.300)	<.001
logHealthResponse	—	—	–0.255 (0.416)	.54	–0.681 (0.443)	.12
logLeisureResponse	—	—	–0.455 (0.515)	.38	–0.297 (0.542)	.58
logAge	—	—	–0.005 (0.306)	.99	–0.125 (0.336)	.71
Gender (female)	—	—	0.035 (0.312)	.91	–0.027 (0.328)	.93
Goal to lose weight	—	—	–0.574 (0.259)	.03	–0.660 (0.277)	.02
Goal to manage diabetes	—	—	0.099 (0.433)	.82	0.202 (0.483)	.68
Goal to beat Hepatitis C virus	—	—	0.816 (0.417)	.05	0.351 (0.433)	.42
Goal for cancer treatment	—	—	0.830 (0.438)	.06	0.929 (0.451)	.04

[a]Log-likelihood function, –407.5; Bayesian information criterion, 783.324.

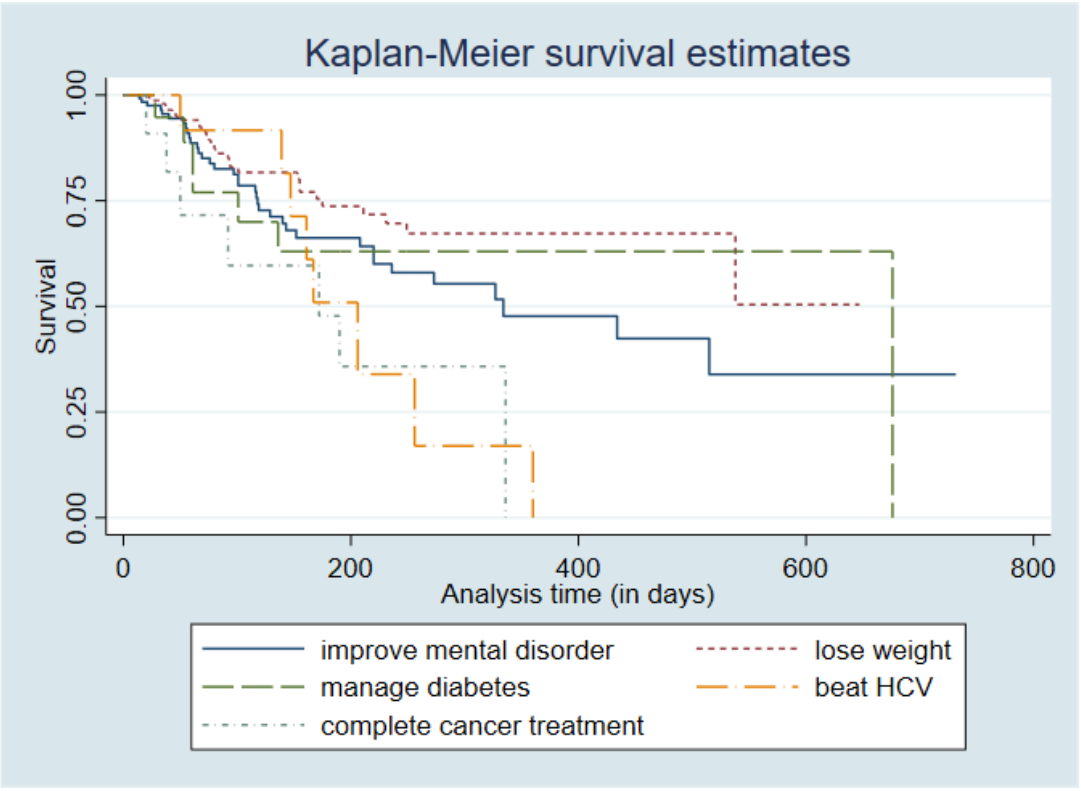
[b]Log-likelihood function, –362.0; Bayesian information criterion, 838.935.

[c]13 values were missing for age and gender.

[d]Log-likelihood function, –334.9; Bayesian information criterion, 752.851.

[e]Not available.

Figure 3. Kaplan-Meier survival estimates by goal types. HCV: hepatitis C virus.



Discussion

Principal Findings

In this study, we found that the received emotional support and health-related self-reflection in a web-based patient community increase a person's probability of reaching a health goal. The findings also confirmed the detrimental effect of leisure-oriented content on goal attainment. We now discuss the implications for research and practice and limitations.

Implications for Research and Practice

Our study makes several research contributions. First, this study adds new insights to health goal attainment in web-based patient communities. We conducted a survival study on the likelihood of health goal attainment with the presence of web-based social support and self-reflection. The design of the study allowed us to examine the effects of both environmental (peer support) and cognitive (self-reflection) factors on patients' behavior related to goal pursuit. To the best of our knowledge, this is one of the first studies investigating health goal attainment in the social media setting. Second, although extant studies recognize the crucial effect of social support on patients' health [26], it is unclear whether social support is beneficial for patients' self-striving behaviors (eg, health goal attainment). Our study contributes to the literature by revealing the different effects of social support (ie, informational and emotional support) on patients' goal attainment. More importantly, we found that emotional support is more beneficial than informational support in helping patients in their health goal pursuit. This means that crowd-powered positive effect may be more crucial than health information provided by the crowd (ie, those who are not health care professionals) for patients' goal management. Additionally, this research echoes the previous research [38] on patients' cognitive capabilities (eg, the amount of information to be processed) and pushes it further to connect it with a more specific health outcome, goal attainment, which provides building blocks for future researchers who are interested in patients' health outcomes enabled by web-based cognitive processes.

This study also provides important practical implications. First, we suggest that patients in web-based communities setting their goals in a more specific manner as a purposeful, manageable objective are likely to benefit more from social/cognitive influencers in terms of goal attainment. Second, although setting health goals, especially health goals for chronic conditions, does not always result in winning the battle with a disease, being able to finish a goal has significant meaning in patients' disease management and overall health. Therefore, if web-based patient communities can be integrated with offline professional health coaching methods [39], patients may realize better goal results by taking advantage of both professional-based informational support and peer-based emotional support. Lastly, if hospitals and physicians consider adding peer social interaction functionality to their web-based portals (through patient login), their patients may have a more focused and trustworthy group of peers with whom they can exchange information. Better goal performance can be expected as one of the potential benefits of using such a platform.

Limitations and Future Research

This study has several limitations. First, we collected data from groups with active conditions to observe goal-striving activities. However, the less active groups may carry their unique characteristics that were not uncovered by our data set. Because additional desirable data are truly hard to obtain, more robustness checks may need to be conducted to help verify the results of the survival analysis. Second, although several control variables were included in this study, future studies should consider more control variables (eg, number of support givers, number of strong ties and weak ties) to reduce the omitted variable bias. Third, this study focused on goal attainment, which is the ending point of a goal. To gain more insights on goal attainment in web-based patient communities, future research can shed light on the entire process of health goal management. For example, future research can investigate whether web-based patient communities play a role in goal creation and goal progress. Such studies will help researchers and practitioners in web-based patient communities better understand how a patient's goal proceeds over time and how to provide intervention during the process. Fourth, to the best of our knowledge, this study is among the first to empirically include patient self-reflection in web-based patient community research. An important reason for this factor to be excluded from the prior studies is that self-reflection is highly personal and cognition-oriented. Put simply, a large portion of a person's self-reflection process for self-managed goals is not web-based and is undocumented. In this study, we were restricted to capturing only the observable portion of self-reflection through patients' web-based activities. For future studies, multiple data sources (eg, web-based, clinical settings, interview, survey) should be implemented to reduce this issue.

Acknowledgments

The authors thank Dr. Terry Byrd for providing comments and suggestions at the early stage of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details about the selected goals and disease groups.
[DOCX File , 16 KB - jmir_v22i9e19895_app1.docx]

Multimedia Appendix 2
Content Analysis.
[DOCX File , 19 KB - jmir_v22i9e19895_app2.docx]

Multimedia Appendix 3
Other related statistics.
[DOCX File , 23 KB - jmir_v22i9e19895_app3.docx]

References

1. CDC. Chronic diseases in America. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm [accessed 2020-04-17]
2. Wu T, Chen Z, Zhang D, Wu X, Wang R. Correction: Predictors of Patients' Loyalty Toward Doctors on Web-Based Health Communities: Cross-Sectional Study. J Med Internet Res 2019 Nov 07;21(11):e16942 [FREE Full text] [doi: 10.2196/16942] [Medline: 31697641]
3. Skousen T, Safadi H, Young C, Karahanna E, Safadi S, Chebib F. Successful Moderation in Online Patient Communities: Inductive Case Study. J Med Internet Res 2020 Mar 17;22(3):e15983 [FREE Full text] [doi: 10.2196/15983] [Medline: 32181743]
4. Welltok. New national survey shows that seniors' desire to improve health doesn't fade with age. PR Newswire. 2015. URL: http://www.prnewswire.com/news-releases/new-national-survey-shows-that-seniors-desire-to-improve-health-doesnt-fade-with-age-300222245.html [accessed 2020-08-27]
5. Bandura A. Health promotion by social cognitive means. Health Education & Behavior 2004 Apr;31(2):143-164 [FREE Full text] [doi: 10.1177/1090198104263660] [Medline: 15090118]
6. Pressman S, Cohen S. Does positive affect influence health? Psychol Bull 2005 Nov;131(6):925-971. [doi: 10.1037/0033-2909.131.6.925] [Medline: 16351329]
7. Maes S, Karoly P. Self-Regulation Assessment and Intervention in Physical Health and Illness: A Review. Applied Psychology 2005 Apr;54(2):267-299 [FREE Full text] [doi: 10.1111/j.1464-0597.2005.00210.x]
8. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
9. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
10. Kiresuk T, Sherman RE. Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. Community Ment Health J 1968 Dec;4(6):443-453 [FREE Full text] [doi: 10.1007/bf01530764]
11. Cameron D, Bertenshaw EJ, Sheeran P. Positive affect and physical activity: Testing effects on goal setting, activation, prioritisation, and attainment. Psychol Health 2018 Feb;33(2):258-274. [doi: 10.1080/08870446.2017.1314477] [Medline: 28436287]
12. Huang S, Aaker J. It's the journey, not the destination: How metaphor drives growth after goal attainment. J Pers Soc Psychol 2019 Oct;117(4):697-720. [doi: 10.1037/pspa0000164] [Medline: 31233318]
13. Malec JF. Goal Attainment Scaling in Rehabilitation. Neuropsychological Rehabilitation 1999 Jul;9(3-4):253-275 [FREE Full text] [doi: 10.1080/096020199389365]
14. Nápoles AM, Santoyo-Olsson J, Chacón L, Stewart AL, Dixit N, Ortiz C. Feasibility of a Mobile Phone App and Telephone Coaching Survivorship Care Planning Program Among Spanish-Speaking Breast Cancer Survivors. JMIR Cancer 2019 Jul 09;5(2):e13543 [FREE Full text] [doi: 10.2196/13543] [Medline: 31290395]
15. Locke E, Latham GP. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist 2002 Sep;57(9):705-717 [FREE Full text] [doi: 10.1037/0003-066x.57.9.705]
16. Goes P, Guo C, Lin M. Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange. Information Systems Research 2016 Sep;27(3):497-516 [FREE Full text] [doi: 10.1287/isre.2016.0635]
17. Thoits P. Stress, Coping, and Social Support Processes: Where Are We? What Next? Journal of Health and Social Behavior 1995;35:53-79 [FREE Full text] [doi: 10.2307/2626957]
18. Wang YC, Kraut R, Levine JM. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012 Feb 11 Presented at: ACM 2012 Conference on Computer Supported Cooperative Work; Feb 11, 2012; New York, NY p. 833-842. [doi: 10.1145/2145204.2145329]
19. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychological Bulletin 1985 Sep;98(2):310-357 [FREE Full text] [doi: 10.1037/0033-2909.98.2.310]
20. Thoits P. Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 2011 Jun;52(2):145-161. [doi: 10.1177/0022146510395592] [Medline: 21673143]

21. Bambina A. Online social support: the interplay of social networks and computer-mediated communication. Youngstown, NY: Cambria Press; 2007.
22. Sherif M, Hovland CI. Social judgment: assimilation and contrast effects in communication and attitude change. Oxford, England: Yale University Press; 1961.
23. Perloff RM. The dynamics of persuasion: communication and attitudes in the 21st century (4th ed.). New York, NY: Routledge; 2010.
24. Yaniv I. Receiving other people's advice: Influence and benefit. Organizational Behavior and Human Decision Processes 2004 Jan;93(1):1-13 [FREE Full text] [doi: 10.1016/j.obhdp.2003.08.002]
25. Huang K, Chengalur-Smith I, Ran W. Not Just for Support: Companionship Activities in Healthcare Virtual Support Communities. CAIS 2014;34:561-594 [FREE Full text] [doi: 10.17705/1cais.03429]
26. Yan L, Tan Y. Feeling Blue? Go Online: An Empirical Study of Social Support Among Patients. Information Systems Research 2014 Dec;25(4):690-709 [FREE Full text] [doi: 10.1287/isre.2014.0538]
27. Zimmerman BJ. Attaining self-regulation: a social cognitive perspective. In: Boekaerts M, Pintrich PR, Zeidner M, editors. Handbook of Self-Regulation. New York, NY: ScienceDirect: Academic Press; 2000:13-39.
28. Miller GA, Galanter E, Pribram KH. Plans and the structure of behavior. New York, NY: Adams Bannister Cox; 1986.
29. Scheier MF, Carver CS. Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. Health Psychology 1985;4(3):219-247 [FREE Full text] [doi: 10.1037/0278-6133.4.3.219]
30. Huang K, Chengalur-Smith I, Pinsonneault A. Sharing Is Caring: Social Support Provision and Companionship Activities in Healthcare Virtual Support Communities. MISQ 2019 Jan 1;43(2):395-423 [FREE Full text] [doi: 10.25300/misq/2019/13225]
31. Febbraro G, Clum GA. Meta-analytic investigation of the effectiveness of self-regulatory components in the treatment of adult problem behaviors. Clinical Psychology Review 1998 Mar;18(2):143-161 [FREE Full text] [doi: 10.1016/s0272-7358(97)00008-1]
32. Bandura A. Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):248-287 [FREE Full text] [doi: 10.1016/0749-5978(91)90022-l]
33. Terry ML, Leary MR. Self-compassion, self-regulation, and health. Self and Identity 2011 May 25;10(3):352-362. [doi: 10.1080/15298868.2011.558404]
34. Diehl M, Semegon AB, Schwarzer R. Assessing attention control in goal pursuit: a component of dispositional self-regulation. J Pers Assess 2006 Jun;86(3):306-317 [FREE Full text] [doi: 10.1207/s15327752jpa8603_06] [Medline: 16740114]
35. Cox D. Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;34(2):187-202 [FREE Full text] [doi: 10.1111/j.2517-6161.1972.tb00899.x]
36. große Deters F, Mehl MR, Eid M. Narcissistic power poster? On the relationship between narcissism and status updating activity on Facebook. Journal of Research in Personality 2014 Dec;53:165-174 [FREE Full text] [doi: 10.1016/j.jrp.2014.10.004]
37. Morf C, Rhodewalt F. Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model. Psychological Inquiry 2001 Oct;12(4):177-196 [FREE Full text] [doi: 10.1207/s15327965pli1204_1]
38. Yan L, Peng J, Tan Y. Network Dynamics: How Can We Find Patients Like Us? Information Systems Research 2015 Sep;26(3):496-512 [FREE Full text] [doi: 10.1287/isre.2015.0585]
39. Olsen J, Nesbitt BJ. Health Coaching to Improve Healthy Lifestyle Behaviors: An Integrative Review. Am J Health Promot 2010 Sep;25(1):e1-e12 [FREE Full text] [doi: 10.4278/ajhp.090313-lit-101]

Abbreviations

VIF: variance inflation factor

Edited by G Eysenbach; submitted 05.05.20; peer-reviewed by T Skousen, H Safadi, MA Bahrami; comments to author 18.06.20; revised version received 03.07.20; accepted 22.07.20; published 11.09.20.

Please cite as:
Song J, Xu P, Paradice DB
Health Goal Attainment of Patients With Chronic Diseases in Web-Based Patient Communities: Content and Survival Analysis
J Med Internet Res 2020;22(9):e19895
URL: http://www.jmir.org/2020/9/e19895/
doi:10.2196/19895
PMID:32915152

Original Paper

Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study

François-Emery Cotté[1], PharmD, PhD, MPH; Paméla Voillot[2], MSc; Bryan Bennett[3], PhD; Bruno Falissard[4,5,6,7], MD, PhD; Christophe Tzourio[8,9], MD, PhD; Pierre Foulquié[2], MSc; Anne-Françoise Gaudin[1], PharmD; Hervé Lemasson[1], MSc; Valentine Grumberg[10], MSc; Laura McDonald[11], PhD; Carole Faviez[2], MSc; Stéphane Schück[2], MD

[1]Bristol-Myers Squibb, Rueil-Malmaison, France
[2]Kap Code, Paris, France
[3]Bristol-Myers Squibb, Stockport, United Kingdom
[4]Paris-Sud University, Paris, France
[5]Paris-Descartes Universitiy, Paris, France
[6]AP-HP, Paris, France
[7]INSERM U1178, Paris, France
[8]Bordeaux University, Bordeaux, France
[9]Inserm U1219, Bordeaux, France
[10]UFR Pharmacy, Grenoble Alpes University, Grenoble, France
[11]Bristol-Myers Squibb, London, United Kingdom

Corresponding Author:
Paméla Voillot, MSc
Kap Code
28 rue d'Enghien
Paris, 75010
France
Phone: 33 624058742
Email: pamela.voillot@kapcode.fr

Abstract

Background: Immune checkpoint inhibitors (ICIs) are increasingly used to treat several types of tumors. Impact of this emerging therapy on patients' health-related quality of life (HRQoL) is usually collected in clinical trials through standard questionnaires. However, this might not fully reflect HRQoL of patients under real-world conditions. In parallel, users' narratives from social media represent a potential new source of research concerning HRQoL.

Objective: The aim of this study is to assess and compare coverage of ICI-treated patients' HRQoL domains and subdomains in standard questionnaires from clinical trials and in real-world setting from social media posts.

Methods: A retrospective study was carried out by collecting social media posts in French language written by internet users mentioning their experiences with ICIs between January 2011 and August 2018. Automatic and manual extractions were implemented to create a corpus where domains and subdomains of HRQoL were classified. These annotations were compared with domains covered by 2 standard HRQoL questionnaires, the EORTC QLQ-C30 and the FACT-G.

Results: We identified 150 users who described their own experience with ICI (89/150, 59.3%) or that of their relative (61/150, 40.7%), with 137 users (91.3%) reporting at least one HRQoL domain in their social media posts. A total of 8 domains and 42 subdomains of HRQoL were identified: Global health (1 subdomain; 115 patients), Symptoms (13; 76), Emotional state (10; 49), Role (7; 22), Physical activity (4; 13), Professional situation (3; 9), Cognitive state (2; 2), and Social state (2; 2). The QLQ-C30 showed a wider global coverage of social media HRQoL subdomains than the FACT-G, 45% (19/42) and 29% (12/42), respectively. For both QLQ-C30 and FACT-G questionnaires, coverage rates were particularly suboptimal for Symptoms (68/123, 55.3% and 72/123, 58.5%, respectively), Emotional state (7/49, 14% and 24/49, 49%, respectively), and Role (17/22, 77% and 15/22, 68%, respectively).

Conclusions: Many patients with cancer are using social media to share their experiences with immunotherapy. Collecting and analyzing their spontaneous narratives are helpful to capture and understand their HRQoL in real-world setting. New measures

of HRQoL are needed to provide more in-depth evaluation of Symptoms, Emotional state, and Role among patients with cancer treated with immunotherapy.

(J Med Internet Res 2020;22(9):e19694) doi:10.2196/19694

KEYWORDS

health-related quality of life; immunotherapy; patients with cancer; social media use; measures; real world

Introduction

Health-related quality of life (HRQoL) is a complex subjective concept pertaining to multiple domains including physical, emotional, social, professional, and functional well-being [1,2]. The number of cancer cases is continuing to rise across Europe with approximately 8% of people currently living with cancer [3] and with the side effects of cancer treatment (eg, hair loss, pain, fatigue, nausea) [4]. Patients and their families also often experience psychological distress as a result of a cancer diagnosis and its treatment, including stress, anxiety, and depression [5]. Therefore, HRQoL is particularly important to understand and to be quantified among people living with cancer or caring for someone with cancer. Several cancer-related HRQoL self-administered questionnaires, such as the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) [6] and the Functional Assessment of Cancer Therapy - General (FACT-G) [7], have been used and validated in oncology patient populations, but are limited in that they are required to be complete at a predefined time.

Recently, the internet has provided an opportunity for introducing new sources of clinically relevant information related to patients' perception of illness and its burden [8]. Specifically, patients and their relatives are increasingly using forums, blogs, and social media to obtain health-related information and support [9]. Indeed, the information generated online represents an alternative way to understand patients' and relatives' health state compared with self-administered questionnaires. These patient-generated health data are produced spontaneously—thus not limited to medical consultations, for instance—mostly anonymously, and may better correspond to patients' and relatives' feelings than close-ended questions. Moreover, text mining techniques applied to analyze social media data can be used with relative ease [10] and have opened up new opportunities to bridge the gap between qualitative and quantitative data [11].

In parallel to the development of new strategies for collecting information on HRQoL of patients with cancer, treatment with immunotherapy has emerged as an innovative curative approach that, instead of destroying cancer cells directly, stimulates the immune system making it capable of identifying and selectively attacking cancer cells. The body develops an internal defense mechanism which can lead to reduced side effects and improved HRQoL [12]. Recent studies have assessed HRQoL of patients undergoing immunotherapy using standard questionnaires, but exclusively only in clinical trial settings [13-17]. In addition to prolonged survival, HRQoL results showed that immune checkpoint inhibitors (ICIs), such as nivolumab, maintained or improved baseline HRQoL levels in patients with advanced melanoma [18], advanced renal cell carcinoma [19], or advanced squamous non-small-cell lung cancer [20]. However, HRQoL of patients treated with ICI remains largely unknown under real-world conditions and it is not known whether the EORTC QLQ-C30 and the FACT-G questionnaires capture the full range of HRQoL domains and experiences relevant to these specific patients. The use of social media data offers a novel opportunity to explore this.

This study assessed and compared conceptual coverage of ICI-treated patients' HRQoL between standard questionnaires and users' ICI-related experiences described in social media. We hypothesized that, given the evolving and dynamic nature of the HRQoL concept, new HRQoL subdomains would emerge from social media posts going beyond the coverage of existing questionnaires, especially in a population of patients treated with new drugs such as ICI.

Methods

Study Design and Population

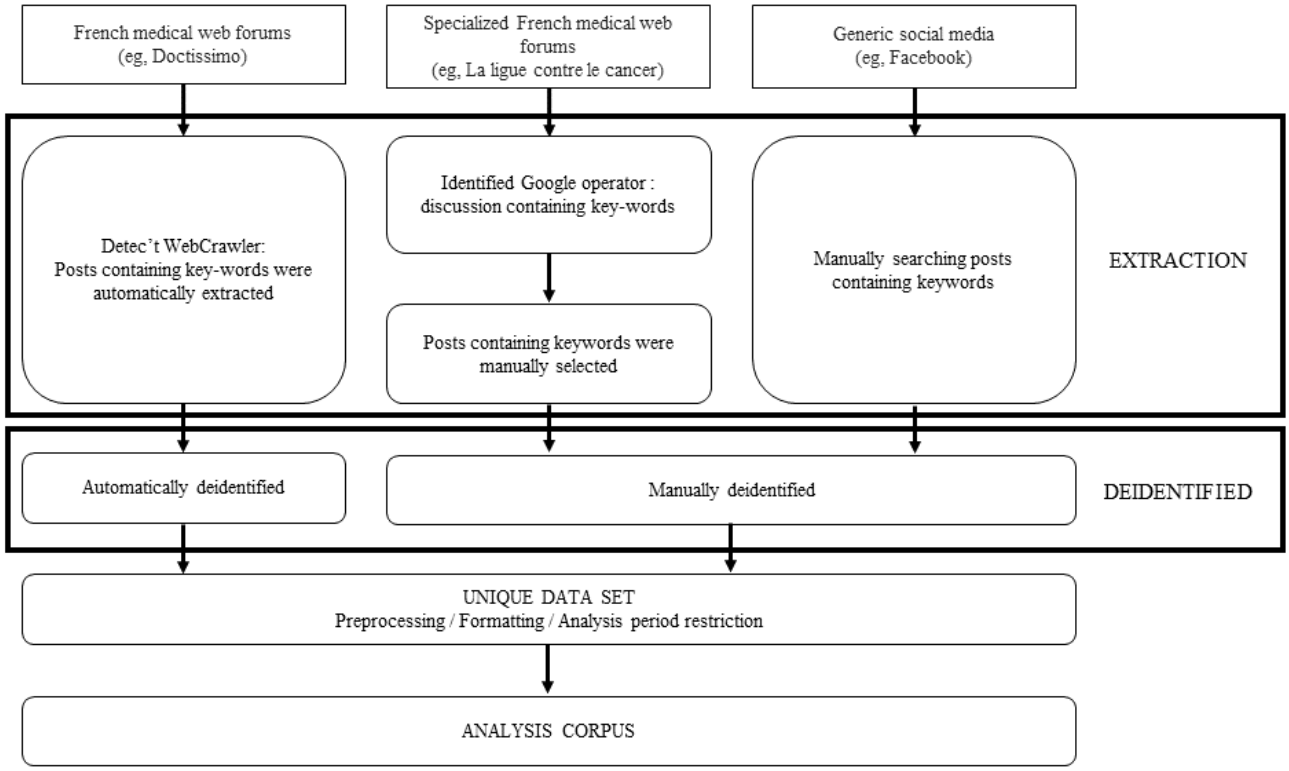
This was a retrospective study using a text mining approach to retrieve information from social media posts written by French language internet users between January 1, 2011, and August 31, 2018. The start date of January 2011 was selected because it corresponds to the date of marketing authorization for selling the first available immuno-oncology treatment in France. Included posts (comprising forum posts or comments on videos or photos) had to mention past or current patients' experience with ICI (ie, ipilimumab, nivolumab, pembrolizumab, atezolizumab, or those available through early access schemes or clinical trials, such as durvalumab, tremelimumab, and avelumab). Posts referring to treatments other than ICI were excluded. Posts could be authored by patients themselves or by their relatives, here interchangeably referred to as *patients*. Posts were retrieved from the following social media: 12 generic French medical web forums; 4 cancer-specialized French medical web forums; and 3 generic social media (Facebook, YouTube, and Instagram). These social media were screened from the Detec't database [21]. Description of forums is provided in Multimedia Appendix 1. All posts had to be publicly available and include at least one of the predetermined keywords with their synonyms (see the list in Multimedia Appendix 2). Ambiguous posts or duplicates (ie, similar posts posted by the same user in different social media) were excluded through a manual review.

Data Extraction

Automatic and manual extraction methods were utilized, depending on the source of the posts. For generic French medical web forums, posts were automatically extracted using the Detec't WebCrawler [22]. Collection of the posts was performed

according to the HTML structure of each forum. All posts containing one of the predefined keywords were automatically retrieved from discussion threads and deidentified (signature and quote withdrawal). The deidentification of posts was performed by using an in-house algorithm based on regular expression to automatically identify specific sequences of characters (proper names, phone numbers, postal codes, mail addresses, etc). For specialized French medical web forums, posts containing predefined keywords were identified using a Google operator searching these keywords in selected forum websites (Site: +URL). Identified discussion threads were manually explored, and posts containing one of the predefined keywords were manually retrieved and deidentified. Finally, for the 3 generic social media, posts were identified and retrieved by manually searching the predefined keywords using the social media embedded search fields. These posts were also manually deidentified. Data from these 3 collection methods were then grouped in a unique data set (the analysis corpus), which went through several steps of cleaning (preprocessing by removing French accents, unnecessary spaces, and punctuation and lowercasing all words; removal of stop words; and stemming, based on Porter's algorithm [23]) and formatting (transformation of the corpus into a matrix; creation of tokens and measure of their frequency; exclusion of hardly used, ambiguous, and misspelled words; and document-term matrix weighting; Figure 1).

Figure 1. Diagram of the steps for creating the analysis corpus.



Study Variables

The analysis corpus contained information on data source (name of the forum or social media), post characteristics (URL of the page or discussion where the post was published; date of the publication; pseudonym or alias of the user; keyword associated with the post leading to its extraction), patient characteristics (age, gender, type of cancer), and user status among "patient," "relative," and "unspecified." Regarding content of retrieved posts, for each patient "associated HRQoL domains" and "associated subdomains" were collected manually. "Associated HRQoL domains" were grouped considering the classification of domains provided in existing measures (ie, the EORTC QLQ-C30 and the FACT-G). For each domain, "associated subdomains" mentioned by patients were also collected even if not included in standard measures in order to allow new subdomains to emerge.

Comparison With HRQoL Standard Questionnaires

The EORTC QLQ-C30 [6] and the FACT-G [7] are among the most widely used questionnaires to capture HRQoL of patients with cancer in research and clinical settings [24,25]. Compared with other questionnaires, they are not limited to a specific cancer type [26,27] and cover the highest number of HRQoL domains across all cancer-specific questionnaires [28]. Both questionnaires were first developed in 1993 [29] and are validated in the French language [30,31].

The EORTC QLQ-C30 is a 30-item questionnaire composed of multi-item scales and single items, and comprises 5 functional scales (physical activity, role, emotional state, cognitive state, and social state), 3 symptoms scales (fatigue, nausea and vomiting, and pain), and a global health status and HRQoL scale. The remaining single items assess additional symptoms commonly reported by patients with cancer: dyspnea, insomnia, lack of appetite, constipation, diarrhea, and financial difficulties. The FACT-G is a 27-item questionnaire divided into 4

well-being subscales: physical well-being, social/family well-being, emotional well-being, and functional well-being [32]. Coverage of domains identified in social media and related subdomains measured by the EORTC QLQ-C30 and the FACT-G was manually assessed independently by 2 operators and compared through a concept mapping approach [33].

Data Analysis

Each post corresponded to a statistical unit. Frequentist analysis was performed on extracted posts to characterize the whole analysis corpus through the following indicators: number of posts, number of patients, occurrence of HRQoL domain(s), keywords (including the number of extracted posts), data source, and users' characteristics. A Venn diagram was generated through the CRAN package "VennDiagram." The list of identified subdomains, including the number of patients and posts, was presented in a descriptive format using Microsoft Excel. For each subdomain, coverage by one or several items of the 2 questionnaires was assessed. For each domain, coverage rates were calculated by dividing the sum of subdomain occurrences covered by questionnaires by the total number of occurrences in the social media. The diagram indicating the coverage rates of each HRQoL domain retrieved from social media posts through the standard questionnaires EORTC QLQ-C30 and FACT-G was generated through Microsoft Excel.

Results

Description of the Population and Posts

The final analysis corpus included 267 social media posts meeting the inclusion criteria, with a maximum of 11 posts from 1 patient and a median of 2 posts per patient. Through the manual extraction, we identified 150 patients (posters) who described their personal experience with ICI (89/150, 59.3%) or that of their relative (61/150, 40.7%). A majority of patients were women (82/150, 54.7%) and gender was undetermined for only 8/150 patients (5.3%). The type of related cancer was identified for 123/150 patients (82.0%): the most frequent cancers were lung cancer and melanoma (Table 1). The majority of posts were retrieved from 1 cancer-specific patient forum (La ligue contre le cancer, 78 posts by 43 patients) and 1 generic medical forum (Doctissimo, 76 posts by 43 patients). The most frequently identified keyword was "immunotherapy" (72/150, 48.0%) followed by "nivolumab" (31/150, 20.7%) and "ipilimumab" (20/150, 13.3%).

Table 1. Patients characteristics.

Characteristics	Patients mentioning their quality of life (N=150)
Gender, n (%)	
Women	82 (54.7)
Men	60 (40.0)
Undetermined	8 (5.3)
Type of cancer, n (%)	
Lung cancer	46 (30.7)
Melanoma	41 (27.3)
Others	36 (24.0)
Undetermined	27 (18.0)
Type of immunotherapy, n (%)	
Nivolumab	31 (20.7)
Ipilimumab	20 (13.3)
Pembrolizumab	16 (10.7)
Others	11 (7.3)
Undetermined	72 (48.0)
Posts number, n	267
Per patient, median (min-max)	2 (1-11)
Social media, n (%)	
Doctissimo	43 (28.7)
La ligue contre le cancer	43 (28.7)
Facebook	16 (10.7)
Others	48 (32.0)

Domains and Subdomains of HRQoL

Of the study population, 91.3% (137/150) mentioned at least one HRQoL domain in their posts. We identified 8 HRQoL domains, with 83.9% (115/137) patients mentioning Global health, 55.5% (76/137) mentioning Symptoms, 35.8% (49/137) mentioning Emotional state, 16.1% (22/137) mentioning Role, 9.5% (13/137) mentioning Physical activity, 6.6% (9/137) mentioning Professional situation, 1.5% (2/137) mentioning Social state, and 1.5% (2/137) mentioning Cognitive state. A Venn diagram illustrates the distribution of the HRQoL domains and the relationship among them (Figure 2). Verbatim examples for each HRQoL domains are presented in Table 2.

A total of 42 subdomains of HRQoL were mentioned in the social media posts and were distributed as follows: 1 for Global health, 13 for Symptoms, 4 for Physical activity, 10 for Emotional state, 7 for Role, 3 for Professional situation, 2 for Cognitive state, and 2 for Social state. Table 2 provides the list and occurrences of each social media subdomain. Among 115 patients mentioning their Global health, a majority (63/115, 54.8%) considered it to be poor, while 27.8% (32/115) and 17.4% (20/115) perceived it to be stable and good, respectively. As for Symptoms, 76 patients mentioned symptoms 123 times, which were grouped into 13 subdomains. The most addressed subdomain of Symptoms was Fatigue/Tiredness (34/123, 27.6%). A total of 49 patients mentioned at least one subdomain concerning Emotional state. With 15 occurrences, Hope was the most recurrent subdomain (15/49, 31%). The most frequent subdomains of Role were normal life/reduced activities and pace of life (12/22, 55%). Only 13 patients mentioned Physical activity and the most reported subdomains pertaining to this larger domain were minimal or no physical activity/maintained activity (5/13, 38%) and difficulty walking/eating (5/13, 38%). Professional situation was mentioned by 9 patients with the most frequent subdomain being sick leave (6/9, 67%). Two subdomains were mentioned for the Cognitive state and the Social state, respectively.

Table 2 also synthetizes coverage of each social media subdomain by items of the QLQ-C30 and the FACT-G questionnaires. The QLQ-C30 showed a wider coverage of social media subdomains than the FACT-G (45% [19/42] versus 29% [12/42], respectively).

Figure 2. Venn diagram showing the distribution of patients according to their posts mentioning HRQoL domains. HRQoL: health-related quality of life.

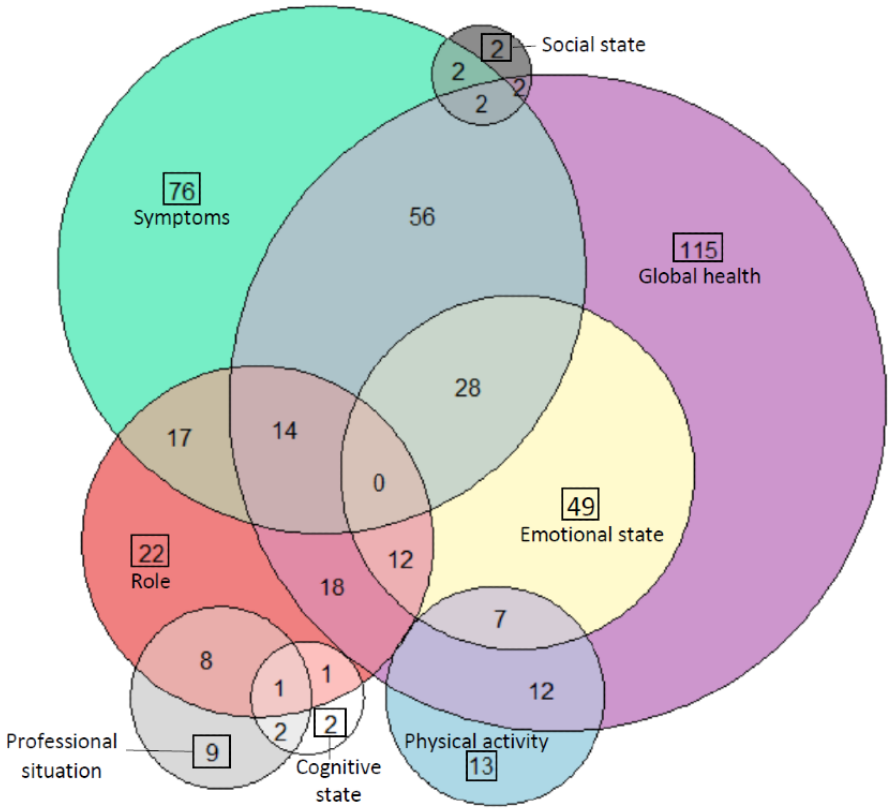


Table 2. Verbatim examples of posts related to HRQoL[a] of patients with cancer.

HRQoL domain and specific subdomain	Verbatim example (translation from French)
Global health	
Deteriorated	"Hello, we started immunotherapy 3 weeks ago, we didn't combine it with [Drug], his health is deteriorating more and more. Yesterday he was re-hospitalized...I don't think his condition could get any worse...I strongly hope he gets better, that he can eat and move..."
Symptoms	
Weight loss	"When I was diagnosed with melanoma in May 2010, the [hospital] immediately referred me to the [protocol] ([Drug] treatment). (...) it must be said that I lost 9 kilos after the 1st injection, then 2 kilos per month, until we stopped this treatment, in [month] 2011, for a total weight loss of 28 kilos. Since the beginning of the injections I had to stop working."
Emotional state	
Optimism	"I have lung cancer and metastasis (...). I have had immuno with [drug] for 69 weeks, I am quite well and the mass sometimes decreases slightly and sometimes I have almost no nodule on the lungs. I have almost no reactions as I do not lose my hair, no nausea, just a little bit. I am at stage 4, but I trust my treatment and I have a lot of hope."
Exhaustion	"I have been fighting lung cancer with bone metastases for more than three years. Since the beginning of July, I have been trying immunotherapy. If people have had this treatment, I would like to know what the side effects were. As for me, I am very tired, have no appetite and my moral is deteriorating."
Physical activity	
Minimal or no physical activity/maintained activity	"Cancer was diagnosed in [month] 2017. An operation was not possible. In [month] after the immunotherapy the saturation level was very low and the oxygen allowed maintaining a very minimal physical activity... a few steps..."
Role	
Normal life	"I have been receiving this treatment for 1 year and a half, I don't know if it will cure me but thanks to [Drug] I have been in remission for more than a year. Relatively non-toxic, it allows to lead a normal life in parallel."
Reduced activities and pace of life	"Hello. I have melanoma stage 4; for 3 years, interferon, [Drug], [Drug] currently every 15 days. I am tired, so exhausted that I can no longer take care of my house, but I have the rage to live."
Professional situation	
Maintained work activity	"In short, of course I'm afraid of dying, but that's out of the question. This type of cancer is treatable if taken in time. And I was in stage 4 Clark. Immunotherapy is not always cool, but I still work, I manage the house, the children and I am still smiling."
Cognitive state	
Concentration disorders	"Here I am, dean of this post: a year under [drug]. And no trace of melanoma! (...) If I had to relive it I would stop working earlier because I have the impression that [drug] has mostly prevented me from recovering, from sleeping well, with a cumulative effect on the end where I ended the year on the kneecaps wondering if I was not depressed so much I could not concentrate on anything."
Social state	
No longer participates in family parties	"On the first scan, increased lung metastases...but no new ones, so that's something. (...) We won't do much for the holidays. Until recently we used to go to my in-laws, there were a lot of us, but since my mother-in-law died and since I'm no longer in Olympic shape, we haven't been moving around!"

[a]HRQoL: health-related quality of life.

Coverage of HRQoL Domains in Social Media by the QLQ-C30 and FACT-G Questionnaires

As shown in Table 3, Global health was entirely covered by both the QLQ-C30 and the FACT-G. Physical activity, Professional situation, Cognitive state, and Social state were also fully covered by the QLQ-C30. For Symptoms, the EORTC QLQ-C30 covered a majority (68/123, 55.3%) of the subdomains identified in social media posts, and so did the FACT-G (72/123, 58.5%). Coverage was lower for the Emotional state domain: 14% (7/49) by the EORTC QLQ-C30 and 49% (24/49) by the FACT-G. Finally, the EORTC QLQ-C30 covered 77% (17/22) of subdomains for Role, whereas the FACT-G covered 68% (15/22). Of these domains, the FACT-G fully covered only the Professional situation, whereas it covered 46% (6/13) of the Physical activity, with no coverage of neither Cognitive state nor Social state.

Table 3. HRQoL[a] subdomains in social media posts and their coverage by EORTC QLQ-C30[b] and FACT-G[c] questionnaires.

HRQoL domains and subdomains in social media (number of occurrences in posts)	EORTC QLQ-C30 coverage (question number)	FACT-G coverage (question number)
Global health		
Deteriorated/stable/better (115)	✓ (Q29)	✓ (GF7)
Symptoms		
Fatigue/Tiredness (34)	✓ (Q10, Q12, Q18)	✓ (GP1)
None/Unspecified adverse events (23)	X	✓ (GP5)
Fever (8)	X	X
Pain/Joint pain/Leg pain/Liver pain (15)	✓ (Q9, Q19)	✓ (GP4)
Cough (7)	X	X
Loss of appetite/Weight loss (11)	✓ (Q13)	X
Rash/itch (6)	X	X
Respiratory trouble/Breathlessness (6)	✓ (Q18)	X
Headache (3)	X	X
Thyroid disorders (4)	X	X
Heavy legs (2)	X	X
Hair loss (2)	X	X
Diarrhea (2)	✓ (Q17)	X
Emotional state		
Optimism/Hope (15)	X	✓ (GE3)
Exhaustion (7)	X	X
Distress (6)	X	X
Good mental health/Morale (5)	X	✓ (GE1, GE2)
Fear (4)	✓ (Q22)	✓ (GE5, GE6)
Depression (3)	✓ (Q24)	X
Psychological disorders (3)	X	X
Stable health (3)	X	X
Isolate oneself (2)	X	X
Emotional exhaustion due to side effects (1)	X	X
Role		
Normal life/Reduced activities and pace of life (12)	✓ (Q6, Q7)	✓ (GF6, GF3)
Time constraints regarding medical care (3)	X	X
Financial problems (2)	✓ (Q28)	X
Maintained activities (2)	✓ (Q6, Q7)	✓ (GF6)
Home care constraints (1)	X	X
Improved HRQoL (1)	✓ (Q30)	✓ (GF7)
Ability to drive again (1)	X	X
Physical activity		
Minimal or no physical activity/maintained activity (6)	✓ (Q1, Q2, Q3, Q4)	✓ (GP1, GP7)
Difficulty walking/eating (5)	✓ (Q2, Q3, Q5)	X
Difficulty climbing stairs (1)	✓ (Q2, Q3, Q5)	X
Difficulty getting into bed alone (1)	✓ (Q2, Q3, Q5)	X
Professional situation		

HRQoL domains and subdomains in social media (number of occurrences in posts)	EORTC QLQ-C30 coverage (question number)	FACT-G coverage (question number)
Sick leave (6)	✓ (Q6)	✓ (GF1, GF2)
Maintained work activity (2)	✓ (Q6)	✓ (GF1, GF2)
Reduction in working time (1)	✓ (Q6)	✓ (GF1, GF2)
Cognitive state		
Concentration problems (1)	✓ (Q20)	X
Memory problems (1)	✓ (Q25)	X
Social state		
Requiring family assistance for medical care (1)	✓ (Q26)	X
Inability to participate in family celebrations (1)	✓ (Q26)	X

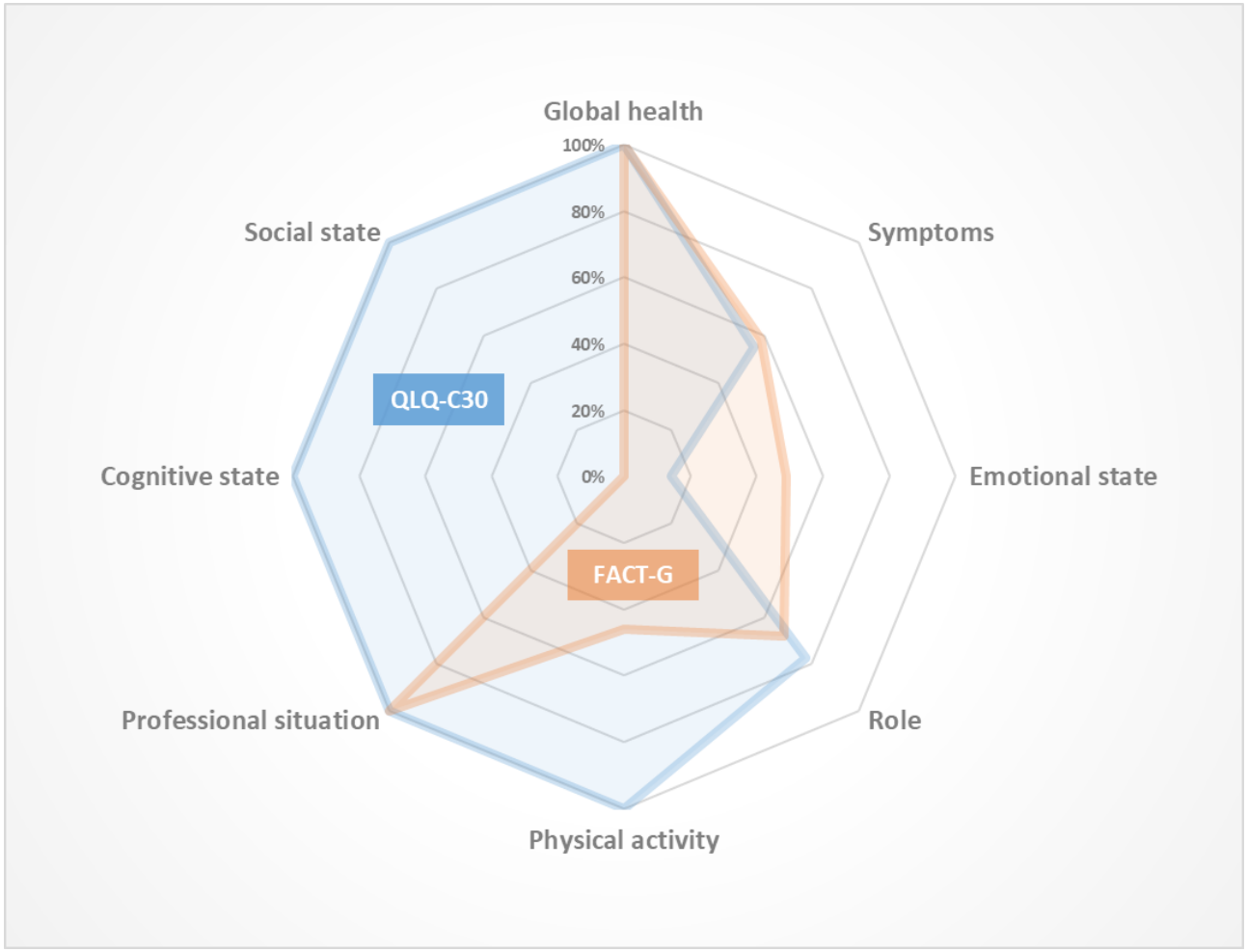
[a]HRQoL: health-related quality of life.

[b]EORTC QLQ-C39: European Organization for Research and Treatment of Cancer QLQ-C30.

[c]FACT-G: Functional Assessment of Cancer Therapy - General.

Specific subdomains which were not covered by both the EORTC QLQ-C30 and the FACT-G were fever, cough, rash/itch, headache, thyroid disorders, heavy legs, and hair loss for the Symptoms domain; exhaustion, distress, psychological disorders, stable health, emotional exhaustion due to side effects, and isolation for the Emotional state domain; and time constraints regarding medical care and ability to drive again for the Role domain (Figure 3).

Figure 3. Diagram indicating the coverage rates of each HRQoL domains retrieved in social media by the standard questionnaires EORTC QLQ-C30 and FACT-G. FACT-G: Functional Assessment of Cancer Therapy - General; QLQ-C30: Quality of Life Questionnaire Core 30.



Discussion

Principal Findings

This novel approach of using social media identified that some of the content posted by patients with cancer and caregivers with experience of ICI overlaps with concepts captured in the 2 most frequently used HRQoL questionnaires. However, the main findings also included the fact that there are a large number of concepts which are not captured in these 2 HRQoL questionnaires. These results confirmed our hypothesis by underlining the emergence of new subdomains of HRQoL in patients treated with immunotherapy. In particular, we observed that retrieved social media posts frequently addressed specific subdomains of the HRQoL domains of Symptoms, Emotional state, and Role, which are not fully covered by the EORTC QLQ-C30 and the FACT-G. Reasons for not including these subdomains in standard questionnaires are varied. First, questionnaires such as the EORTC QLQ-C30 and the FACT-G are designed to be short, effective, and time-saving, thus reducing burden on patients completing these questionnaires. As a consequence, they are limited and do not cover the whole range of the issues impacting patients' HRQoL. Second, these questionnaires were developed more than 25 years ago, which may explain why the domains might not fully capture the impact of recently developed therapies, such as ICI. The concept of HRQoL is evolving and several users have already raised the problem of "partial covering" [33], which relates to the complexity of measuring this broad ranging concept through a robust methodology [34]. Third, subdomains retrieved by our study might be specific to immunotherapy and, therefore, not measured by generic standard questionnaires. Finally, accurately quantifying an individual's HRQoL is, per se, a debated question, because standardized questionnaires might restrict a patient's choice and limit their spontaneity, thus not using a patient-centered approach [35].

We observed a remarkable usage in the occurrences of the keyword "immunotherapy" which likely results from the growing availability of ICI, the steadily increased use of ICI in the French health care system, and, in parallel, the growing proportion of social media users.

Comparison With the Literature on HRQoL and Immunotherapy

The conceptual and psychometric measurement properties of the EORTC QLQ-C30 and the FACT-G have not yet been systematically examined in ICI-treated patient populations and the results of this study cast some doubt on the content validity of these measures in ICI-treated patients. Indeed, content validity is the extent to which an instrument measures the important aspects of concepts most significant and relevant to a patient's condition and its treatment [36]. Because there are subdomains which do not appear in the 2 questionnaires that participants completed, we can assume that the 2 questionnaires lack content validity for this specific patient population. This does underline the need for new or adapted patient-reported outcomes in patients treated with immunotherapy.

Existing studies using these questionnaires in patients with cancer treated with immunotherapy are still often limited to research settings [17,18,37]. For instance, the work by Long and colleagues [18] has demonstrated that the use of nivolumab maintained baseline HRQoL levels to provide long-term quality of survival benefit among 418 patients with advanced melanoma. Cella and colleagues [19] have confirmed the association between nivolumab treatment and HRQoL improvement using the FACT-G among 847 patients with advanced renal cell carcinoma.

Although HRQoL has already been assessed with standard questionnaires in a number of clinical trials of immunotherapy, covered domains were pre-established and limited. Instead, in our study various new spontaneous subdomains emerged, such as fever, time constraints of treatment, difficulty in driving, and isolation. Furthermore, certain subdomains covered by the questionnaires might be inadequately designed for patients treated with ICI. For example, in the FACT-G, hope is collected in a negative way (ie, "I am losing hope in the fight against my illness"), whereas social media users mainly referred to hope in an optimistic way (ie, "regaining hope").

Limitations of the Study

This study was not without limitations. First, selection bias was a major limitation because analyses were restricted to selected data sources and available contents. The population under study was composed of social media users who might not necessarily reflect the characteristics of all patients with cancer receiving immunotherapy. However, because social media are increasingly used by patients [38], especially in France [39], retrieved posts should pertain to an important section of the French population. We also collected data from relatives who provided immunotherapy-related experiences of patients with cancer who were not active on social media. However, given the small number of patients and relatives in our sample, we were not able to distinguish their posts within the analysis corpus. Our results should then be interpreted considering this further limitation. HRQoL self-assessed by patients might be different from the evaluation provided by relatives, as shown in previous research [40]. Similarly, the limited size of our analysis corpus did not allow an analysis per type of cancer. Melanoma, for instance, has long been treated with immunotherapy [41], which means that HRQoL of patients with melanoma might be different from HRQoL related to other cancers.

Second, an extraction bias is also possible because we only considered posts containing predefining keywords. If users expressed their experiences with immunotherapy and consequent impact on HRQoL by using other nonspecific words, their posts were not included in the final analysis corpus. To mitigate this bias, the set of keywords was as comprehensive as possible.

Third, because this study was based on secondary use of data published in social media, it was impossible to get additional data and information from patients (only identified by their pseudonym). The analysis was then restricted to what users mentioned, which can lead to missing data or incomplete capture of the patients' full experience. In particular, subdomains emerging from social media were spontaneously addressed by users versus items from standard questionnaires. The fact that

some items of the EORTC QLQ-C30 or the FACT-G were not mentioned in social media posts does not mean that patients were not concerned by them. For this, our conclusions should be considered with a certain degree of caution.

Fourth, because immunotherapy is a fairly recent therapeutic approach, few posts could be identified. A larger analysis corpus is needed to obtain more robust results and to validate our initial findings. As demonstrated in this study, the number of posts concerning immunotherapy is increasing year after year and new studies will benefit from this expanding analysis corpus. In particular, posts should be compared across countries where the EORTC QLQ-C30 and the FACT-G are usually administered to capture HRQoL. Exploration of potential cross-country differences in subdomains mentioned in social media would be noteworthy.

Fifth, social media represent an ideal place where patients can freely and spontaneously discuss their experiences with their therapy, thus providing valuable information on their HRQoL. However, this observation should be interpreted cautiously, because social media data may include a higher frequency of erroneous information, and patients posting on social media forums may not be representative of the wider patient population [10].

Finally, biases related to semantic analyses must be considered. Given the low number of posts within our analysis corpus, we were obliged to retrieve and code the mentions manually and could not apply automated analysis, for example, using topic modeling.

Implications and Future Research

We were able to include users' subjective narratives in the evaluation of the impact of ICI on patients' HRQoL. The results of our study suggest that commonly used measures such as the EORTC QLQ-C30 and the FACT-G may require updating to improve their coverage and applicability of HRQoL domains under real-world conditions. The challenge in measuring HRQoL lies in its uniqueness to individuals [35] and questionnaires such as the EORTC QLQ-C30 and the FACT-G might not take account of this by imposing standardized models of HRQoL. For this reason, as already demonstrated in studies concerning other diseases than cancer [42-46], posts in online forums and social media should be integrated in the assessment of patients' HRQoL, because they can help either detect adverse events or characterize patient experience in a more individualized and spontaneous way.

In summary, this study suggested to explore further specific HRQoL domains related to patients treated with ICI to potentially enrich existing standard questionnaires with new items that are more relevant for these patients in their daily confrontation with disease and treatment.

Conclusion

Patients with cancer and their relatives are using social media to share their experiences with immunotherapy and its impact on HRQoL, particularly with regard to Global health and Symptoms. Emotional state and Role are also increasingly referenced in online forums and social media. Collecting and analyzing these spontaneous narratives can be helpful to capture how immunotherapy affects patients' HRQoL in a more individualized way, thus obtaining information on more facets of life that are important for patients. While standard questionnaires can provide objective scores, which are easily interpretable from a clinical and research point of view, mining social media posts might better inform health care professions and patients of the impact of immunotherapy on patients' HRQoL under real-world conditions. Future research is required to corroborate our findings and propose new individualized measures covering HRQoL more in depth than existing standard questionnaires.

Acknowledgments

The medical writing support was provided by Ilaria Montagni-Duchemin of the microenterprise Scientific Medical Writer IMD. The study was funded by Bristol-Myers Squibb.

Conflicts of Interest

F-EC, BB, LM, HL, and A-FG have disclosed that they are employees of Bristol-Myers Squibb. VG was a recipient of funding for studentships from Bristol-Myers Squibb. PV, PF, CF, and SS have disclosed that they are employees of Kap Code, a CRO, and are contracted with Bristol-Myers Squibb to carry out this study. BF and CT are contracted with Bristol-Myers Squibb to participate in this study.

Multimedia Appendix 1
Description of the websites used to retrieve the posts under study.
[DOCX File , 15 KB - jmir_v22i9e19694_app1.docx]

Multimedia Appendix 2
List of predefined keywords for posts inclusion.
[DOCX File , 13 KB - jmir_v22i9e19694_app2.docx]

References

1. Kaplan RM. Health status and health policy: Quality of life in health care evaluation and resource allocation. Qual Life Res 1994 Dec;3(6):457-458. [doi: 10.1007/bf00435398]
2. Noble H. Quality of life and health related quality of life – is there a difference? Evidence-Based Nursing. London: BMJ; 2014 Jan 27. URL: https://blogs.bmj.com/ebn/2014/01/27/quality-of-life-and-health-related-quality-of-life-is-there-a-difference/ [accessed 2020-08-28]
3. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 2018 Nov;103:356-387. [doi: 10.1016/j.ejca.2018.07.005] [Medline: 30100160]
4. International Agency for Research on Cancer. Cancer Tomorrow. Geneva: World Health Organization; 2019.
5. Heydarnejad MS, Hassanpour DA, Solati DK. Factors affecting quality of life in cancer patients undergoing chemotherapy. Afr Health Sci 2011 Jun;11(2):266-270 [FREE Full text] [Medline: 21857860]
6. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
7. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
8. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What Patients Can Tell Us: Topic Analysis for Social Media on Breast Cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
9. Brigo F, Igwe SC, Nardone R, Orioli A, Otte WM. Cancer information disparities on the internet: An infodemiological study. Journal of Cancer Policy 2016 Jun;8:33-37. [doi: 10.1016/j.jcpo.2016.04.002]
10. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med 2019 Jan 16;17(1):11 [FREE Full text] [doi: 10.1186/s12916-018-1247-8] [Medline: 30646913]
11. Wiedemann G. Opening up to big data: Computer-assisted analysis of textual data in social sciences. Forum: Qualitative Social Research (Forum Qualitative Sozialforschung) 2013;14(2):Art13 [FREE Full text] [doi: 10.17169/fqs-14.2.1949]
12. Hoos A, Britten CM. The immuno-oncology framework: Enabling a new era of cancer therapy. Oncoimmunology 2012 May 01;1(3):334-339 [FREE Full text] [doi: 10.4161/onci.19268] [Medline: 22737609]
13. Ramirez RA, Lu J, Thomas KEH. Quality of life for non-small cell lung cancer patients in the age of immunotherapy. Transl Lung Cancer Res 2018 Apr;7(Suppl 2):S149-S152 [FREE Full text] [doi: 10.21037/tlcr.2018.03.10] [Medline: 29782567]
14. Brahmer JR, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Health-related quality-of-life results for pembrolizumab versus chemotherapy in advanced, PD-L1-positive NSCLC (KEYNOTE-024): a multicentre, international, randomised, open-label phase 3 trial. The Lancet Oncology 2017 Dec;18(12):1600-1609. [doi: 10.1016/s1470-2045(17)30690-3]
15. Cohen EEW, Bell RB, Bifulco CB, Burtness B, Gillison ML, Harrington KJ, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of squamous cell carcinoma of the head and neck (HNSCC). J Immunother Cancer 2019 Jul 15;7(1):184 [FREE Full text] [doi: 10.1186/s40425-019-0662-5] [Medline: 31307547]
16. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar 21;145(3):765-774 [FREE Full text] [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
17. Anagnostou V, Yarchoan M, Hansen AR, Wang H, Verde F, Sharon E, et al. Immuno-oncology Trial Endpoints: Capturing Clinically Meaningful Activity. Clin Cancer Res 2017 Aug 31;23(17):4959-4969. [doi: 10.1158/1078-0432.ccr-16-3065]
18. Long G, Atkinson V, Ascierto P, Robert C, Hassel J, Rutkowski P, et al. Effect of nivolumab on health-related quality of life in patients with treatment-naïve advanced melanoma: results from the phase III CheckMate 066 study. Ann Oncol 2016 Oct;27(10):1940-1946 [FREE Full text] [doi: 10.1093/annonc/mdw265] [Medline: 27405322]
19. Cella D, Grünwald V, Escudier B, Hammers HJ, George S, Nathan P, et al. Patient-reported outcomes of patients with advanced renal cell carcinoma treated with nivolumab plus ipilimumab versus sunitinib (CheckMate 214): a randomised, phase 3 trial. The Lancet Oncology 2019 Feb;20(2):297-310. [doi: 10.1016/s1470-2045(18)30778-2]
20. Reck M, Taylor F, Penrod JR, DeRosa M, Morrissey L, Dastani H, et al. Impact of Nivolumab versus Docetaxel on Health-Related Quality of Life and Symptoms in Patients with Advanced Squamous Non-Small Cell Lung Cancer: Results from the CheckMate 017 Study. J Thorac Oncol 2018 Feb;13(2):194-204 [FREE Full text] [doi: 10.1016/j.jtho.2017.10.029] [Medline: 29129758]
21. Kürzinger ML, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, et al. Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis. J Med Internet Res 2018 Nov 20;20(11):e10466 [FREE Full text] [doi: 10.2196/10466] [Medline: 30459145]
22. Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach. J Med Internet Res 2018 Mar 14;20(3):e85 [FREE Full text] [doi: 10.2196/jmir.9222] [Medline: 29540337]

23. Ramasubramanian C, Ramya R. Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm. Int J Adv Res Computer and Commun Eng 2013;2(12):4536-4538 [FREE Full text]
24. Luckett T, King M, Butow P, Oguchi M, Rankin N, Price M, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. Ann Oncol 2011 Oct;22(10):2179-2190 [FREE Full text] [doi: 10.1093/annonc/mdq721] [Medline: 21339384]
25. Kuijpers W, Giesinger JM, Zabernigg A, Young T, Friend E, Tomaszewska IM, et al. Patients' and health professionals' understanding of and preferences for graphical presentation styles for individual-level EORTC QLQ-C30 scores. Qual Life Res 2016 Mar 9;25(3):595-604 [FREE Full text] [doi: 10.1007/s11136-015-1107-3] [Medline: 26353905]
26. Chopra I, Kamal KM. A systematic review of quality of life instruments in long-term breast cancer survivors. Health Qual Life Outcomes 2012 Jan 31;10(1):14 [FREE Full text] [doi: 10.1186/1477-7525-10-14] [Medline: 22289425]
27. Wheelwright S, Darlington A, Hopkinson JB, Fitzsimmons D, White A, Johnson CD. A systematic review of health-related quality of life instruments in patients with cancer cachexia. Support Care Cancer 2013 Sep 25;21(9):2625-2636. [doi: 10.1007/s00520-013-1881-9] [Medline: 23797577]
28. Ferrans CE. Differences in what quality-of-life instruments measure. J Natl Cancer Inst Monogr 2007 Oct 18;2007(37):22-26. [doi: 10.1093/jncimonographs/lgm008] [Medline: 17951227]
29. Fayers P, Bottomley A. Quality of life research within the EORTC—the EORTC QLQ-C30. Eur J Cancer 2002 Mar;38:125-133. [doi: 10.1016/s0959-8049(01)00448-8]
30. Conroy T, Mercier M, Bonneterre J, Luporsi E, Lefebvre J, Lapeyre M, et al. French version of FACT-G: validation and comparison with other cancer-specific instruments. Eur J Cancer 2004 Oct;40(15):2243-2252. [doi: 10.1016/j.ejca.2004.06.010]
31. Uwer L, Rotonda C, Guillemin F, Miny J, Kaminsky M, Mercier M, et al. Responsiveness of EORTC QLQ-C30, QLQ-CR38 and FACT-C quality of life questionnaires in patients with colorectal cancer. Health Qual Life Outcomes 2011 Aug 22;9(1):70 [FREE Full text] [doi: 10.1186/1477-7525-9-70] [Medline: 21859485]
32. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health Qual Life Outcomes 2003 Dec 16;1:79 [FREE Full text] [doi: 10.1186/1477-7525-1-79] [Medline: 14678568]
33. van Nieuwenhuizen C, Schene AH, Koeter MWJ, Huxley PJ. The Lancashire Quality of Life Profile: modification and psychometric evaluation. Soc Psychiatry Psychiatr Epidemiol 2001 Jan 12;36(1):36-44. [doi: 10.1007/s001270050288] [Medline: 11320806]
34. The Whoqol Group. Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. Psychol. Med 1998 May 01;28(3):551-558. [doi: 10.1017/S0033291798006667]
35. Carr AJ, Higginson IJ. Are quality of life measures patient centred? BMJ 2001 Jul 02;322(7298):1357-1360 [FREE Full text] [doi: 10.1136/bmj.322.7298.1357] [Medline: 11387189]
36. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. Value Health 2011 Dec;14(8):978-988 [FREE Full text] [doi: 10.1016/j.jval.2011.06.013] [Medline: 22152166]
37. Cella D, Grünwald V, Nathan P, Doan J, Dastani H, Taylor F, et al. Quality of life in patients with advanced renal cell carcinoma given nivolumab versus everolimus in CheckMate 025: a randomised, open-label, phase 3 trial. The Lancet Oncology 2016 Jul;17(7):994-1003. [doi: 10.1016/s1470-2045(16)30125-5]
38. De Martino I, D'Apolito R, McLawhorn AS, Fehring KA, Sculco PK, Gasparini G. Social media for patients: benefits and drawbacks. Curr Rev Musculoskelet Med 2017 Mar 21;10(1):141-145 [FREE Full text] [doi: 10.1007/s12178-017-9394-7] [Medline: 28110391]
39. 39 J. Les usages de l'Internet santé : vers une parentalité connectée ? La prévention toujours en re-création. À l'école la PMI 2016:155-186 [FREE Full text] [doi: 10.3917/eres.suess.2016.01.0155]
40. Minaya P, Baumstarck K, Berbis J, Goncalves A, Barlesi F, Michel G, et al. The CareGiver Oncology Quality of Life questionnaire (CarGOQoL): development and validation of an instrument to measure the quality of life of the caregivers of patients with cancer. Eur J Cancer 2012 Apr;48(6):904-911. [doi: 10.1016/j.ejca.2011.09.010] [Medline: 22033328]
41. Lugowska I, Teterycz P, Rutkowski P. Immunotherapy of melanoma. Contemp Oncol (Pozn) 2018 Mar;22(1A):61-67 [FREE Full text] [doi: 10.5114/wo.2018.73889] [Medline: 29628796]
42. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol Drug Saf 2013 Mar 16;22(3):256-262 [FREE Full text] [doi: 10.1002/pds.3365] [Medline: 23322591]
43. Schück S, Foulquié P, Mebarki A, Voillot P, Texier N, Faviez C. Détection automatique du mésusage des neuroleptiques dans le trouble anxieux et la démence à partir des réseaux sociaux. Revue d'Épidémiologie et de Santé Publique 2018 Jun;66:S211-S212. [doi: 10.1016/j.respe.2018.04.048]
44. Chen X, Faviez C, Schuck S, Lillo-Le-Louët A, Texier N, Dahamna B, et al. Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate. Front Pharmacol 2018;9:541 [FREE Full text] [doi: 10.3389/fphar.2018.00541] [Medline: 29881351]

45. Foulquié P, Voillot P, Faviez C, Mebarki A, Chen X, Texier N, et al. Étude de l'usage du méthylphénidate sur les réseaux sociaux. Revue d'Épidémiologie et de Santé Publique 2018 Jun;66:S225. [doi: 10.1016/j.respe.2018.04.030]
46. Schück S, Voillot P, Foulquié P, Faviez C, Mebarki A, Texier N, et al. Que nous apportent les réseaux sociaux quant à la crise sanitaire du Levothyrox ® d'août 2017 ? Revue d'Épidémiologie et de Santé Publique 2018 Jun;66:S225. [doi: 10.1016/j.respe.2018.04.029]

Abbreviations

EORTC: European Organization for Research and Treatment of Cancer
FACT-G: Functional Assessment of Cancer Therapy - General
HRQoL: health-related quality of life
ICI: immune checkpoint inhibitors
QLQ-C30: Quality of Life Questionnaire Core 30

Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by A Benetoli, M El Tantawi; comments to author 04.07.20; revised version received 10.07.20; accepted 26.07.20; published 11.09.20.

Please cite as:
Cotté FE, Voillot P, Bennett B, Falissard B, Tzourio C, Foulquié P, Gaudin AF, Lemasson H, Grumberg V, McDonald L, Faviez C, Schück S
Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study
J Med Internet Res 2020;22(9):e19694
URL: http://www.jmir.org/2020/9/e19694/
doi:10.2196/19694
PMID:32915159

Original Paper

Defining the Digital Self: A Qualitative Study to Explore the Digital Component of Professional Identity in the Health Professions

Brandon Ruan[1], BHSc; Yusuf Yilmaz[1,2], PhD; Daniel Lu[3], MD; Mark Lee[1], BHSc, DipCLS; Teresa M Chan[1,4,5], MD, FRCPC, MHPE, DRCPSC

[1]McMaster Education Research, Innovation, and Theory (MERIT), Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[2]Department of Medical Education, Faculty of Medicine, Ege University, Izmir, Turkey

[3]Psychiatry Residency Training Program, Department of Psychiatry, University of British Columbia, Vancouver, ON, Canada

[4]McMaster Program for Faculty Development / Office of Continuing Professional Development, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[5]Division of Emergency Medicine / Division of Education and Innovation, Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Teresa M Chan, MD, FRCPC, MHPE, DRCPSC
Division of Emergency Medicine / Division of Education and Innovation, Department of Medicine
Faculty of Health Sciences
McMaster University
237 Barton St E, McMaster Clinic, Room 255
Hamilton General Hospital
Hamilton, ON, L8L 2X2
Canada
Phone: 1 9055212100
Email: teresa.chan@medportal.ca

Abstract

Background: Recent medical education literature pertaining to professional identity development fails to reflect the impact social media has on professional identity theory. Social media is transforming the field of medicine, as the web-based medium is now an avenue for professional development and socialization for medical students and residents. Research regarding identity development in social media has been primarily confined to electronic professionalism through best practice guidelines. However, this neglects other potential aspects pertinent to digital identity that have not yet been explored.

Objective: This study aims to define the properties and development of the digital self and its interactions with the current professional identity development theory.

Methods: A qualitative study was conducted using thematic analysis. A total of 17 participants who are social media education and knowledge translation experts were interviewed. The initial participants were from emergency medicine, and a snowball sampling method was used following their respective web-based semistructured interviews to enable global recruitment of other participants from interprofessional disciplines. The research team consisted of a diverse group of researchers including one current social media knowledge translation physician clinician educator, one postdoctoral researcher who is regularly engaged in social media knowledge translation, and 3 nonphysician research assistants who are not social media users. Half of the team conducted the initial coding and analysis, whereas the other 2 investigators audited the procedures followed.

Results: A total of 4 themes were identified that pertain to digital identity. In the first theme, origins of initial digital identity formation were found to be derived from perceived needs in professional roles (eg, as a medical student or resident). The second theme consisted of the cultivation of digital identity, in which digital identity was developed parallel to professional identity. The third theme that emerged was the management between the professional and personal components of digital identity. Participants initially preferred keeping these components completely separate; however, attempts to do so were inadequate while the integration of both components provided benefits. The fourth theme was the management of real-life identity and digital identity. Participants preferred real-life identity to be wholly represented on the web. Instances of misalignment resulted in identity conflict, compromising one of the identities.

Conclusions: Social media introduces new features to professional identity in the digital world. The formation of digital identity, its development, and reconciliation with other identities were features captured in our analysis. The virtual component of

professional identity must not be neglected but instead further explored, as educational institutions continue to give more importance to navigating professional identity development.

(J Med Internet Res 2020;22(9):e21416) doi:10.2196/21416

KEYWORDS

professional identity; social media; digital identity; health care professionals; e-professionalism

Introduction

Professional identity has become an increasingly prominent phenomenon that is thought to be crucial to consider when educating health profession trainees [1-5]. Professional identity development is a recurring and adaptive process; personal identity that is constructed by one's internal values and morals molds and projects into the formation of the professional self, which encompasses role expectations and professional ethics [6]. For medical education, facilitating the strengthening of physician identity and internalizing physician values bring confidence into clinical practice and creates humanistic, compassionate, and ethical physicians [1,4]. The medical education literature has proposed models for navigating professional identity development and has documented its resiliencies [7,8]; however, recent work in this area does not explore the impact of new forms of expression, such as social media, on professional identity formation or development.

Opportunities for professional development through collaboration and networking have become more accessible, as social media platforms such as Twitter facilitate more continuous communication [9,10]. Furthermore, medical education literature suggests that the digital space also acts as an avenue for informal learning and socialization for medical trainees [11-13]. To navigate these spaces effectively, best practice guidelines have been published [14-18]. However, these guidelines often focus on confining trainees to a professional frame, often with an emphasis on preventing lapses in professionalism [19-21].

With social media altering the methods of learning, communication, and collaboration, it is imperative to investigate how professional identity fully manifests in the digital space. To continue guiding the next generation of physicians through professional identity development in an era of social media, institutions must aim to train them in the growing digital space in conjunction with the physical space.

Although there is an importance in practicing appropriate professional conduct on social media, continuously expressing the *professional self* may conflict with personal identity development—forcing an intricate interplay between personal identity and the professional self in an open forum that has previously been impossible [22]. It is within this area of tension that the components of professional identity may manifest uniquely in the digital space and that the establishment of one's professional self within a digital world may currently be unexplored.

The purpose of this study is to explore the digital component of professional identity: digital identity. We aim to extend the literature by defining the characteristics, properties, and development of digital identity and its interactions with the current professional identity development theory.

Methods

Research Paradigm

Through interviewing research experts and collecting their responses, we conducted a qualitative study to identify the underlying themes and phenomena guiding identity development using generic thematic analysis [23]. We applied an interpretivist lens to examine the recollections of our participants as they pertained to their personal and professional identities.

Research Team

The research team comprised 5 individuals: 1 physician (TMC) experienced in knowledge of the background literature on social media, teaching, and learning and in qualitative analysis, a postdoctoral researcher (YY) who is regularly engaged in social media knowledge translation and trained in qualitative analysis, and 3 nonphysician research assistants (BR, DL, and ML) trained in qualitative analysis but are not social media users. Half of the research team participated in the interpretation and analysis of the transcripts (TMC, BR, and DL), whereas the other half of the research team conducted an audit trail of the analysis and procedure (YY and ML).This dynamic of the research team was constructed to require the lead supervisor (TMC) to justify her interpretations of the themes to the other members (DL and BR) who hold no stake in the social media, teaching, and learning spheres.

Context

The targeted context pertinent to this study was the web-based community of social media–based educators and knowledge translation experts who are engaged in health profession education disciplines. The study initially started with social media educationalists from the emergency medicine field; however, the snowball sampling technique enabled nominations from interprofessional fields outside of emergency medicine [24]. The initial selection of web-based knowledge translation experts from the emergency medicine field was justified owing to their likelihood of having substantial commentary and experience on their digital identity origin and development based on the influential following they have built [25].

Sampling

Recruitment was initiated using a randomized selection of 10 physicians from a previously published list of the top 100 most influential emergency medicine physicians on Twitter [25]. Initial participants were contacted via institutional email or social media to participate in a semistructured interview that pertains to the individual's social media activities and origins

of social media scholarship. A snowball sampling technique was followed upon completing the initial participants' interview [24]. As the use of social media by knowledge translation experts and educators is rapidly evolving, the list by Riddell et al [25] may not fully capture the current social media sphere, therefore making the population ill-defined and an approach of snowball sampling to be appropriate [24]. The snowball sampling technique serves to better define our population of interest by requiring participants to identify and nominate other potential influencers within their social networks, who are also likely to have extensive experiences regarding digital identity origin and development. In the context of our study, the snowball sampling technique also enabled nominations from interprofessional fields, allowing us to capture social media knowledge translation experts in the health profession education disciplines beyond emergency medicine. Nominations were requested upon the completion of each participant's interview until thematic sufficiency was reached.

Ethics Approval

Ethics approval was granted by our institutional review board, the Hamilton Integrated Research Ethics Board (HIREB-5609).

Data Collection: Methods and Analysis

Interviews were conducted by the research assistants (BR and AM). The research assistants were trained in semistructured interviewing through simulations with the project lead. Semistructured interview prompts were constructed into the interview guide (Multimedia Appendix 1), and adaptations owing to missing gaps within the delivery of these prompts were made after the initial transcripts were reviewed. Participants did not review transcripts following their synthesis. All interviews were conducted and recorded using the Zoom (Zoom Video Communications, Inc) calling interface version 4.5.6, with audio being recorded [26].

Data Processing

Interviews were recorded, and the audio files were sent to a third-party professional medical transcriptionist. Transcripts were verified or corrected, if required by the interviewer and the analysis team, to confirm the accuracy of the transcripts.

Data Analysis

The research team reviewed the transcripts using a generic thematic analysis, deriving themes related to the identity formation of the participants. A total of 17 transcripts were obtained and analyzed for themes, 2 to 4 batches at a time over several months. Each coding session outputted a codebook that was saved, archived, and updated for the next session. The codebook was updated and restructured every session until thematic sufficiency was reached in concepts pertinent to digital identity.

Enhancing Rigor and Trustworthiness

Following the completion of transcript analysis, an audit was conducted by YY and ML to assess our analysis trail. Both auditors were given the full codebook and archives in addition to full primary transcripts. Owing to the ongoing COVID-19 pandemic [27,28], and the demographics of our participants being frontline health care workers, the initial plan to conduct a final member check of our analysis was revised. This study on digital identity is an umbrella study that consists of a larger project with 2 phases of interviews. For this particular study, only the first phase of the interview data was used for analysis. The division of this study was defined as *a priori*. Data analyses were separated to ensure that each unique component of the umbrella study reached data sufficiency. We adhered to the *Standards for Reporting Qualitative Research* checklist for reporting [29].

Results

Demographics

A total of 17 international social media experts were recruited and interviewed (male, 10/17, 59%; female, 7/17, 41%). Data surrounding their personal and professional identities were collected, including their Twitter followership. The mean interview duration was 30.6 (SD 7.98) min. The range of interview length was 18.6 min to 52.1 min. A total of 189 pages of transcripts were analyzed and reviewed. Most interviewees were partnered with dependents (9/17, 53%) or single with no dependents (5/17, 29%). Details on participant demographics are shown in Table 1.

Countries of academic or clinical practice included 47% (8/17) from the United States, 29% (5/17) from Canada, 12% (2/17) from Australia, and 12% (2/17) from other countries. The participants also varied across academic ranks: 35% (6/17) were assistant professors, 24% (4/17) were associate professors, 18% (3/17) were adjunct scientists, and 6% (1/17) were assistant professors. All participants identified their academic roles as educators (17/17, 100%). In addition, most participants also identified their academic roles as either a clinician (15/17, 88%) or a research investigator (13/17, 77%).

A multitude of social media platforms were identified by interviewees for use in both personal and professional contexts. The personal use of social media consisted of a majority that use Twitter (14/17, 82%), Facebook (14/17, 82%), and WhatsApp (11/17, 65%). In conjunction with their personal use, Twitter was also identified by all the participants for professional use (17/17, 100%), and most participants used Google Scholar (10/17, 59%). All social media platforms identified by interviewees for personal or professional use can be found in Table 2.

Major themes pertaining to digital identity are summarized in Table 3 and further elucidated below.

Table 1. Participants' demographics.

Demographic characteristics	Values
Gender, n (%)	
Female	7 (41)
Male	10 (59)
Characteristics of family and caregiving, n (%)	
Single, with no children or dependents	5 (29)
Single, with children or dependents	1 (6)
Partnered, with no children or dependents	2 (12)
Partnered, with children or dependents	9 (53)
Social media presence, median (IQR)	
Number of Twitter followers	6800 (3391-10,500)

Table 2. Frequencies of social media platforms for personal and professional use.

Social media platform used	Frequency	
	Personal use, n (%)	Professional use, n (%)
Twitter	14 (82)	17 (100)
Facebook	14 (82)	8 (47)
Google Scholar	10 (59)	10 (59)
WhatsApp	11 (65)	6 (35)
Slack	7 (41)	8 (47)
LinkedIn	7 (41)	7 (41)
ResearchGate	5 (29)	8 (47)
ORCID	5 (29)	8 (47)
Instagram	9 (53)	3 (18)
Reddit	3 (18)	3 (18)
Snapchat	2 (12)	1 (6)
Academia.edu	1 (6)	2 (12)

Table 3. Major themes in the construct of digital identity.

Themes	Summary	Representative quote
Initial formation of the digital identity	Most participants identified an initial hook into social media when starting their digital identity	"So, when I started off on social media, I was only on Twitter. I actually got into social media as almost a bet."
Cultivating digital identity	Digital identity evolves rapidly and can develop in parallel to professional identity	"Yeah, I guess my identity has changed on social media as my [professional] identity has changed throughout my training."
Real-life identity versus digital identity management	Participants noted the importance of representing themselves holistically on the web	"I am who I am. And I feel comfortable being who I am and representing myself, as such."
The professional and personal dimensions of social media	Some participants attempt to separate their professional and personal identity on the web; however, others demonstrate convergence is inevitable	"So, I used to keep things more separate and then and as my presence in social media has evolved, they have tended to grow closer together both the personal and the professional and merge into one because well that is who I am…"

Initial Formation of Digital Identity

Participants were asked to identify their origin of digital identity and how they began engaging in social media. Identified incentives guiding their initial use of social media were categorized into the following: (1) meeting an educational need and (2) facilitated via role responsibilities. These incentives acted as a hook for our participants into social media, as they continued their web-based activity once initialized into the digital space.

Meeting an Educational Need

Participants highlighted the educational value found in social media resources in their workplace environment:

> *I think [social media] now also supports structured education in the emergency department as well as just-in-time practice...* [Participant F1]

Other participants who harnessed social media as an academic resource utilized it during their professional identity as a medical student. In one participant's experience, social media initially aided in further bolstering engagement with their academic lecture and conference material:

> *Honestly, I started as a way to pay attention during lectures and conferences while staying engaged. It was a tool to essentially take notes. And do something that would kind of keep my ADD at bay while actually still engaging in the talks.* [Participant M1]

As Participant M1 continued to immerse into social media, their use of social media transformed outside their initial intentions from an academic tool into an avenue to share and disseminate content:

> *...and then that became a good way, also I found that it was also a good way to kind of share things that I was learning.* [Participant M1]

Another participant expressed social media as a platform for engagement despite their early hierarchical status as a medical student. Moreover, social media satiated their desire to engage with the community of practice in the early professional development stage as a medical student:

> *I was just a medical student...I think initially I was just excited and overwhelmed by the number of resources that were available and the number of people who were willing to engage with me as a medical student. And so, I was probably mostly excited about content at that phase and using [social media] to consume content.* [Participant F2]

Role Responsibilities

Other participants identified their origin of digital identity being formed through various pressures related to role responsibilities directed by supervisors. In these instances, it was their supervisors or superiors who engaged with our participants that encouraged them to start their digital identity:

> *My program director had asked that I start writing weekly pearls, the [Residency Program] pearls are still generated but that got me in the habit of writing something in an electronic format weekly.* [Participant M2]

Another participant experienced unconventional pressures from their supervisors through competition and boisterous wagers that ignited their digital identity's birth. In such cases, the result of long-term social media activity still prevails:

> *I made a bet with the VP who was in charge of operations and public relations.I said that in three months that I could have a bigger following than the hospital system and as a physician. And I think it took me five months. But I definitely did get more and still have more followers than the hospital system.* [Participant F3]

Cultivating Digital Identity

Socialization Into the Web-Based Community of Practice

As our participants engage further into the digital realm, the norms of the digital space are learned. By learning these rules, the digital identity of the participants begin to solidify. Participants are also able to increase their precision in navigating the digital space, allowing them to recognize which content garners them a stronger following. Resulting from this are networking opportunities and further engagement in the community of practice, as mentioned:

> *I made an account and very soon after six months I went to the #1 [spot]...I created an infographic because I just wanted to see if I could make one. And I put it out there and it got...300, 400, then 500 likes [and] retweets...So, it's like people started seeing my name. And then I got asked to do some training and talks: like hey you are good at this...And it really just snowballed.* [Participant F4]

Growth Via Parallel Development

The development of digital identity was noted by Participant F2, who explained that, in their experience, digital identity developed parallel to their professional identity development in real life. Namely, one participant's (F2) identity from being a medical student to being a resident and her progressive development in expertise was similarly reflected in her digital identity, as she explains:

> *...So, I think my position in the community has changed as I have developed more expertise and figured out really who I am as an academic and as a member of this community.* [Participant F2]

The Characteristics of Digital Identity: Fluid and Dynamic

Many participants expressed the different roles digital identity manifested, including "doing more critical appraisals and reviews" to "educating followers on areas of expertise" or "sharing [your] own work." Interestingly, despite the many roles that exist in the academic digital space, they are not singular nor mutually exclusive, as participants noted their "role as having multiple components." The evolution of digital identity was identified by participants to be fluid, transitive, and dynamic, where roles could be added over time or compounded:

> *Yeah, so I think there are a few different roles, so one is as a producer of content for social media and that is in our podcast... I also [identify] myself as an occasional contributor to other social media sites. And then I see myself as a consumer and sometimes curator.* [Participant F1]

The Professional and Personal Dimensions of Digital Identity

Desire to Be Professional

When we asked our participants about their contexts of social media use, many responded "primarily using it professionally" to the public. Other participants preferred a defined "dichotomous approach," separating professional use from personal use on social media utilizing strategies such as "using different platforms" for different identities. The observed management of identity separation on social media demonstrates an implicit expectation to confine to a purely professional image on the web.

Challenges to Disentangle Personal and Professional

Although some participants expressed the importance of maintaining a professional public frame on social media at all times, other interviewed experts expressed challenges in doing so. When browsing through social media content on platforms such as Twitter, one comes into contact with various types of content to which "you can't help but come across things that are of interest and not strictly professional." Blurring the lines between the personal and professional components of digital identity also extended to colleague relationships as elucidated by one of our male participants (M3):

> *I definitely started with very strict bounds where my only content that I really used personally was Facebook and everything else was professional. But that is essentially impossible to keep up because as you spend more time professionally [on social media], you become really good friends with a number of your colleagues. And therefore, they become part of your personal circle.* [Participant M3]

Convergence Is Inevitable

The active divide in splitting professional and personal components of digital identity is ineffective. Even common strategies such as explicitly stating the distinction of accountability in the profile biography between the professional and personal identity are viewed by our participants to be insufficient. This strategy of explicitly stating this distinction is intended to safeguard organizations from personal stances and behaviors that may not align with the organization's values and professionalisms; however, one of our male participants criticized as follows:

> *People like to put disclaimers in their Twitter bio saying that they don't represent their employers. I think that [it] is kind of nonsense. It is not going to help you legally if you get into trouble with your employer. It is not a get out of jail free card. So, I do mention where I work. And so far, that hasn't caused any problem with my employer.* [Participant M4]

Benefits of Convergence

Our interviewed participants have given diverse responses when managing personal and professional identities on social media; however, participants like M4 have experienced no consequences integrating both. Furthermore, researchers (such as F5) bolster their professional practice by harnessing their personal identity on the web:

> *So, with Twitter, I don't really have the separation because it is more about the sort of medical, networking side of things and a little bit less about my very personal experience of the world. Although there is a bit of that [personal experience] as well. And in my opinion, my experience is enhanced by being by presenting myself as a human and having a kind of complex life outside of academia or pure clinical roles.* [Participant F5]

Real Life and Social Media Alignment

Observation of the Alignment

When asking participants about their behavior and identity in the digital space with respect to real life, most participants expressed an alignment between these identities. One participant engaged with other scientists on the web as if they were their colleagues in real life:

> *I just have normal conversations as I would if I was having a conversation at a conference or some other professional context with people.* [Participant M5]

As expected, such an alignment further extended to academic interests and roles as well, demonstrating social media as an extension to developing the identity of these researchers. In one participant's case (M6), their interest in medications and advocacy for safe and established treatments directed to the scientific community reflected their real-life academic identity:

> *One of my main interests is medications, and so I would like to encourage what I perceive to be safer drug therapy. I also like to give followers insight into clinical cases or also sometimes even personal stuff.* [Participant M6]

Importance of Holistic Representation

Participants identified an inherent importance of centralizing their digital identity to their true identity in real life. One participant highlights a state of comfort when representing their full digital identity by balancing their professional and personal identity on the web as they refuse "artificially keeping [professional and personal identity] separate." Another participant supported this view in avoiding artificial separation of personal and professional identities when engaging with other educators on the web*:*

> *I don't generally have interactions that I wouldn't have on professional contacts on my Twitter channel, which as I noted I use for personal and professional purposes.* [Participant M5]

Consequence of Misalignment

When immersing further into the community of practice, opportunities for misalignment between the physical and digital identity may occur in which the physical identity lags:

> *It has taken a while for my [online identity] to also become reflective in my public identity"* [Participant F6]

In another participant's case, the rift was formed because of a noted, enhanced capability of networking on social media:

> *I feel like I occasionally get some type of imposter syndrome by this group of people online and that is the network of people that I have met online and have worked with on a number of projects. I think they see me as really valuable and this person that they want to work with and really value what I bring to the table. I don't always feel that same way at my home institution that I work at.* [Participant F2]

In this instance, where digital identity outpaces the physical professional identity, the misalignment of these identities caused this participant to compromise or restructure one of her identities to fit the latter:

> *And so, then I wonder: is this a bit of a disconnect between who I am in real life and who I am online? Or is it actually that there are people in other organizations that actually value what I bring to the table a little bit more than where I am at.* [Participant F2]

Discussion

Through the collected experiences from our interviewed social media experts, our team has investigated the surrounding interactions and explored the construct of a digital identity. For our experts, the initial formation of digital identity materialized through needs present in the early stages of professional identity in medical students. Proceeding initial growth, the development of digital identity molded in conjunction with real-life professional identity development and expertise. The importance of representing oneself holistically on the web between the physical and digital identities and between the personal and professional identities was highlighted.

Our main finding is that the fluid and dynamic nature of digital identity are similar to those of professional identity's fluidity [30-32]. Participants often transformed, even interchangeably, through different roles from education bloggers, critical appraisers, or even extending to personal characters. Interestingly, the transitive property of easily interchanging roles in our findings of digital identity contrasts the disruption faced when frequently transitioning between different professional identities [33]. Our analysis also captured the behavior of models by Lave and Wenger [34] on situated learning and community of practice. Engagement at the initial identity formation stage starts peripheral as our participants begin developing their content [34]. Through the feedback process of socialization of continually deeper engagement, the identity becomes more refined until becoming a member of the community of practice [34]. In the case of our findings on digital identity, web-based metrics such as followers or increased web traffic were successful markers of socialization into the digital community and guided our participants to refine their digital identity by adhering to engagement strategies [35] and content that increased followings.

Although similarities exist between the characteristics of digital identity and professional identity development, differences are highlighted through the pace of development via the digital space. Social media has overcome physical barriers and allows health professionals to have a global reach to engage with national committees and remote stakeholders [12,21,36-39]. This allowed our interviewed participants who faced physical barriers such as practicing rurally to otherwise come across opportunities that normally would not exist. Occasionally, identity mismatch would occur as professional identity lags behind digital identity development.

To resolve the tension between identity mismatch, it was found to be important to capture a holistic identity and represent it digitally. This finding contrasted with identity management techniques observed by Cho and Jimerson [39], where professional identity was publicly expressed whereas personal identity was compartmentalized in separate social media accounts. Their study additionally found that professional identity was further fragmented between different organizational social media accounts as an engagement strategy to tailor to different audiences [40]. These differences between our findings may highlight deviations in identity management strategies between organization and individually owned social media accounts. Although the participants in the study by Cho and Jimerson [39] responded to organizational and administrative needs, our participants utilized social media to collaborate with the community of practice to further advance their scholarly work. These interactions were preferred to be organic, utilizing a natural blend between personal and professional identities. Such a preference for convergence aligns with findings by DeCamp et al [22], who argue that physicians are not intended to avoid personal interactions, as it is sometimes unavoidable. Decamp et al [22] further affirm that appropriate personal engagement on the web can bolster web-based professional interactions, to which our participants agree.

Our findings showed that digital identity formation initially begins because of a professional or academic need that can take the form of a medical student seeking to understand a specialty [7,21], or a new leader looking to connect more with stakeholders. Many of our interviewed participants explained that their digital identity began while they were concurrently developing their professional identity as trainees. This finding is in line with previous literature; a qualitative study interviewing medical students demonstrated their desire to be given a voice and engage with the online community of practice [12]. Other studies extend this narrative, showing that engagement in social media enables networking, mentorship, and content learning [36,37,39,41-44]. As Mather et al [45] highlight the benefits social media provide call the need for educational institutions to engage and facilitate the development of professional identity for learners within the digital space.

The impact of social media as a disseminator of health-related information and misinformation is one that has been undoubtedly demonstrated by its utilization during the COVID-19 pandemic [46-48]. Further exploration of digital identity development will allow educational institutions and organizations to cultivate physicians that are increasingly familiar with the digital space. By equipping these physicians with the tools to build web-based stakeholdership and develop their digital role as digital responders and disseminators of

information on the web, physicians will be more ready to combat future threats of web-based health misinformation [46,47].

The current digital space serves as an extension to physical space, and its ability to capture and express our personal and professional identities is, thus, also extended in the same manner [49]. Our findings provide insight into the model of digital identity development and its interactions with existing literature regarding professional identity development. We present digital identity as an additional dimension to professional identity development within health care professionals. With social media's dominant role as an accelerated medium for dissemination and identity development, further investigation of the interactions between digital identity formation upon professional identity development in individuals of various key stages including medical students, residents, and junior academic cohorts are warranted. These findings will be important to guide medical trainees and students through their professional identity formation in both the real and digital worlds.

Limitations

This study had a number of limitations. The research lead investigator had expertise in web-based pedagogy and knowledge translation. Preventative measures were taken to reduce any skewed interpretations of participant answers and words. For example, the research assistant and transcriptionist interviewed all participants and transcribed all transcripts, respectively. In addition, all transcripts were deidentified before the qualitative analysis, preventing the lead investigator from deriving the identity of the participants. During the analysis phase, a variety of strategies were used with the knowledge that the lead investigator of the research team was an expert in social media knowledge translation with respect to the other members.

We focused on those who had established reputations and track records for their academic work and, therefore, likely skewed our interviews toward those who are further along in their web-based professional identity formation. Of note, we only had 1 senior trainee participant, and therefore, there may be some limitations in extrapolating our key findings to more junior learners. However, many of our participants had good recollections about their experiences with digital identity formation and origin during their training, which may allow for transferable findings to be relevant to educators seeking to counsel junior learners.

Conclusions

Social media introduces new features to professional identity in the digital space. The formation of digital identity, its development, and interactions that require identity management were features captured in our study. Moreover, the fluid and dynamic characteristics of digital identity in conjunction with its accelerated capacity of growth yield differences from professional identity development that can potentially be harnessed. Navigating the identity development of young or upcoming health care professionals is a priority for institutions now and in the post–COVID-19 world. Today, digital identity can no longer be neglected. Digital citizenship can no longer be ignored as a key facet of one's professional responsibility. If we are to effectively train the next generation of health care professionals in an era of ongoing technological development, digital identity development must be explored and supported.

Acknowledgments

The authors wish to thank Priya Thomas, Emma Bridgwater, and Aisha Mohamed for their assistance in the early phases of this study. The authors also thank Elizabeth Clow for her services as a transcriptionist for their project. TMC reports receiving funding from the Physician Services Incorporated (PSI) Foundation for this work via the 2018 PSI Foundation Graham Farharquason Knowledge Translation grant recipient. YY is the recipient of the Scientific and Technological Research Council of Turkey Postdoctoral Fellowship grant.

Conflicts of Interest

TMC reports affiliations with McMaster University and Academic Life in Emergency Medicine and receives stipends from these organizations. She is the co-founder of the CanadiEM website and declares an intellectual conflict of interest but receives no funding from this organization.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 59 KB - jmir_v22i9e21416_app1.pdf]

References

1. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. Reframing medical education to support professional identity formation. Acad Med 2014 Nov;89(11):1446-1451 [FREE Full text] [doi: 10.1097/ACM.0000000000000427] [Medline: 25054423]
2. Wilson I, Cowin LS, Johnson M, Young H. Professional identity in medical students: pedagogical challenges to medical education. Teach Learn Med 2013;25(4):369-373. [doi: 10.1080/10401334.2013.827968] [Medline: 24112208]
3. Wald HS, Anthony D, Hutchinson TA, Liben S, Smilovitch M, Donato AA. Professional identity formation in medical education for humanistic, resilient physicians: pedagogic strategies for bridging theory to practice. Acad Med 2015 Jun;90(6):753-760. [doi: 10.1097/ACM.0000000000000725] [Medline: 25901874]

4. Monrouxe L. Identity, identification and medical education: why should we care? Med Educ 2010 Jan;44(1):40-49. [doi: 10.1111/j.1365-2923.2009.03440.x] [Medline: 20078755]
5. Cruess SR, Cruess RL, Steinert Y. Supporting the development of a professional identity: general principles. Med Teach 2019 Jun;41(6):641-649. [doi: 10.1080/0142159X.2018.1536260] [Medline: 30739517]
6. Costello C. Professional Identity Crisis: Race, Class, Gender, and Success at Professional Schools. Nashville, TN: Vanderbilt University Press; 2006.
7. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. Acad Med 2015 Jun;90(6):718-725. [doi: 10.1097/ACM.0000000000000700] [Medline: 25785682]
8. Wald HS. Professional identity (trans)formation in medical education: reflection, relationship, resilience. Acad Med 2015 Jun;90(6):701-706. [doi: 10.1097/ACM.0000000000000731] [Medline: 25881651]
9. Paton C, Bamidis PD, Eysenbach G, Hansen M, Cabrer M. Experience in the use of social media in medical and health education. Contribution of the IMIA social media working group. Yearb Med Inform 2011;6:21-29. [Medline: 21938320]
10. Davis WM, Ho K, Last J. Advancing social media in medical education. Can Med Assoc J 2015 May 19;187(8):549-550 [FREE Full text] [doi: 10.1503/cmaj.141417] [Medline: 25852033]
11. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. Acad Med 2013 Jun;88(6):893-901. [doi: 10.1097/ACM.0b013e31828ffc23] [Medline: 23619071]
12. Chretien KC, Tuck MG, Simon M, Singh LO, Kind T. A digital ethnography of medical students who use twitter for professional development. J Gen Intern Med 2015 Nov;30(11):1673-1680 [FREE Full text] [doi: 10.1007/s11606-015-3345-z] [Medline: 25952652]
13. Avcı K, Çelikden SG, Eren S, Aydenizöz D. Assessment of medical students' attitudes on social media use in medicine: a cross-sectional study. BMC Med Educ 2015 Feb 15;15:18 [FREE Full text] [doi: 10.1186/s12909-015-0300-y] [Medline: 25890252]
14. Cleary M, Ferguson C, Jackson D, Watson R. Editorial: social media and the new e-professionalism. Contemp Nurse 2013 Oct;45(2):152-154. [Medline: 24422224]
15. Kind T, Patel PD, Lie D, Chretien KC. Twelve tips for using social media as a medical educator. Med Teach 2014 Apr;36(4):284-290. [doi: 10.3109/0142159X.2013.852167] [Medline: 24261897]
16. Mansfield SJ, Morrison SG, Stephens HO, Bonning MA, Wang S, Withers AH, et al. Social media and the medical profession. Med J Aust 2011 Jun 20;194(12):642-644. [Medline: 21692723]
17. Chretien KC, Kind T. Social media and clinical care: ethical, professional, and social implications. Circulation 2013 Apr 2;127(13):1413-1421. [doi: 10.1161/CIRCULATIONAHA.112.128017] [Medline: 23547180]
18. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
19. Fenwick T. Social media and medical professionalism: rethinking the debate and the way forward. Acad Med 2014 Oct;89(10):1331-1334 [FREE Full text] [doi: 10.1097/ACM.0000000000000436] [Medline: 25076200]
20. Greysen SR, Chretien KC, Kind T, Young A, Gross CP. Physician violations of online professionalism and disciplinary actions: a national survey of state medical boards. J Am Med Assoc 2012 Mar 21;307(11):1141-1142. [doi: 10.1001/jama.2012.330] [Medline: 22436951]
21. O'Regan A, Smithson WH, Spain E. Social media and professional identity: pitfalls and potential. Med Teach 2018 Feb;40(2):112-116. [doi: 10.1080/0142159X.2017.1396308] [Medline: 29172814]
22. DeCamp M, Koenig TW, Chisolm MS. Social media and physicians' online identity crisis. J Am Med Assoc 2013 Aug 14;310(6):581-582 [FREE Full text] [doi: 10.1001/jama.2013.8238] [Medline: 23942675]
23. Percy WH, Kostere K, Kostere S, Percy WH, Kostere K, Kostere R. Generic Qualitative Research in Psychology. - 2015:-.
24. Etikan I. Comparision of snowball sampling and sequential sampling technique. BBIJ 2016 Jan 2;3(1):-. [doi: 10.15406/bbij.2016.03.00055]
25. Riddell J, Brown A, Kovic I, Jauregui J. Who are the most influential emergency physicians on Twitter? West J Emerg Med 2017 Feb;18(2):281-287 [FREE Full text] [doi: 10.5811/westjem.2016.11.31299] [Medline: 28210365]
26. Video Conferencing, Web Conferencing, Webinars, Screen Sharing. Zoom Video Conferencing. URL: https://zoom.us/ [accessed 2020-06-04]
27. Fauci AS, Lane HC, Redfield RR. Covid-19 - navigating the uncharted. N Engl J Med 2020 Mar 26;382(13):1268-1269 [FREE Full text] [doi: 10.1056/NEJMe2002387] [Medline: 32109011]
28. Paules CI, Marston HD, Fauci AS. Coronavirus infections-more than just the common cold. J Am Med Assoc 2020 Jan 23:- epub ahead of print. [doi: 10.1001/jama.2020.0757] [Medline: 31971553]
29. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
30. Holden MD, Buck E, Luk J, Ambriz F, Boisaubin EV, Clark MA, et al. Professional identity formation: creating a longitudinal framework through TIME (transformation in medical education). Acad Med 2015 Jun;90(6):761-767. [doi: 10.1097/ACM.0000000000000719] [Medline: 25853688]

31. Bebeau MJ. Evidence-based character development. In: Kenny N, Shelton W, editors. Lost Virtue. Advances in Bioethics. Volume 10. Melbourne, Australia: Emerald Group Publishing Limited; 2006:47-86.
32. Stets J, Serpe R. Identity theory. In: DeLamater J, Ward A, editors. Handbook of Social Psychology. Netherlands: Springer; 2013:31-60.
33. Gordon L, Rees CE, Jindal-Snape D. Doctors' identity transitions: choosing to occupy a state of 'betwixt and between'. Med Educ 2020 May 13:- epub ahead of print. [doi: 10.1111/medu.14219] [Medline: 32402133]
34. Illeris K. Contemporary Theories of Learning. New York, USA: Routledge; 2009.
35. Kohli MD, Daye D, Towbin AJ, Kotsenas AL, Heilbrun ME. Social media tools for department and practice communication and branding in the digital age. Radiographics 2018 Oct;38(6):1773-1785. [doi: 10.1148/rg.2018180090] [Medline: 30303796]
36. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
37. Chan WS, Leung AY. Use of social network sites for communication among health professionals: systematic review. J Med Internet Res 2018 Mar 28;20(3):e117 [FREE Full text] [doi: 10.2196/jmir.8382] [Medline: 29592845]
38. Peters ME, Uible E, Chisolm MS. A Twitter education: why psychiatrists should tweet. Curr Psychiatry Rep 2015 Dec;17(12):94. [doi: 10.1007/s11920-015-0635-4] [Medline: 26463050]
39. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
40. Cho V, Jimerson JB. Managing digital identity on Twitter. Educ Manag Adm Leadersh 2016 Sep 22;45(5):884-900. [doi: 10.1177/1741143216659295]
41. Hinojo-Lucena F, Aznar-Díaz I, Cáceres-Reche M, Romero-Rodríguez J. Use of social networks for international collaboration among medical students. Educación Médica 2020 Mar;21(2):137-141. [doi: 10.1016/j.edumed.2018.08.009]
42. Popoiu MC, Grosseck G, Holotescu C. What do we know about the use of social media in medical education? Procedia Soc Behav Sci 2012;46:2262-2266. [doi: 10.1016/j.sbspro.2012.05.466]
43. Hazzam J, Lahrech A. Health care professionals' social media behavior and the underlying factors of social media adoption and use: quantitative study. J Med Internet Res 2018 Nov 7;20(11):e12035 [FREE Full text] [doi: 10.2196/12035] [Medline: 30404773]
44. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
45. Mather C, Cummings E, Nichols L. Social Media Training for Professional Identity Development in Undergraduate Nurses. Stud Health Technol Inform 2016;225:344-348 [FREE Full text] [Medline: 27332219]
46. Merchant RM. Evaluating the potential role of social media in preventive health care. J Am Med Assoc 2020 Jan 10:- epub ahead of print. [doi: 10.1001/jama.2019.21084] [Medline: 31922532]
47. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. J Am Med Assoc 2020 Mar 23:- epub ahead of print. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
48. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. J Am Med Assoc 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
49. Lupton D. Digital Sociology. New York, USA: Routledge; 2015.

Abbreviations

PSI: Physician Services Incorporated

Edited by M Focsa; submitted 15.06.20; peer-reviewed by M Peters, C Mather; comments to author 07.08.20; revised version received 11.08.20; accepted 14.08.20; published 29.09.20.

Please cite as:
Ruan B, Yilmaz Y, Lu D, Lee M, Chan TM
Defining the Digital Self: A Qualitative Study to Explore the Digital Component of Professional Identity in the Health Professions
J Med Internet Res 2020;22(9):e21416
URL: http://www.jmir.org/2020/9/e21416/
doi:10.2196/21416
PMID:32990636

Original Paper

Development of a Social Network for People Without a Diagnosis (RarePairs): Evaluation Study

Lara Kühnle[1*], BSc; Urs Mücke[1*], Dr med; Werner M Lechner[2*], DProf, DRING; Frank Klawonn[3,4*], DProf; Lorenz Grigull[5*], DProf, Dr med

[1]Hannover Medical School, Hannover, Germany
[2]KIMedi GmbH, Donauwörth, Germany
[3]Helmholtz Centre for Infection Research, Braunschweig, Germany
[4]Ostfalia University, Wolfenbüttel, Germany
[5]University Hospital Bonn (UKB), Bonn, Germany
[*]all authors contributed equally

Corresponding Author:
Urs Mücke, Dr med
Hannover Medical School
Carl-Neuberg-Straße 1
Hannover, 30625
Germany
Phone: 49 511532 ext 3220
Email: muecke.urs@mh-hannover.de

Abstract

Background: Diagnostic delay in rare disease (RD) is common, occasionally lasting up to more than 20 years. In attempting to reduce it, diagnostic support tools have been studied extensively. However, social platforms have not yet been used for systematic diagnostic support. This paper illustrates the development and prototypic application of a social network using scientifically developed questions to match individuals without a diagnosis.

Objective: The study aimed to outline, create, and evaluate a prototype tool (a social network platform named RarePairs), helping patients with undiagnosed RDs to find individuals with similar symptoms. The prototype includes a matching algorithm, bringing together individuals with similar disease burden in the lead-up to diagnosis.

Methods: We divided our project into 4 phases. In phase 1, we used known data and findings in the literature to understand and specify the context of use. In phase 2, we specified the user requirements. In phase 3, we designed a prototype based on the results of phases 1 and 2, as well as incorporating a state-of-the-art questionnaire with 53 items for recognizing an RD. Lastly, we evaluated this prototype with a data set of 973 questionnaires from individuals suffering from different RDs using 24 distance calculating methods.

Results: Based on a step-by-step construction process, the digital patient platform prototype, RarePairs, was developed. In order to match individuals with similar experiences, it uses answer patterns generated by a specifically designed questionnaire (Q53). A total of 973 questionnaires answered by patients with RDs were used to construct and test an artificial intelligence (AI) algorithm like the k-nearest neighbor search. With this, we found matches for every single one of the 973 records. The cross-validation of those matches showed that the algorithm outperforms random matching significantly. Statistically, for every data set the algorithm found at least one other record (match) with the same diagnosis.

Conclusions: Diagnostic delay is torturous for patients without a diagnosis. Shortening the delay is important for both doctors and patients. Diagnostic support using AI can be promoted differently. The prototype of the social media platform RarePairs might be a low-threshold patient platform, and proved suitable to match and connect different individuals with comparable symptoms. This exchange promoted through RarePairs might be used to speed up the diagnostic process. Further studies include its evaluation in a prospective setting and implementation of RarePairs as a mobile phone app.

(J Med Internet Res 2020;22(9):e21849) doi:10.2196/21849

KEYWORDS

rare disease; diagnostic support tool; prototype; social network; machine learning; artificial intelligence

Introduction

A patient without a diagnosis desperately struggles for help. This holds especially true for those with an undiagnosed rare disease (RD). Although an RD is one that, by definition, only 5 out of 10,000 people suffer from, in total there are approximately 13.5 million people with an RD in the European Union (EU) [1] and approximately 400 million worldwide [2]. Affected patients search for the diagnosis for an average of 8 years. During this time, misdiagnosis and wrong treatments are common, and social isolation and financial damage occur frequently [3-9]. By contrast, patients with a diagnosed RD are highly active in supporting each other, and may serve as experts for their diseases in patient groups. This is an important resource for information and guiding besides the information on RD in the internet.

The internet has grown to be an easily accessible hub for research, even for health care information. Today, almost everybody is *Googling* symptoms before, while, and after a health care visit [10]. The power of internet-based diagnosis was recently underscored by Siempos et al [11], highlighting that 22.1% of correct diagnoses from laymen were due to web searches. Furthermore, doctors themselves similarly consult the internet searching for the correct diagnosis [12]. Here, in 58% of the cases, search engines such as Google helped identify the diagnosis [12].

Besides searching the internet, almost all young Americans aged between 18 and 29 use social media [13]. Communicating via those networks is a daily activity for them, and using social media platforms has become an established way of making personal connections [14]. Online social networks are not tied to a specific time or place, making them even more efficient for communication. There are online social networks that help preserve contacts over distances, as well as networks facilitating meeting people with common interests or issues. Those networks often use matching algorithms containing artificial intelligence (AI) to match for optimum results. Moreover, such matching algorithms are used in marketing to find products and services that fit the needs of a person.

In *RarePairs* we also use a matching algorithm to meet the needs of potential users. This prototype of a new social platform is designed to bring people with and without a diagnosis together, making interaction and supporting to find the right diagnosis possible. Thus, it tries to help find the right people to discuss possible diagnoses, coping strategies, and treatments for the user's symptoms. We decided to focus on the group of RDs as they are still overlooked in cases of diagnosis, care, and treatment. The idea behind RarePairs is to combine already existing resources (social networks, the internet, smart mathematical algorithms, an existing questionnaire/data set), use cases (finding diagnoses, health information, contact to other people with the same condition), and challenges (diagnostic odyssey, a very small global proportion of people with the same condition), and fit them into one tool. The aim of this study was to outline, create, and evaluate a prototype of RarePairs.

Methods

To ensure the quality of the design process for an online social network, we used an ISO norm. This ISO norm is designed to develop a user-centered design software product.

To build our prototype we used ISO 9241-210:2010 and followed the suggested 4 steps: (1) Understand and specify the context of use; (2) Specify the user requirements; (3) Produce design solutions to meet these requirements; and (4) Evaluate the designs against requirements.

To complete steps 1 and 2, we collected known facts based on different materials, such as the German website for information on RDs [15]. Additionally, we used expert knowledge about people with RD which was collected from previous research and discussions with patient groups. Details on the completed steps can be found in Multimedia Appendix 1.

In the second step, this information was discussed with an interdisciplinary team of doctors, computer scientists, and mathematicians to define the context of use and user requirements. To complete step 3, we used commonly known hardware and software (all-day-use laptops, *Adobe Photoshop*, Text editors, *MAMP*, and *GitHub*) for the web design of our prototype. We also used common coding formats, such as HTML, CSS, PHP, MySQL, and R, for designing the prototype. No templates or content management systems were used.

For the most important part of the prototype, the matching algorithm, we resorted to a questionnaire named Q53 which was built during previous research in the working group [16]. Briefly, this questionnaire was built using patients' experience. Individuals with different RDs were interviewed to gain insight into their prediagnostic experiences. These experiences (in daily life) were qualitatively analyzed. In a 7-step process following strict rules we finally ended up with a set of 53 questions. Afterward, larger cohorts of individuals with different RDs (and established diagnoses) were contacted and invited to answer the questions. This approach was based on the idea that most people with different RDs not only experience similar basic symptoms (eg, fatigue, blaming) during their prediagnostic *odyssey*, but also have comparable strategies on finding a diagnosis (eg, consulting various doctors) or coping with daily life (eg, avoidance strategies, intuitive usage of assistive technologies). Based on these experiences from the period (sometimes years) prior to the diagnosis, which were collected through interviews from individuals with a proven RD, 53 questions were identified as being crucial and prototypic for individuals with different RDs. The 53 questions break down into 7 different categories (eg, symptoms, social environment, or looking for the cause), and can be answered with: (1) *No*, (2) *Slightly no*, (3) *Slightly yes*, (4) *Yes*, and (5) *Do not know*. Thus, the questionnaire Q53 not only asks for specific symptoms, but also reflects the challenges and obstacles of individuals with different RDs in daily life, typical circumstances, and certain actions, and the Q53 can be answered without expert knowledge within 15 minutes. Some examples out of the set of 53 questions are (1) Do you withhold information about your complaints from your environment (eg, family, friends, colleagues)? (2) Do you use supportive devices to positively help your daily

routine; (3) Did you suspect—if so since longer—that something is "wrong" with you? (4) Do you use tricks and dodges to master restrictions during your daily life? (5) Would you say that the ambiguity about the cause of your complaints/irritating phenomena was the worst? (see Multimedia Appendix 2 [German] or Multimedia Appendix 3 [English] for the whole Q53 questionnaire).

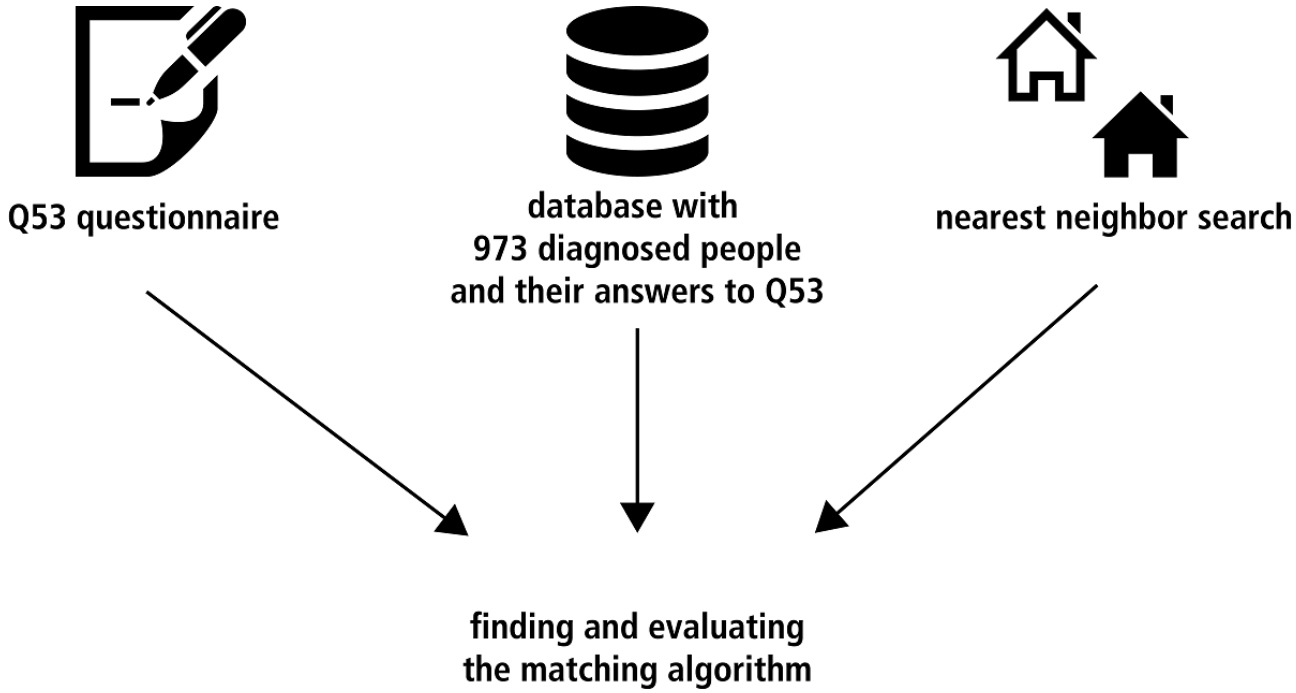
In the previous study [16] which described the development of the questionnaire, the set of questions was then used by a matching algorithm combining the results of support vector machine, random forest, logistic regression, and linear discriminant analysis, to effectively classify a data set of approximately 1000 questionnaires of individuals with different RDs and other disease conditions into 4 different diagnostic categories [16]. Briefly, these diagnostic categories differentiated between RDs, chronic diseases (CDs), psychosomatic disorders, and other disease conditions. Hence, the questionnaire Q53 was originally *not* designed for diagnosing a specific disease, but it proved effective to cluster diseases into diagnostic categories (which might also prove helpful for guiding individuals during a *diagnostic odyssey*). As the aim of RarePairs is bringing together undiagnosed/diagnosed people so as to share their symptoms and coping strategies, such a questionnaire might suit well for this purpose.

In the second step, an existing data set (from previous study [16]) of answered Q53 questionnaires (n=973) from individuals with non-RDs and different RDs, such as neuromuscular diseases (eg, Pompe disease, amyotrophic lateral sclerosis), autoimmune diseases (eg, sarcoidosis, systemic lupus erythematosus), or rare metabolic diseases (eg, glycogen storage diseases), was used for prototypic evaluation of RarePairs.

Of the 973 people, 759 were previously and certainly diagnosed with an RD, 27 were healthy, 34 had an unknown diagnosis, 27 were diagnosed with psychosomatic disorders, and 126 had a CD. They were between 0 and 87 years of age with a mean age of 39.4 years (702 were female and 271 were male). Recruitment was limited to Germany.

This basic process is illustrated in Figure 1. We used the k-nearest neighbor search and different distance calculating methods to find and evaluate matchings for a given set of 973 users with different diagnoses and different diagnostic categories. The k-nearest neighbor search is an easy, but effective, way to find similarities. Different calculating methods from different mathematical groups based on a publication by Cha [17] were compared.

Figure 1. Used material for finding and evaluating the matching algorithm.



In step 4, the prototype with the matching algorithm was evaluated based on leave-one-out cross-validation, which means that 1 data set was *left out* and the algorithm searched for fitting matches in the remaining 972 data sets. The matches for every single data set out of the total 973 answered Q53 questionnaires were analyzed regarding age, gender, latency (ie, *time with* symptoms but no diagnosis), disease group (category of the diagnosis referring to the affected organ or pathophysiology [eg, neuromuscular disease, metabolic disease]), diagnostic system (higher-order category that defines the type of the diagnosis [eg, RD or CD] but does not consider the affected organ [eg, RD of the liver]), and exact diagnosis (ie, exact name of the diagnosis). We considered matching for the same diagnosis in the category *diagnosis* most important, followed by the same *diagnostic system*, *disease group*, and *age*. The same *latency* and *gender* were considered less important for matching. Here, we followed the hypothesis that matching partners benefit most from sharing the same diagnoses or diagnostic categories. Such a matching might enable a helpful dialogue between the matching partners (eg, common experiences, doctors, therapies). By contrast, the same *gender* would not be as helpful. This analysis was performed for every

calculating method, as well as a random sampling. Likewise, a comparison of random matching and similarity-based matching was possible.

Results

Overview of RarePairs User Path

The users' path through RarePairs is illustrated in Figure 2. After the login procedure, the user updates a profile and answers the 53 questions essential for matching (see Figure 2). The landing page of RarePairs is shown in Figure 3 (Further screenshots of the prototype can be found in Multimedia Appendix 4).

Figure 2. Simplified scheme of the clicking path for new and already registered users of RarePairs.

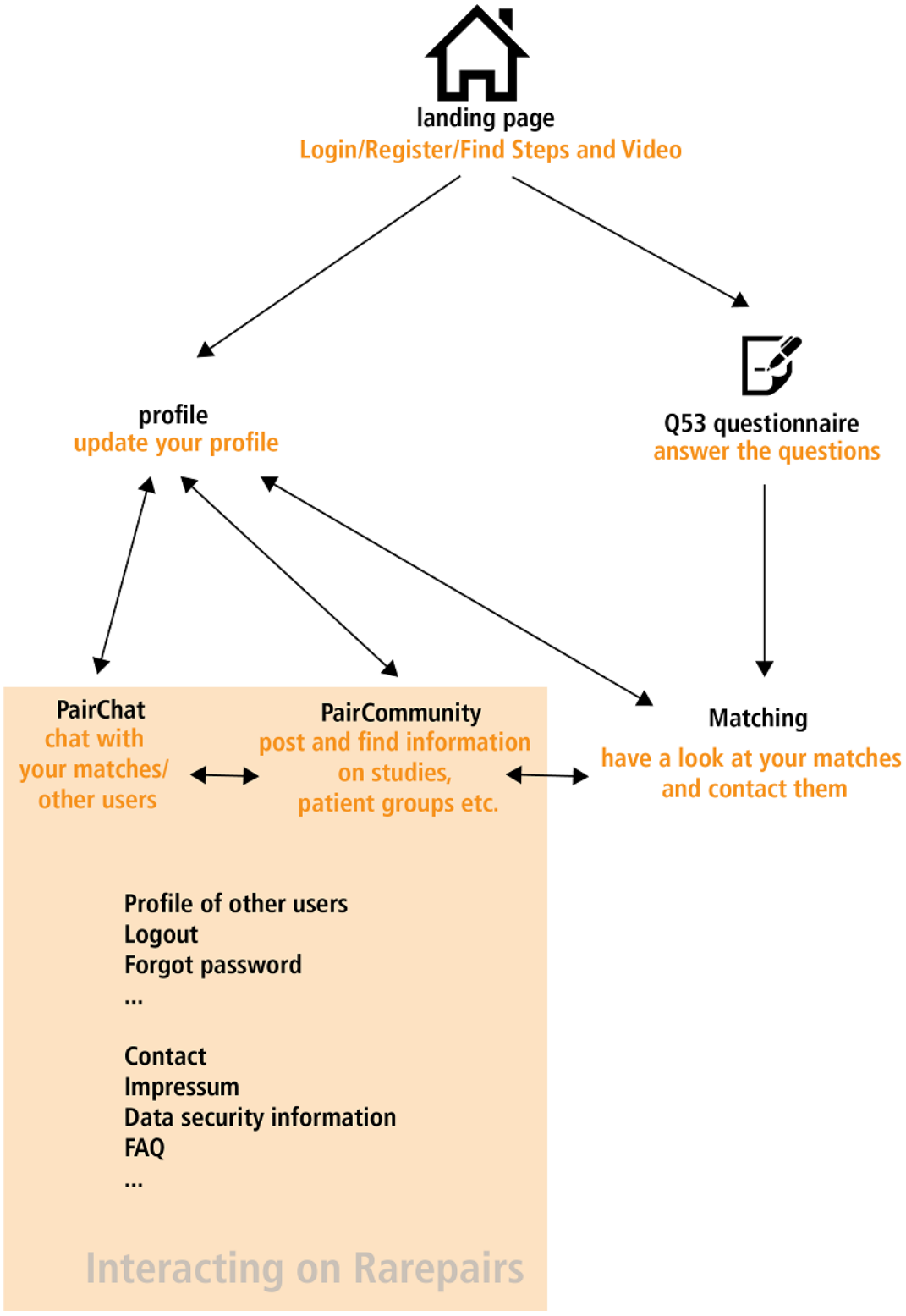
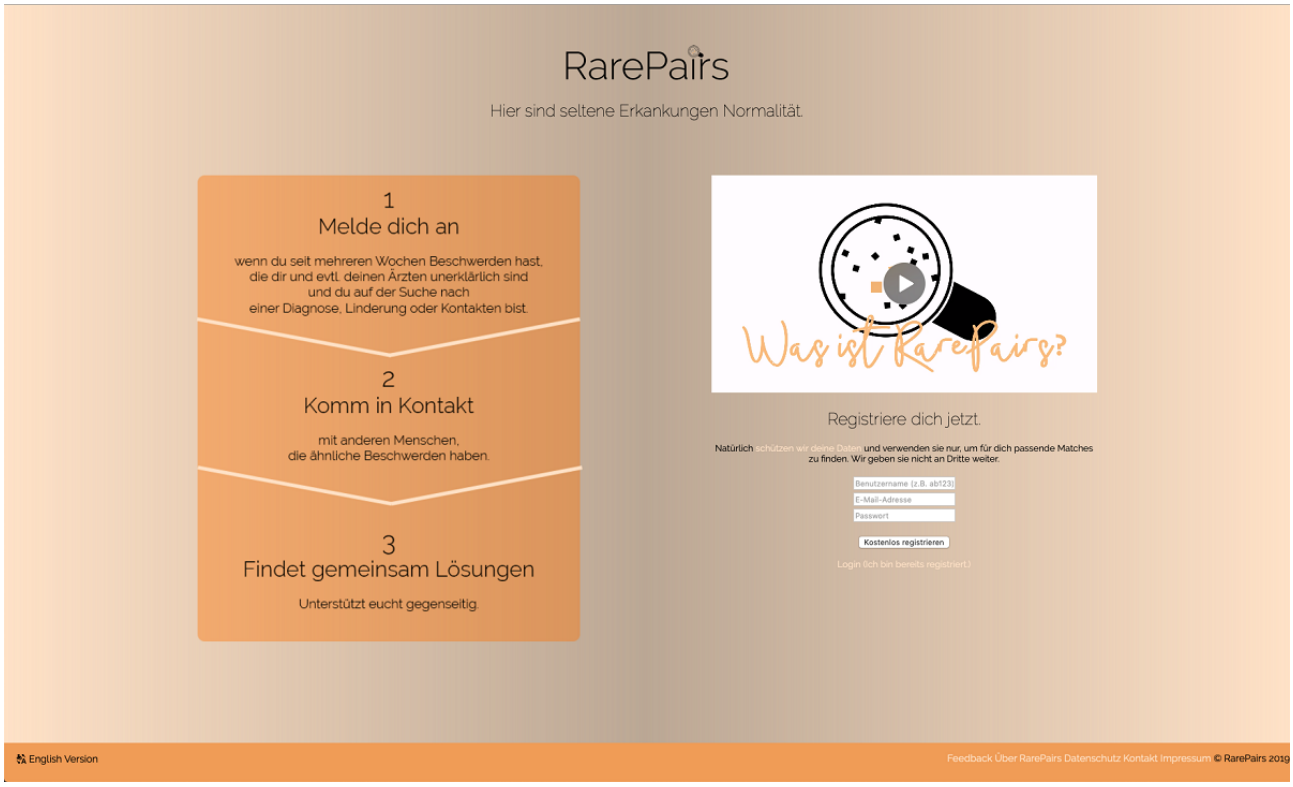


Figure 3. Landing page of RarePairs where users can get information, register, or log in. Users find information by text and a short video addressing aims and scope of RarePairs. Currently, the landing page is in German, an English version is under construction.



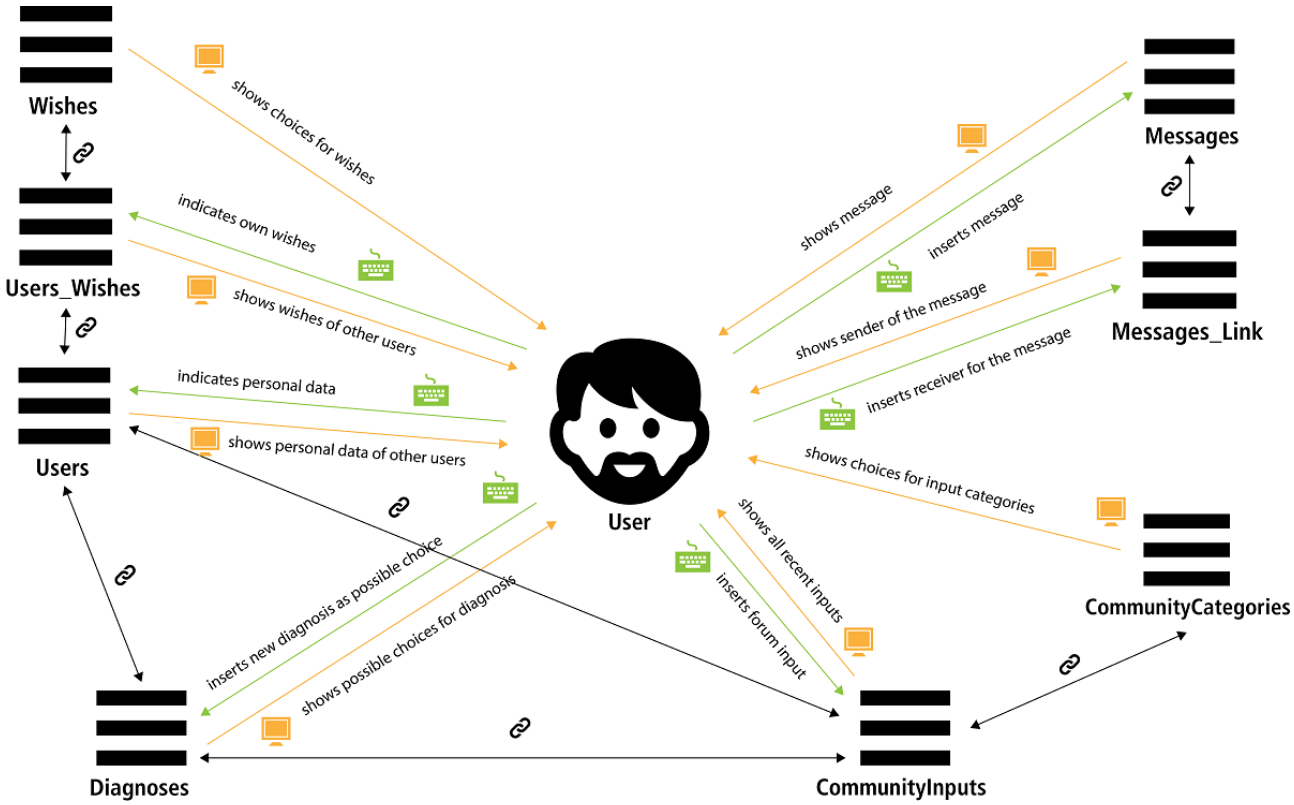
Login/Register

Users register, or, if they already have an account, login into RarePairs. The usual basic security arrangements such as checking email format, checking password, transferring data as POST variables, not allowing multiple accounts with the same email address are made. Moreover, the users can request their password via email, if forgotten. The user data are stored in a MySQL database table titled *users*.

Q53 Questionnaire

New registering users are initially led to the first Q53 questionnaire sheet. The questionnaire is divided into 9 website sheets. The first explains the aims and scope of the Q53 questionnaire. Additionally, users have the opportunity to give supplementary information (eg, hobbies or the aims/wishes of the user) for the account profile. The in-between sheets (numbers 2-8) show the 53 questions and allow answering through PHP forms. Here, the data are transferred via POST variables and stored into the MySQL database table *users* (users aims/wishes are stored in the database 'Wishes' [table 'Users_Wishes']; see Figure 4 for all database tables and possible interactions). It is possible to change the personal profile data later. In the current prototypic version of RarePairs, the answers to the 53 questions of the Q53 questionnaire are static (ie, they cannot be changed later).

Figure 4. Visualization of the tables and possible interactions between them and the user.



Matching

The answer pattern of the Q53 questionnaire forms the basis for matching different users of RarePairs. In this prototype, the nearest neighbor was used to calculate matching users/similar answer patterns in the Q53 questionnaire.

Therefore, we calculated the differences of answers stored in the MySQL database table *users*. For calculating, many different distance (respectively similarity) calculating methods were compared such as Manhattan ($d= \Sigma^{d}_{i=1}|P_i - Q_i)|$) or cosine ($d=[\Sigma^{d}_{i=1}Pi \cdot Qi]/[\sqrt{\Sigma^{d}_{i=1}p^2_i} \cdot \sqrt{\Sigma^{d}_{i=1}q^2_i}]$) from 8 different mathematical groups based on the publication by Cha [17]. As values, we used the numerical equivalents of given answers as explained above: (1) No, (2) Slightly no, (3) Slightly yes, (4) Yes, and (5) Do not know.

Finding the Right Distance Calculating Method

From an AI perspective, it is not a priori clear as to which calculating method works best. Therefore, we compared matching with 24 different methods (see Multimedia Appendix 5 for the exact mathematics). We used the existing database of 973 data sets (containing personal information such as age and gender and all answers to the Q53 questionnaire), and identified 10 matches for every data set (using the leave-one-out method; see Figures 5 and 6 for visualization). We chose k=10 because we assumed it to be a good compromise between proper selection size for the searching user and still not be overwhelming. After finding matches with one calculating method, we repeated the process with another method. Second, we evaluated the matching by comparing the properties of the data set and its 10 matches, and the quality of the matching (eg, accordance of diagnosis = how many of the 10 matches have the same diagnosis as the user under evaluation). The average of the accordance for every calculating method is shown in Table 1 (this table is also added as Multimedia Appendix 6, with additional details). Comparisons of the average values indicate that only a few methods differ significantly ($P>.05$; see Table 1 for exact *P*-values).

Figure 5. Illustration of the first part of identifying a matching method. Screening the data set for 10 "best" matches for one user (using the leave-one-out-method). This scheme only illustrates the basic principle of the simulation. The exact results are shown in Table 1. GBS: Guillain-Barré syndrome; M. Pompe: Morbus Pompe.

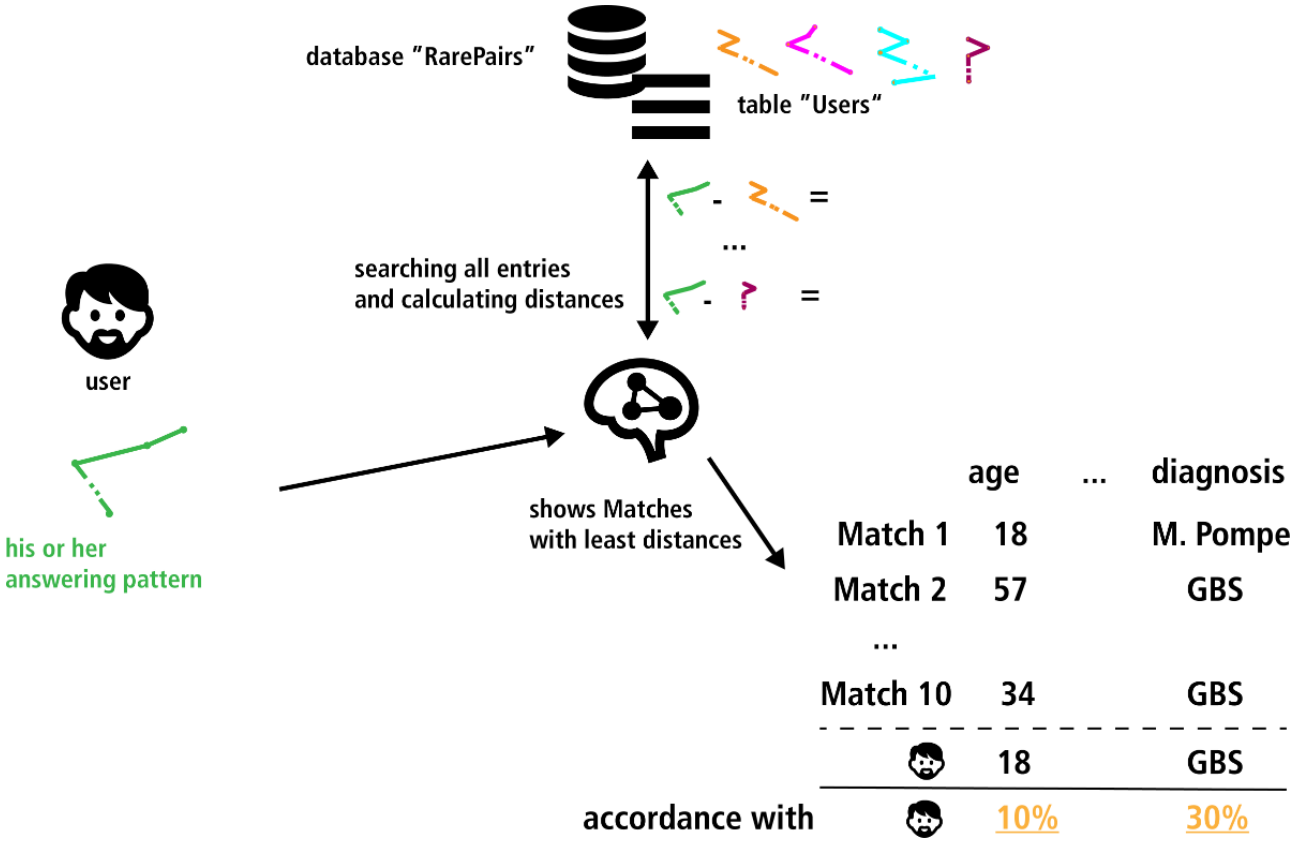


Figure 6. Schematic illustration of the second part: identifying a matching method for a given data set of users with rare diseases. Ten matches for all 973 data sets with one calculating method, calculating the average of the matching accordance of properties. This figure illustrates the basic principle of the simulation (for the complete testing results, see Multimedia Appendix 5 and Table 1).

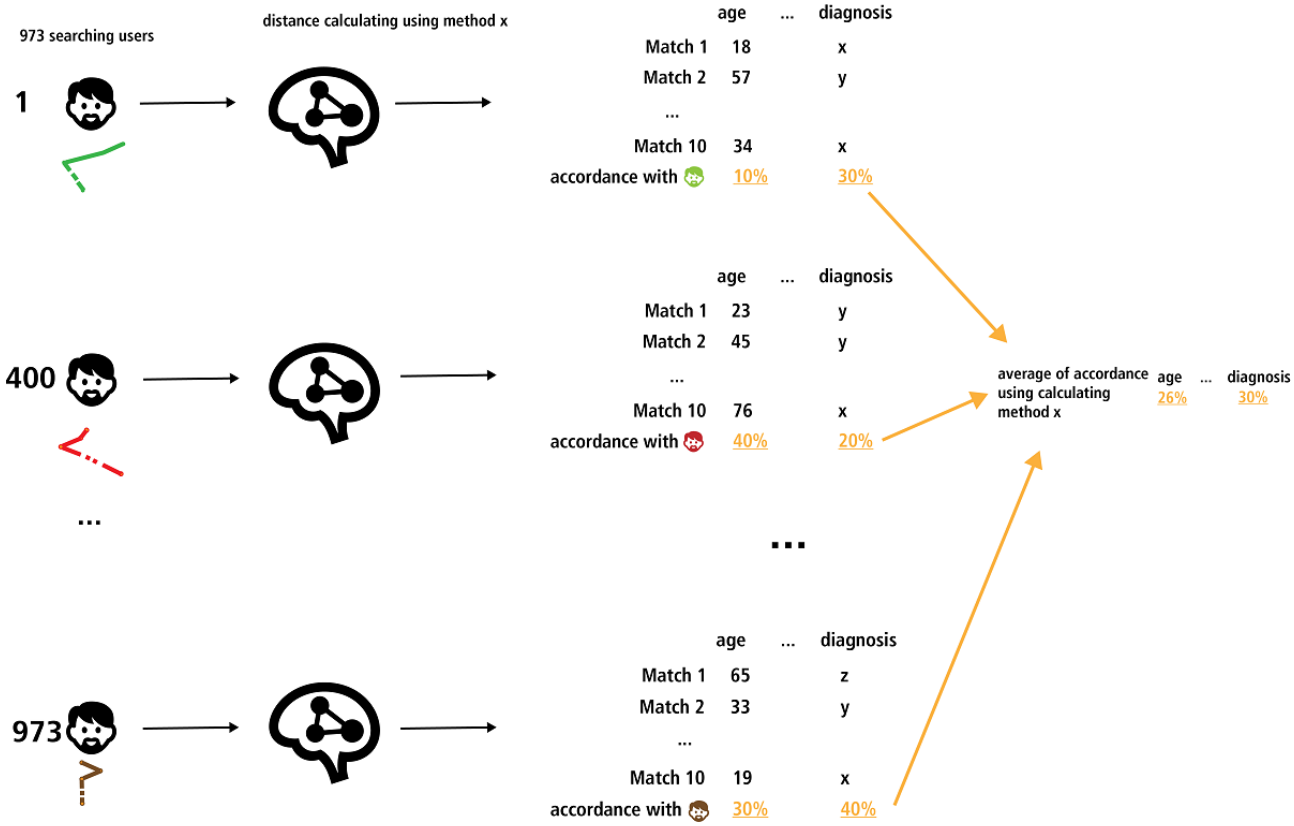


Table 1. Different distance calculating methods simulated using the leave-one-out principle.

Distance calculating method	Matching accordance in percent					
	Gender[a]	Age[b]	Latency[c]	Disease group[d]	Diagnostic system[e]	Exact diagnosis[f]
(Pseudo)random sampling (=negative benchmark)	59.6	15.7	11.9	8.3	31.7	3.8
L_p Minkowski family						
Manhattan	63.9[g]	21.9[g]	14.3	16.2[g]	40.3[g]	9.4[g]
Euclidean	62.4[g]	21.1	15.5[g]	16.0[g]	36.1[g]	9.7 (=positive benchmark)
Minkowski	63.8[g]	21.9[g]	14.4	16.2[g]	40.3[g]	9.4[g]
L_1 family						
Sørensen	64.8[g]	22.1[g]	13.2	15.3[g]	37.8[g]	8.8
Gower	63.8[g]	21.9[g]	14.4	16.2[g]	40.3[g]	9.4[g]
Canberra	64.2[g]	21.5[g]	14.0	15.8[g]	37.7[g]	9.4[g] (does not differ from the positive benchmark; *P*=.56)
Lorentzian	63.5[g]	21.7[g]	13.9	15.9[g]	39.4[g]	9.3[g]
Intersection family						
Wave Hedges	64.1[g]	21.5[g]	14.0	15.9[g]	38.3[g]	9.3
Czekanowski	64.8[g]	22.1[g]	13.2	15.3[g]	37.8[g]	8.8
Tanimoto	64.8 (=positive benchmark)	22.1 (=positive benchmark)	13.2 (differs from negative benchmark; *P*<.001)	15.3[g]	37.8[g]	8.8
Jaccard	63.8[g]	21.3	13.8	14.9	35.4[g]	8.7
Dice	63.8[g]	21.3 (differs from positive benchmark; *P*=.03)	13.8	14.9 (differs from positive benchmark; *P*=.05)	35.4[g]	8.7 (differs from negative benchmark; *P*<.001)
Inner product family						
Cosine	53.6[h] (< negative benchmark)	12.0[h] (< negative benchmark)	7.3[h] (< negative benchmark)	4.7[h] (< negative benchmark)	17.0[h] (< negative benchmark)	1.9[h] (< negative benchmark)
Fidelity family or Squared-chord family						
Bhattacharyya	64.8[g]	19.0 (differs from negative benchmark; *P*<.001)	7.9[h] (< negative benchmark)	10.1[h] (does not differ from negative benchmark; *P*=.70)	32.3 (differs from negative benchmark; *P*<.001; differs from positive benchmark; *P*<.001)	2.8[h] (< negative benchmark)
Hellinger	62.9[g]	19.3	5.8[h] (< negative benchmark)	7.4[h] (< negative benchmark)	52.5 (=positive benchmark)	0.3[h] (< negative benchmark)
Squared-Chord	62.0[g]	21.6[g]	15.3[g]	15.6[g]	35.2[g]	9.6[g]
Squared L_2 family/dX^2 family						

Distance calculating method	Matching accordance in percent					
	Gender[a]	Age[b]	Latency[c]	Disease group[d]	Diagnostic system[e]	Exact diagnosis[f]
Neyman	59.5[h] (< negative benchmark)	20.4	15.2 (differs from positive benchmark; *P*=.02)	14.9 (differs from negative benchmark; *P*<.001)	34.2[g]	9.0
Probabilistic Symmetric	62.2[g]	21.6[g]	15.1	15.8[g]	35.5[g]	9.7[g]
Clark	63.1[g]	21.4[g] (does not differ from positive benchmark; *P*=.42)	14.8	15.3[g]	35.5[g]	9.3
Additive symmetric	61.3[g] (does not differ from positive benchmark; *P*=.08)	21.1	15.7[g]	15.5[g]	34.1[g]	9.5[g]
Shannon's entropy family						
Jeffreys	62.0[g]	21.5[g]	15.3[g]	15.6[g]	35.1[g]	9.6[g]
Jensen difference	62.1[g]	21.5[g]	15.3[g] (does not differ from positive benchmark; *P*=.55)	15.7[g]	35.3[g]	9.6[g]
Combined methods						
Kumar–Johnson	60.8 (differs from positive benchmark; *P*=.02)	21.2	15.7 (=positive benchmark)	15.0[g] (does not differ from positive benchmark; *P*=.08)	33.4[g] (does not differ from positive benchmark; *P*=.14)	9.3[g] (differs from positive benchmark; *P*=.02)
Avg	63.7	21.8[h]	14.4	16.4 (=positive benchmark)	40.3[g]	9.4[g]

[a]Gender of the person.

[b]Age of the person

[c]Time with symptoms but no diagnosis.

[d]Category of the diagnosis referring to the affected organ or pathophysiology (eg, neuromuscular disease, metabolic disease)

[e]Greater category the diagnosis can be assigned to (eg, RD, CD), not especially considering the affected organ

[f]Exact name of the one diagnosis.

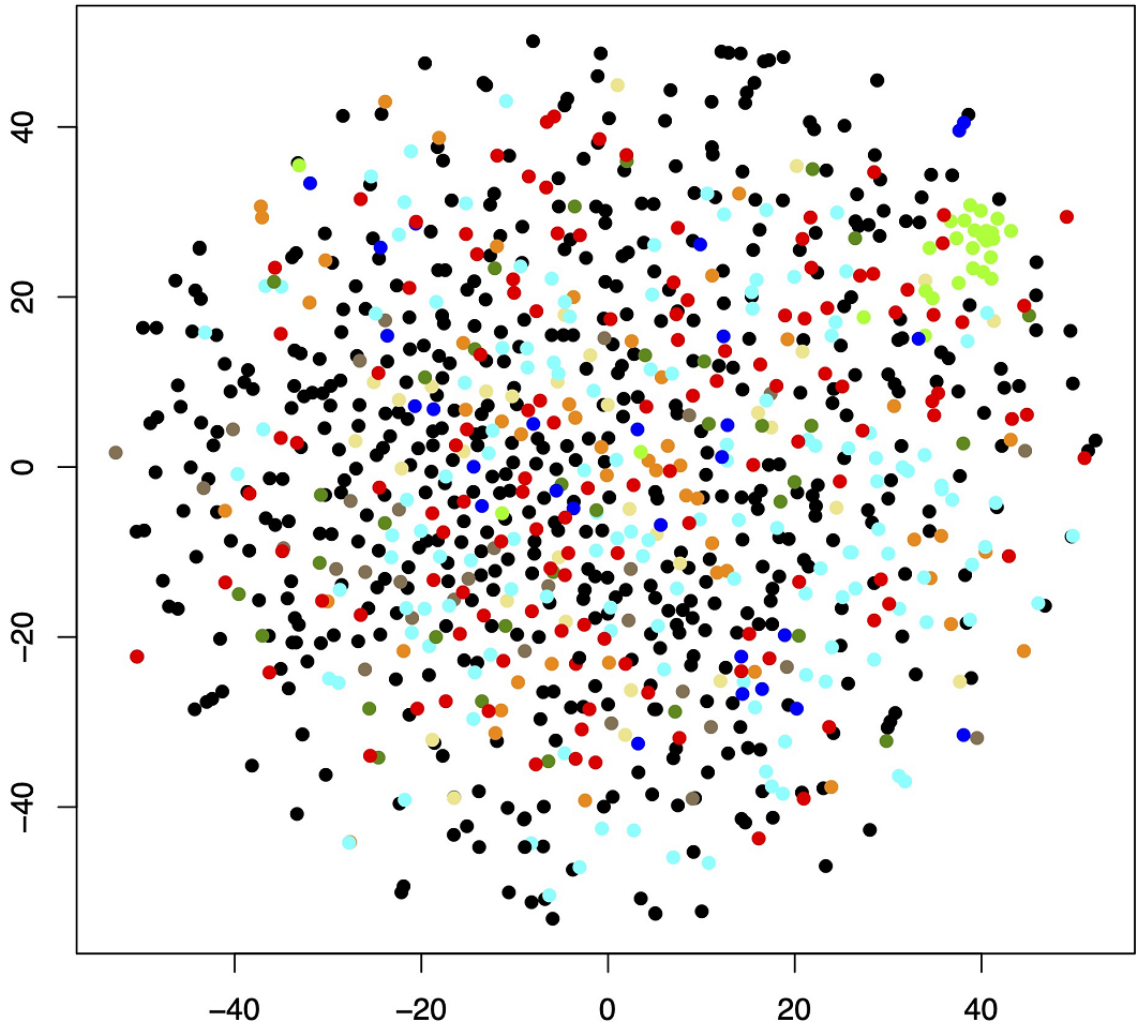
[g]Fields do not differ significantly from the positive benchmark in this category; see *P*-value in those fields. If no *P*-values are mentioned, the matching values lie in between the benchmark and the furthest value, which is only just not differing significantly from this benchmark.

[h]Fields do not differ significantly from the (pseudo)random matching (=negative benchmark); see *P*-value in those fields. If no *P*-values are mentioned, the matching values lie in between the benchmark and the furthest value, which is only just not differing significantly from this benchmark.

Furthermore, we calculated the average values for a random matching and comparison, showing that most of the calculating methods resulted in significantly better results than the random matching ($P \leq .05$; see Table 1 for exact *P*-values). These results support our assumption that the k-nearest neighbor search itself is a robust base for the matching algorithm. Additionally, the cosine method, which mathematically produces matches of people with preferably different answers, showed the unfavorable results as expected. To make sure the results fit the outcome that could be expected from the 973 people data set, we plotted the 973 people by the diagnostic system of their disease using the t-distributed stochastic neighbor embedding

method. The plot (Figure 7) illustrates that there is no clustering except for the group of healthy individuals (green dots). This finding underlines that the nearest neighbor method would produce poor results when used (solely) for classification. In RarePairs, we used this method to find suitable matches (and not to classify our data). The results in Table 1 illustrate that the nearest neighbor search is suitable for matching users answering the 53-item questionnaire under discussion in this study. In this set of data, the Jeffreys, squared-chord, or Jensen difference method proved most powerful.

Figure 7. A t-distributed stochastic neighbor embedding plot showing a possible clustering of the 973 test objects concerning the diagnostic system of their disease. Key: black: rare diseases; red: chronic diseases; dark blue: psychiatric diseases with somatoform part; dark green: unknown diagnosis; light green: healthy individuals; light blue: sarcoidosis; orange: idiopathic pulmonary arterial hypertonia; yellow: syringomyelia; brown: systemic lupus erythematosus.



Interacting on RarePairs

Once the user has found matches, s/he can interact with his/her new matches as well as other users. We implemented a few basic methods to illustrate this function.

PairChat

For interacting with 1 single user, a given user can use the *PairChat* function. PairChat is programmed as a common chat tool with the obvious functionalities of writing and receiving messages and reading them in a messenger window. Written messages are stored in the database table called *Messages*, and the link between messages and their receiver and sender is stored in the table *Messages_Link*. For showing all the messages that were exchanged between 2 users, the PHP script looks for all messages in *Messages_Link* that were sent or received by the active user, and shows them sorted by the corresponding sending or receiving chat partner and date/time.

PairCommunity

We designed the *PairCommunity* as a *broadcast forum* for all RarePairs users. Here they can share important information that might interest all registered users such as invitation to self-help groups or announcements about current scientific events. It is possible to sort the posts by different categories (database table *CommunityCategories*), such as self-help-groups or leisure time groups. The posts themselves are stored in the database table *CommunityInputs*.

Discussion

Principal Findings

In this study we outlined, created, and evaluated a prototype of a social media platform (RarePairs) for individuals with (undiagnosed) RD conditions. Evaluation of RarePairs using a single statistic function on an existing set of data illustrated decent matching results for possible users searching for a diagnosis.

About 35% of Americans use the internet for finding a diagnosis [18] and over half the global population use social media [19]. Hence, the idea of creating an online social network for people looking for a diagnosis seems to be a logical step. In that context, Russell et al [20] installed and observed a Facebook group where parents of disabled children and researchers were linked together and discussed medical studies, everyday struggles, and disease challenges. According to their analysis, 95% of the parents were motivated to join the group for connecting with like-minded people, 78% were using the group to find information, and 73% wanted to receive or give emotional support. Although the focus of that project was more in the context of research and understanding the use of social media in the context of a given medical context, the results indicate that users appreciate beneficial effects of social media in certain medical or medicosocial contexts. A popular example of a diagnostic-support online platform is *CrowdMed*, where professionals and nonprofessionals can engage in trying to help undiagnosed individuals find the right diagnosis. In contrast to our project, where diagnostic support is based on using a questionnaire, in CrowdMed patients share medical information (eg, medical reports, laboratory data). Meyer et al [21] reported first successes for a few participants and 56.9% of participants reported that the hints given by others on CrowdMed led them closer to the right diagnosis [21].

A common motivation brings individuals with various diagnoses together in *real life* self-help groups. Plinsinga et al [22] performed a survey on individuals with osteoarthritis and their interest in self-help-groups [22]. In that study, 307 of the 415 included patients were interested in participating in a self-help group; 54% of the patients reported to be engaged in a patient group, whereas 41% reported participation in an online self-help group (namely through Facebook). Such data illustrate the tremendous motivation among individuals with CD or RD to connect and support each other. RarePairs fills in a gap between individuals without diagnosis and those knowing the name of their RD.

Establishing an online social network for undiagnosed people seems to be an opportunity for younger adults during their diagnostic odyssey. They are even more likely to search for help on the internet and are more accustomed to using the internet. Lee et al [23] stated that people aged 55 or older have more difficulties finding the right (health) information on the internet than younger people.

Especially in the context of RDs, a social platform as illustrated in our study seems beneficial. Here, affected individuals from different countries could be easily connected and inspired to exchange valuable health care information crossing borders and even continents. Such a technology might be especially valuable for ultra-RD conditions with only a handful of affected individuals worldwide.

Individuals with experience in using social media and performing diagnostic research on the internet might find a platform like RarePairs advantageous. However, there might be criticism that *Googling* symptoms results in wrong information. Concerns that patients may have trouble following the doctor's recommendations after reading online information prove mostly wrong [24], and such well-informed (via the internet) patients may even help the doctor with the diagnosis [25]. Besides the opportunity of *Googling* symptoms, there is also a lack of reliable online information [26]. Moreover, wrong or disturbing information from the web might disturb the patient–doctor relationship or produce *cyberchondriacs* [27,28].

Of course, there are indicators that the internet will be a growing source of diagnostic help, but there is a lack of well-designed online tools and websites with relevant and quality-proven information [26]. With RarePairs we address those needs and promote a completely different strategy: by using a simple, but powerful questionnaire (Q53), users are connected without needing profound medical knowledge. In the future, this questionnaire-based tool might be improved using additional information from, for example, wearables. Additionally, the previous study [16] used a combination of different AI methods (support vector machine, random forest, logistic regression, and linear discriminant analysis), with better results in clustering diseases [16]. Perhaps the use of those additional AI methods could also improve the matching algorithm of RarePairs.

Analyzing the matching algorithm was an important result of this study, highlighting that nearest neighbor methods worked significantly better than a random matching. For improving the test results, there is still room for improvements by including and evaluating other AI methods in combination with the nearest neighbor search in future research projects.

Today, there are almost daily new highlights addressing improvements in the field of diagnostic support through AI (the number of publications on PubMed containing the words AI are 10 times as high as in the 1990s, as per our manual search on the database in 2020). The fields where such algorithms already are in use are mainly within visual diagnostics (radiology, pathology, dermatology, microbiology). For example, support for doctors is tested during a polyp screening [29], predicting a coronary artery disease noninvasively, or diagnosing a sepsis in an early state [30,31]. Nevertheless, in

everyday clinical use, there are also challenges such as server infrastructure or computer power that have to be overcome [32].

Data security of (disease-specific) personal data is an important issue. In RarePairs the stored data are protected by an SSL connection and cannot be linked to an address or even a name, and as such there is a guarantee for basic data safety for all users. Encrypting the stored data could be a next step for the time RarePairs will be used as a real marketable product. Also building a decentralized database could be a very elegant way to make the user data safer during the next steps of development.

Limitations

The complete realization of RarePairs is currently only prototypic. Accordingly, there are several aspects about RarePairs that have to be addressed in the future: We set the user requirements and context of use based on information and selected assumptions. In future studies, the user satisfaction needs evaluation. Besides, it is essential to test the prototype with a larger target audience in the future and reevaluate the results.

A second limitation is the small database used for this prototype. These data might not automatically reflect the community of possible users of such a network. Consequently, the results can only be regarded as a milestone. Additionally, all records of this study belonged to individuals having their diagnosis fixed. The prototype was not tested with real individuals during their *diagnostic journey*. An evaluation with more diverse data will be performed during the next steps of RarePairs' development. Furthermore, one could question the quality criteria of a *perfect match* (same gender, same age, same diagnosis) because we do not know which persons would profit from each other *in reality*. That is why we suggest planning a prototype test phase, and until then ask the participating persons to evaluate the quality of the suggested matches. Constant adjustment of the matching methods will be part of RarePairs while in daily use.

Another limitation of this analysis is that we only evaluated the data set and matching quality by just randomized matching. One might criticize that such an evaluation is only a low bar challenge. Assuming that the user perspective on the matching quality is completely unknown, we decided that this testing fulfills the evaluation of a prototype. Further steps for evaluating the algorithm have to follow.

Additionally, as the Q53 questionnaire was designed and developed from a German perspective, its success and performance must be re-evaluated in different cultural contexts. Today, translated versions are available in English, Chinese, Portuguese, and Finnish, but a systematic trans-cultural evaluation was not yet performed.

From the technical perspective, the current prototype is restricted to usage on a personal computer (and therefore we did not use a CSS Template). We are well aware that in the context of constant growth and increasing numbers of RarePairs users, adoption of the code, and focus on the development of a cross-platform app/website, as most people prefer to use social media on their mobile device [33], a template would be useful. This prototypic evaluation was only designed to be a scientific *proof of concept* and therefore uses an easy-to-use and flexible programming technique. The development of a mobile app is the next logical step.

For the use of forms, JavaScript could also be of advantage because it makes the forms more interactive and the entries can be checked more easily. Concerning the details of the prototype, there are a few functionalities that should be implemented in the future:

- Answers to the Q53 questionnaire should be changeable if the user gets new knowledge about their disease, or if the experience of the symptoms/disease changes. There must be a possibility to get new matches based on these changed answers.
- The display of the matches should contain a form of ranking, possibly showing how many questions were answered similarly or which questions had the biggest effect for the matching (that is how the diagnostic app *Ada* [34] does it.).
- It should be possible to search for other users even if they are not a fitting match (eg, to find friends from real life).
- The *PairCommunity* should be searchable for posts concerning different diagnoses.
- A contact form should be implemented.
- The chat could be using the XMPP (Extensible Messaging and Presence Protocol) to make cross-platform chatting possible.

Besides, any tool in the context of diagnostic support must obviously prevent individuals in despair from raising hope for easy solutions or diagnoses using the internet or a given platform. This cannot be offered by RarePairs in its current structure, and consequently is only one piece in a larger puzzle to support individuals during a difficult diagnostic search. RarePairs therefore strongly underlines that it is not designed as a diagnostic tool (eg, via disclaimers during the registering process).

Next steps

Our next goal is to present RarePairs to real users (eg, in self-help groups) and collect feedback systematically. For a first round with approximately 500 users, we would need additional resources for legal advice, implementing new functions and providing more (data) security.

Acknowledgments

We thank Sandra Mehmecke, Ann-Katrin Rother, Susanne Blöß, and Ulrike Schumacher for their shared expertise in the field of diagnostic search of individuals with undiagnosed conditions and RDs. We also thank those patient groups that supported our work with their insight point of view. The Robert Bosch Stiftung funded a fair part of previous research and therefore made this prototype possible which is gratefully acknowledged.

Conflicts of Interest

WL, FK, UM, and LG are cofounders of KIMedi GmbH, Deutschland. The other authors have no conflicts of interest to disclose. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the words. Of note, the manuscript has not been published, or submitted for publication elsewhere.

Multimedia Appendix 1
Analytic steps for the preparation of RarePairs.
[DOCX File , 14 KB - jmir_v22i9e21849_app1.docx]

Multimedia Appendix 2
Questionnaire Q53 [german].
[PDF File (Adobe PDF File), 970 KB - jmir_v22i9e21849_app2.pdf]

Multimedia Appendix 3
Questionnaire Q53 [english].
[PDF File (Adobe PDF File), 58 KB - jmir_v22i9e21849_app3.pdf]

Multimedia Appendix 4
Collection of RarePairs Screenshots.
[PDF File (Adobe PDF File), 1185 KB - jmir_v22i9e21849_app4.pdf]

Multimedia Appendix 5
Methods for distance calculating.
[PDF File (Adobe PDF File), 100 KB - jmir_v22i9e21849_app5.pdf]

Multimedia Appendix 6
Different distance calculating methods simulated using the leave-one-out-principle.
[PDF File (Adobe PDF File), 326 KB - jmir_v22i9e21849_app6.pdf]

References

1. Aymé S, Schmidtke J. Networking for rare diseases: a necessity for Europe. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2007 Dec;50(12):1477-1483. [doi: 10.1007/s00103-007-0381-9] [Medline: 18026888]
2. Global GP. Rare Facts. Published. 2019. URL: https://globalgenes.org/rare-facts/ [accessed 2020-08-03]
3. Limb L, Nutt S, Sen A. Experiences of rare diseases: An insight from patients and families. London, UK: Rare Disease UK; 2010. URL: https://www.raredisease.org.uk/media/1594/rduk-family-report.pdf [accessed 2020-09-20]
4. Hendriksz CJ. Rare disease impact report: Insights from patients and the medical community. 2013. URL: https://www.researchgate.net/publication/236982217_Rare_Disease_Impact_Report_Insights_from_patients_and_the_medical_community [accessed 2020-09-20]
5. Blöß S, Klemann C, Rother A, Mehmecke S, Schumacher U, Mücke U, et al. Diagnostic needs for rare diseases and shared prediagnostic phenomena: Results of a German-wide expert Delphi survey. PLoS One 2017;12(2):e0172532 [FREE Full text] [doi: 10.1371/journal.pone.0172532] [Medline: 28234950]
6. Pierucci P, Lenato GM, Suppressa P, Lastella P, Triggiani V, Valerio R, et al. A long diagnostic delay in patients with Hereditary Haemorrhagic Telangiectasia: a questionnaire-based retrospective study. Orphanet J Rare Dis 2012 Jun 07;7:33 [FREE Full text] [doi: 10.1186/1750-1172-7-33] [Medline: 22676497]
7. Brown P, Lyson M, Jenkins T. From diagnosis to social diagnosis. Soc Sci Med 2011 Sep;73(6):939-943. [doi: 10.1016/j.socscimed.2011.05.031] [Medline: 21705128]
8. Lin C, Kalb SJ, Yeh W. Delay in Diagnosis of Spinal Muscular Atrophy: A Systematic Literature Review. Pediatr Neurol 2015 Oct;53(4):293-300 [FREE Full text] [doi: 10.1016/j.pediatrneurol.2015.06.002] [Medline: 26260993]
9. Richards J, Korgenski EK, Srivastava R, Bonkowsky JL. Costs of the diagnostic odyssey in children with inherited leukodystrophies. Neurology 2015 Sep 29;85(13):1167-1170 [FREE Full text] [doi: 10.1212/WNL.0000000000001974] [Medline: 26320197]
10. Statistisches Amt der Europäischen Union (Eurostat). Einzelpersonen - Internet-Aktivitäten. Eurostat. 2018. URL: https://ec.europa.eu/eurostat/de/web/products-datasets/-/ISOC_CI_AC_I [accessed 2018-12-20]
11. Siempos II, Spanos A, Issaris EA, Rafailidis PI, Falagas ME. Non-physicians may reach correct diagnoses by using Google: a pilot study. Swiss Med Wkly 2008 Dec 13;138(49-50):741-745. [Medline: 19130327]
12. Tang H, Ng JHK. Googling for a diagnosis--use of Google as a diagnostic aid: internet based study. BMJ 2006 Dec 2;333(7579):1143-1145. [doi: 10.1136/bmj.39003.640567.AE] [Medline: 17098763]

13. Perrin A. Social Networking Usage: 2005-2015. Pew Research Center. 2015 Oct. URL: http://www.pewinternet.org/2015/10/08/2015/Social-Networking-Usage-2005-2015/ [accessed 2015-10-08]
14. Smith A. Why Americans use social media. Pew Research Center. 2011 Nov 15. URL: https://www.pewresearch.org/internet/2011/11/15/why-americans-use-social-media/ [accessed 2011-11-15]
15. Zentrales Informationsportal über seltene Erkrankungen. ZIPSE. portal-se.de. URL: https://www.portal-se.de [accessed 2020-08-03]
16. Grigull L, Mehmecke S, Rother A, Blöß S, Klemann C, Schumacher U, et al. Common pre-diagnostic features in individuals with different rare diseases represent a key for diagnostic support with computerized pattern recognition? PLoS One 2019 Oct 10;14(10):e0222637 [FREE Full text] [doi: 10.1371/journal.pone.0222637] [Medline: 31600214]
17. Cha SH. Comprehensive survey on distance/similarity measures between probability density functions. International Journal of Mathematical Models and Methods in Applied Sciences 2007;1:300-307 [FREE Full text]
18. Fox S, Duggan M. Health online 2013. 2013. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-09-18]
19. Broadband Commission for Sustainable Development. The State of Broadband: Broadband catalyzing sustainable development. 2017 Sep. URL: https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.18-2017-PDF-E.pdf [accessed 2020-09-20]
20. Russell DJ, Sprung J, McCauley D, Kraus de Camargo O, Buchanan F, Gulko R, et al. Knowledge Exchange and Discovery in the Age of Social Media: The Journey From Inception to Establishment of a Parent-Led Web-Based Research Advisory Community for Childhood Disability. J Med Internet Res 2016 Nov 11;18(11):e293 [FREE Full text] [doi: 10.2196/jmir.5994] [Medline: 27836818]
21. Meyer AND, Longhurst CA, Singh H. Crowdsourcing Diagnosis for Patients With Undiagnosed Illnesses: An Evaluation of CrowdMed. J Med Internet Res 2016;18(1):e12 [FREE Full text] [doi: 10.2196/jmir.4887] [Medline: 26769236]
22. Plinsinga ML, Besomi M, Maclachlan L, Melo L, Robbins S, Lawford BJ, et al. Exploring the Characteristics and Preferences for Online Support Groups: Mixed Method Study. J Med Internet Res 2019 Dec 03;21(12):e15987 [FREE Full text] [doi: 10.2196/15987] [Medline: 31793893]
23. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
24. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
25. Bouwman MG, Teunissen QGA, Wijburg FA, Linthorst GE. 'Doctor Google' ending the diagnostic odyssey in lysosomal storage disorders: parents using internet search engines as an efficient diagnostic strategy in rare diseases. Arch Dis Child 2010 Aug;95(8):642-644. [doi: 10.1136/adc.2009.171827] [Medline: 20418338]
26. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
27. Hart A, Henwood F, Wyatt S. The role of the Internet in patient-practitioner relationships: findings from a qualitative research study. J Med Internet Res 2004 Sep 30;6(3):e36 [FREE Full text] [doi: 10.2196/jmir.6.3.e36] [Medline: 15471762]
28. Smith PK, Fox AT, Davies P, Hamidi-Manesh L. Cyberchondriacs. Int J Adolesc Med Health 2006;18(2):209-213. [Medline: 16894859]
29. Bell LTO, Gandhi S. A comparison of computer-assisted detection (CAD) programs for the identification of colorectal polyps: performance and sensitivity analysis, current limitations and practical tips for radiologists. Clin Radiol 2018 Jun;73(6):593.e11-593.e18. [doi: 10.1016/j.crad.2018.02.009] [Medline: 29602538]
30. Alizadehsani R, Hosseini MJ, Khosravi A, Khozeimeh F, Roshanzamir M, Sarrafzadegan N, et al. Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries. Comput Methods Programs Biomed 2018 Aug;162:119-127. [doi: 10.1016/j.cmpb.2018.05.009] [Medline: 29903478]
31. Schinkel M, Paranjape K, Nannan Panday RS, Skyttberg N, Nanayakkara P. Clinical applications of artificial intelligence in sepsis: A narrative review. Comput Biol Med 2019 Dec;115:103488 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.103488] [Medline: 31634699]
32. Yoshida H, Näppi J. CAD in CT colonography without and with oral contrast agents: progress and challenges. Comput Med Imaging Graph 2007;31(4-5):267-284. [doi: 10.1016/j.compmedimag.2007.02.011] [Medline: 17376650]
33. 2016 U.S. Cross-platform Future in Focus . comScore. 2016. URL: https://www.comscore.com/Insights/Presentations-and-Whitepapers/2016/2016-US-Cross-Platform-Future-in-Focus [accessed 2020-09-18]
34. Miller S, Gilbert S, Virani V, Wicks P. Patients' Utilization and Perception of an Artificial Intelligence-Based Symptom Assessment and Advice Technology in a British Primary Care Waiting Room: Exploratory Pilot Study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]

Abbreviations

AI: artificial intelligence

CD: chronic disease
EU: European Union
RD: rare disease

Edited by G Eysenbach; submitted 26.06.20; peer-reviewed by A Meyer, E Alanazi; comments to author 30.07.20; revised version received 13.08.20; accepted 18.08.20; published 29.09.20.

Please cite as:
Kühnle L, Mücke U, Lechner WM, Klawonn F, Grigull L
Development of a Social Network for People Without a Diagnosis (RarePairs): Evaluation Study
J Med Internet Res 2020;22(9):e21849
URL: http://www.jmir.org/2020/9/e21849/
doi:10.2196/21849
PMID:32990634

Original Paper

Factors Engaging Users of Diabetes Social Media Channels on Facebook, Twitter, and Instagram: Observational Study

Elia Gabarron[1], PhD; Dillys Larbi[1], MSc, MPH; Enrique Dorronzoro[2], PhD; Per Erlend Hasvold[3], MSc; Rolf Wynn[4], MD, PhD; Eirik Årsand[1,5], PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[2]Department of Electronic Technology, University of Seville, Seville, Spain

[3]Medical Division, University Hospital of North Norway, Tromsø, Norway

[4]Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway

[5]Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Elia Gabarron, PhD
Norwegian Centre for E-health Research
University Hospital of North Norway
Sykehusveien. 23
Tromsø,
Norway
Phone: 47 77754030
Fax: 47 77754099
Email: elia.gabarron@ehealthresearch.no

Abstract

Background: Diabetes patient associations and diabetes-specific patient groups around the world are present on social media. Although active participation and engagement in these diabetes social media groups has been mostly linked to positive effects, very little is known about the content that is shared on these channels or the post features that engage their users the most.

Objective: The objective of this study was to analyze (1) the content and features of posts shared over a 3-year period on 3 diabetes social media channels (Facebook, Twitter, and Instagram) of a diabetes association, and (2) users' engagement with these posts (likes, comments, and shares).

Methods: All social media posts published from the Norwegian Diabetes Association between January 1, 2017, and December 31, 2019, were extracted. Two independent reviewers classified the posts into 7 categories based on their content. The interrater reliability was calculated using Cohen kappa. Regression analyses were carried out to analyze the effects of content topic, social media channel, and post features on users' engagement (likes, comments, and shares).

Results: A total of 1449 messages were posted. Posts of interviews and personal stories received 111% more likes, 106% more comments, and 112% more shares than miscellaneous posts (all $P<.001$). Messages posted about awareness days and other celebrations were 41% more likely to receive likes than miscellaneous posts ($P<.001$). Conversely, posts on research and innovation received 31% less likes ($P<.001$), 35% less comments ($P=.02$), and 25% less shares ($P=.03$) than miscellaneous posts. Health education posts received 38% less comments ($P=.003$) but were shared 39% more than miscellaneous posts ($P=.007$). With regard to social media channel, Facebook and Instagram posts were both 35 times more likely than Twitter posts to receive likes, and 60 times and almost 10 times more likely to receive comments, respectively ($P<.001$). Compared to text-only posts, those with videos had 3 times greater chance of receiving likes, almost 4 times greater chance of receiving comments, and 2.5 times greater chance of being shared (all $P<.001$). Including both videos and emoji in posts increased the chances of receiving likes by almost 7 times ($P<.001$). Adding an emoji to posts increased their chances of receiving likes and being shared by 71% and 144%, respectively ($P<.001$).

Conclusions: Diabetes social media users seem to be least engaged in posts with content topics that a priori could be linked to greater empowerment: research and innovation on diabetes, and health education. Diabetes social media groups, public health authorities, and other stakeholders interested in sharing research and innovation content and promoting health education on social media should consider including videos and emoji in their posts, and publish on popular and visual-based social media channels, such as Facebook and Instagram, to increase user engagement.

International Registered Report Identifier (IRRID): RR2-10.1186/s12913-018-3178-7

(J Med Internet Res 2020;22(9):e21204) doi:10.2196/21204

KEYWORDS

social media; Facebook; Twitter; Instagram; diabetes; engagement

Introduction

Patient associations and patient groups from around the world are increasingly more present on social media. Being both ubiquitous and freely accessible, social media channels allow patient associations to share content and connect with individuals interested in their health condition. Representing one of the most prevalent chronic diseases worldwide, diabetes associations and diabetes patient groups can also be found on social media [1-4].

Belonging to health-related groups on social media has been linked to several benefits for users, including a reduction in feelings of isolation [5,6], an increased sense of belonging [5,7], positive confirmation of their own situation [5,8], an enhanced sense of well-being [9,10], increased feelings of empowerment [2,11-15], and better health outcomes for users of diabetes-specific social media [11,16-19].

However, although active participation and engagement in diabetes-specific social media groups is mostly linked to positive effects, very little is known about the content that is shared on these channels or what features of posts engage their users the most. In a previous study [20], we surveyed followers of the Norwegian Diabetes Association's social media channels, and we found that almost all the respondents wanted more content about research and innovation on diabetes in social media groups, preferably in text format. However, other previous studies have reported that social media groups for patients with diabetes mostly shared content about diabetes self-management [1-4], scientific content [3,4], health care services [3,4], diabetes awareness [3,4], personal stories [2], or humor [2].

One way of assessing if posted messages engage users of diabetes-related social media is by measuring the posts' received feedback in the form of likes, shares, and comments. Likes and shares are a form of communication that allows social media users to provide feedback to other users with a simple click [21,22]. This quick interaction (ie, likes and shares) signals the user's agreement with the published content [21], and is perceived as a way of supporting the post [22]. Writing comments on social media, which requires more effort than a simple click, has been associated with either strong agreement or strong disagreement with the post [21,23]. Previous studies have reported that Facebook posts including media (ie, pictures, videos, or emoji), providing links, or expressing positive sentiments engage users the most [24], while posts including links and expressing negative sentiments are the least shared [25]. Social media posts dealing with diabetes management and expressing negative sentiments seem to receive more likes when the post is text-only, and less likes when the post includes images [25].

The objective of this study was to analyze the content topic and features of posts that were shared over 3 years in the 3 diabetes-related social media channels (Facebook, Twitter, and Instagram) of the Norwegian Diabetes Association, as well as the users' engagement with these posts. This study is part of a participatory research project on the use of social media for health promotion in diabetes [26]. This project is carried out in collaboration with Diabetesforbundet, the main diabetes association in Norway [27]. By January 2020, the association had more than 34,000 followers on Facebook, more than 7000 followers on Instagram, and more than 3000 followers on Twitter.

Methods

Data Extraction

All social media posts from the Norwegian Diabetes Association (on Facebook, Twitter, and Instagram) published between January 1, 2017, and December 31, 2019, were extracted and included in the study (no posts were excluded or removed from the analysis). The social media posts were extracted using the manager tool for Facebook, manually for Instagram, and using the standard application programming interface (API) for Twitter. Using the appropriate functions, the standard API allows us to gather all tweets from a specific timeframe. In this way, a PHP script was programmed to query Twitter for all the tweets made by the Norwegian Diabetes Association during the study's timeframe. The Twitter data, including the text and the tweet metadata (ie, date of publishing, likes, retweets, etc), were then exported into a Microsoft Excel file document. The following information was extracted from each post: text message. post features (ie, use of emoji, picture, and/or video), and number of likes, comments, and shares. For Facebook, we collected the total number of likes, including the reactions, for each post. For Instagram, we only extracted likes and comments because shares were not an available option at the time of the study.

Code Categories

We classified the content topics into 7 categories. These categories were based on findings from our previous studies [3,4,20], and consisted of (1) health education (including self-management and self-monitoring, information about the condition, and promotion of exercise), (2) research and innovation on diabetes (where results of an investigation were reported), (3) diabetes-related technology (including information about apps, blood-glucose monitors, and insulin pumps, but unrelated to politics), (4) interviews and personal stories, (5) awareness days and other celebrations, (6) recipes and food-related information, and (7) miscellaneous (including information about politics related to diabetes, announcements of conferences, courses, meetings, and events).

Two independent reviewers classified the text message of each post according to its main topic. When a post was considered to fall into more than one category, reviewers were trained and instructed to choose the main topic among the 7 possible options. Discrepancies in the posts' classification were discussed with a third reviewer until reaching consensus. The inter-rater reliability was calculated using Cohen kappa analysis.

Statistical Analyses

All descriptive and regression analyses were performed using SPSS software (version 25; IBM Corp). The dependent variables in the regression analyses were the number of likes, comments, and shares, which were count data and non-normally distributed. Negative binomial regression models emerged as most appropriate based on the overdispersion parameters and the goodness-of-fit indices. For each of the dependent variables, we performed multilevel negative binomial regression, with the predictors being the independent nominal variables. The largest category in each group was used as the reference: content topic (reference group: miscellaneous), social media channel (reference group: Twitter), and post features (reference group: text only). We determined the interaction between the content topic, social media channel, and post features. The level of significance was set at $P<.05$.

Ethics

The study protocol was exempted from requiring ethical approval by the Norwegian Regional Ethics Committee (2017/764/REK Sør-ØstC), as it falls outside the scope of the Norwegian Health Research Act. The treatment of personal information was approved by the Data Protection Officer at the University Hospital of North Norway (ref 0720).

Results

Sample

During the 3-year period of the study, the Norwegian Diabetes Association posted a total of 1449 messages on their social media channels: 718 (49.55%) were posted on Twitter, 530 (36.58%) were posted on Facebook, and 201 (13.87%) were posted on Instagram. The number of posts on each social media channel according to post features (text only, inclusion of picture, video, and/or emoji) is summarized in Table 1.

Table 1. Number of posts on 3 social media channels according to post features.

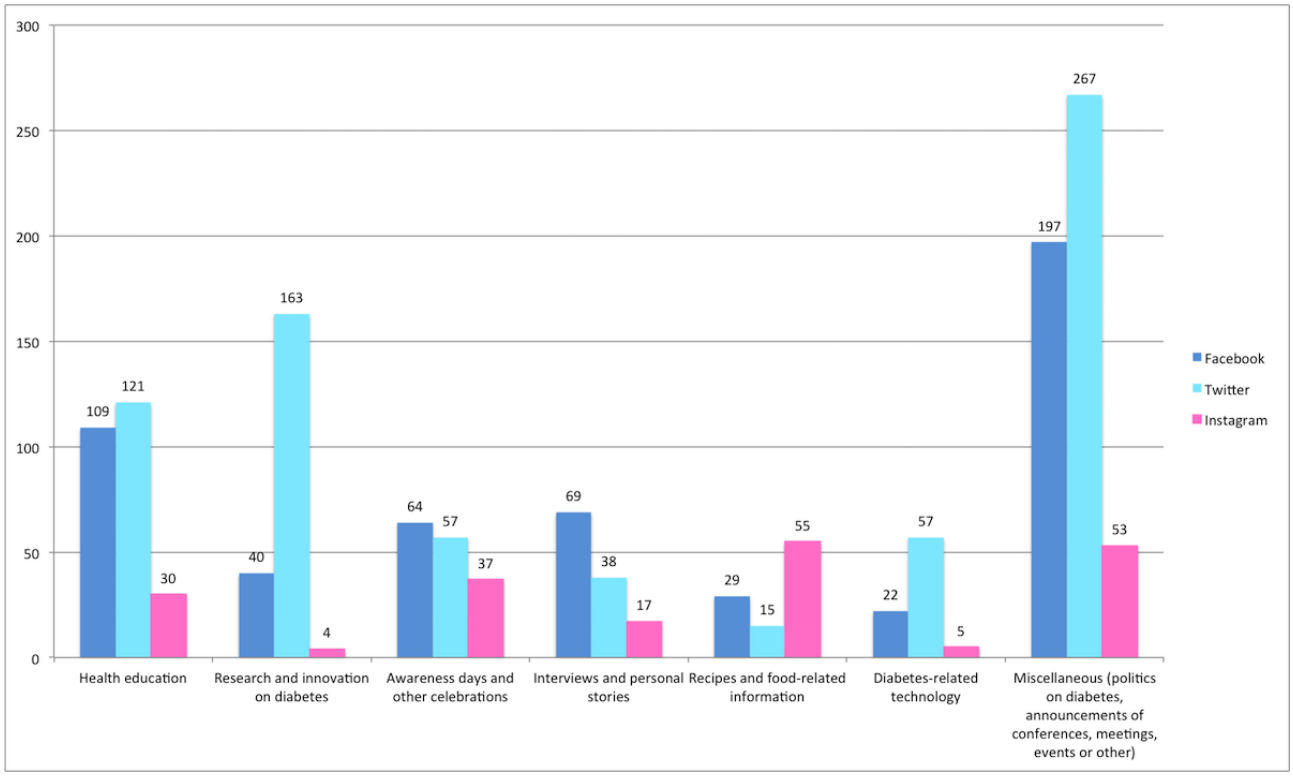
	Without emoji	With emoji	Total
Facebook	410	120	530
Text post	247	73	320
Text + picture	105	30	135
Text + video	58	17	75
Twitter	698	20	718
Text post	455	8	463
Text + picture	225	11	236
Text + video	18	1	19
Instagram	15	186	201
Text post	N/A[a]	N/A	N/A
Text + picture	12	169	181
Text + video	3	17	20
Total	1123	326	1449

[a]N/A: not applicable

Content Topic Classification

The interrater agreement of the posts' main topic was found to be substantial (κ=0.695, κ=0.780, and κ=0.789, for Twitter, Facebook, and Instagram posts, respectively) [28]. Most of the social media posts fell into the miscellaneous category (517/1449, 35.68%), followed by health education (260/1449, 17.94%), and research and innovation on diabetes (207/1449, 14.29%). With only 84 posts in the 3-year period, diabetes-related technology was the least represented category (5.80%). The total number of posts on each social media channel according to content is shown in Figure 1.

Figure 1. Number of diabetes-related posts on 3 social media channels according to content (January 1, 2017, to December 31, 2019).



Engagement: Likes, Comments, and Shares

The effect of content topic, social media channel, and post features on the measures of users' engagement was analyzed using negative binomial regression.

The regression analysis showed that posts of interviews and personal stories received 111% more likes, 106% more comments, and 112% more shares than miscellaneous posts (P<.001 for all). Posts on the topics of awareness days and other celebrations were 41% more likely to receive likes than miscellaneous posts (P<.001). On the other hand, posts of recipes and food-related information and posts discussing research and innovation on diabetes received 47% and 31% less likes, respectively, than miscellaneous posts (both P<.001). The posts that received fewer comments than miscellaneous posts were those with recipes and food-related information (59% less comments, P<.001), posts discussing research and innovation on diabetes (35% less comments, P=.02), and posts related to health education (38% less comments, P=.003). Health education posts were shared 39% more often than miscellaneous posts (P=.007), while posts of recipes and food-related information and posts of research information and innovation on diabetes had 43% and 25% less shares, respectively, than miscellaneous posts (P=.02 and P=.03, respectively).

With regard to social media channel, Facebook posts were 35 times more likely than Twitter posts to receive likes, 60 times more likely to receive comments, and 13 times more likely to be shared (P<.001). Instagram posts were 35 times more likely than Twitter posts to receive likes, and more than 9 times more likely to receive comments (P<.001).

In terms of post features, posts that included videos were 3 times more likely to receive likes, almost 4 times more likely to receive comments, and 2.5 times more likely to be shared than text-only posts (all P<.001). The addition of both a video and an emoji to a post increased its chances of receiving likes by almost 7 times (P<.001), but no effect on comments and shares was observed. Including a picture in the post increased the chances of it receiving likes by 86% and of being shared by 124% (both P<.001), but it did not affect the number of comments. By including only an emoji to the text, the chances of posts receiving likes and being shared increased by 71% and 144%, respectively (P<.001).

Table 2 shows the negative binomial regression analyses of the effects of content topic, social media channel, and post features as predictors of users' engagement (likes, comments, and shares).

Table 2. Effect of content topic, social media channel, and post features on users' engagement (likes, comments, and shares).

Independent variables	n	Likes		Comments		Shares	
		OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Content topic							
Interviews and personal stories	124	2.11 (1.66-2.68)	<.001	2.06 (1.37-3.08)	<.001	2.12 (1.58-2.96)	<.001
Awareness days and other celebrations	158	1.41 (1.12-1.76)	.002	1.31 (0.91-1.89)	.15	1.23 (0.88-1.70)	.22
Recipes and food-related information	99	0.53 (0.40-0.70)	<.001	0.41 (0.25-0.65)	<.001	0.57 (0.35-0.92)	.02
Diabetes-related technology	84	0.78 (0.59-1.04)	.09	1.33 (0.83-2.14)	.23	0.73 (0.49-1.06)	.10
Research and innovation on diabetes	207	0.69 (0.57-0.85)	<.001	0.65 (0.45-0.93)	.02	0.75 (0.58-0.97)	.03
Health education	260	0.87 (0.72-1.04)	.13	0.62 (0.45-0.85)	.003	1.39 (1.09-1.78)	.01
Miscellaneous	517	1[b]		1[b]		1[b]	
Social media channel							
Facebook	530	35.41 (30.62-40.96)	<.001	60.37 (46.77-77.93)	<.001	12.79 (10.59-15.45)	<.001
Instagram	201	34.99 (25.13-48.72)	<.001	9.68 (5.43-17.28)	<.001	N/A[c]	
Twitter	718	1[b]		1[b]		1[b]	
Post features							
Emoji (no picture, no video)	81	1.71 (1.29-2.26)	<.001	1.09 (0.69-1.70)	.70	2.44 (1.69-3.51)	<.001
Emoji and picture	210	1.30 (0.96-1.77)	.09	0.65 (0.37-1.12)	.12	1.33 (0.80-2.19)	.27
Emoji and video	35	6.83 (4.21-11.09)	<.001	1.29 (0.66-2.56)	.45	1.82 (0.91-3.65)	.09
Picture (no emoji)	342	1.86 (1.58-2.19)	<.001	1.12 (0.83-1.49)	.46	2.24 (1.81-2.77)	<.001
Video (no emoji)	79	3.29 (2.26-4.39)	<.001	3.87 (2.37-6.31)	<.001	2.50 (1.73-3.61)	<.001
Text only (no emoji, no picture, no video)	702	1[b]		1[b]		1[b]	

[a]OR: odds ratio.

[b]Reference group in the corresponding independent variable.

[c]N/A: not applicable.

Discussion

Summary

Between January 1, 2017, and December 31, 2019, the Norwegian Diabetes Association posted a total of 1449 messages on their social media channels. The posts that were most engaging to users were those that featured interviews and personal stories. In fact, those posts were twice as likely to receive likes, comments, and shares. On the other hand, posts containing recipes and food-related information, and posts on research and innovation, were the least engaging to social media users. Both types of content were less likely to receive likes, comments, and shares. Regarding the social media channel, Facebook and Instagram posts were both 35 times more likely to receive likes, and 60 times and almost 10 times more likely to receive comments, respectively, than posts on Twitter. Video and emoji were the most engaging post features. Posts with video had 3 times increased chance of receiving likes, almost 4 times increased chance of receiving comments, and 2.5 times increased chance of being shared. Including both video and emoji increased the chances of receiving likes by almost 7 times. The addition of an emoji to a post increased its chances of receiving likes and being shared by 71% and 144%, respectively.

Content Topic Engagement

We found that diabetes social media posts that engaged the most users were the ones that featured interviews and personal stories, and those that mentioned awareness days and other celebrations. Our results indicate that posting content based on interviews and personal stories on diabetes social media channels offers the highest chances of receiving likes, comments, and shares. This finding contrasts with the results of our previous survey [20], in which interviews and personal stories were the least preferred type of content by diabetes social media users. These findings suggest that there might be a discrepancy between what users say they want and what they actually like. Such a discrepancy might be related to issues such as self-presentation

[29], where users prefer to think of themselves as less interested in personal stories than they actually are (and respond accordingly on questionnaires). Researchers have used social media liking patterns to judge users' personality types, and shown that these judgements are more accurate than those made by users' close contacts [30]. This could suggest that social media liking patterns more accurately reflect users' actual interests than do their responses to questionnaires about their preferences.

Posts containing content about awareness days and other celebrations also showed a higher number of likes. Similarly, in a previous study [3] that analyzed 2 diabetes Facebook groups, one open and one closed, posts about awareness days received more likes.

Diabetes social media posts dealing with content that promotes empowerment have been previously linked with higher engagement [25,31]. A study by Harris et al [31] found that tweets that included information about diabetes-related health problems were positively and significantly associated with engagement. Another study [25] reported that posts dealing with diabetes control received more likes, and posts on diabetes' consequences were associated with greater sharing. In our study, posts that a priori could be linked to greater empowerment, such as those dealing with research and innovation on diabetes or health education, were predictors of less engagement (fewer likes, comments, and shares for research and innovation posts; and fewer comments and shares for health education posts). These results are also in discrepancy with our survey findings [20], in which 78% of respondents indicated that they would like to find more content on research and innovation on diabetes on social media channels. This discrepancy between what users say they want in a survey and what they actually like in real life might also be related to self-presentation issues [29], where users prefer thinking and saying that they are more interested in research than they actually are. Another possible reason for this discrepancy could be that this user group really wants to read about research and innovation but does not feel competent or able to comment or acknowledge by liking and sharing this kind of information.

Post Features Engagement

The use of videos predicted higher chances of receiving likes, comments, and shares. The inclusion of pictures and emoji also predicted an increased number of likes and shares. Our results are in concordance with previous publications [24,32] that report the use of videos as one of the key features for attracting the greatest amount of user engagement. Technically, videos compel users to stop scrolling for a brief time to perceive and digest the content, which may also be conducive to engagement. Moreover, videos can convey a message in and of themselves and do not need to be accompanied by text, which may increase their effectiveness in communication and their ability to engage users.

Our results on the effect of pictures on engagement are also in line with the findings of a previous study [25] that analyzed 10 diabetes-related Facebook pages, and with a study [24] that focused on engagement with health agencies on Facebook. These two studies [24,25] reported higher rates of liking and sharing of posts with images. Likewise, the use of emoji in social media posts (which are linked to a more positive sentiment) has also been linked to higher levels of user engagement in previous research [32,33].

Health Implications of Social Media Groups

Disease-specific social media groups, such as the one we analyzed in this study, are recognized as trusted sources of information. Patient associations on social media reach and engage a considerable number of people, which can potentially benefit their users at many levels, including with respect to health outcomes [16,17,19]. These channels could supplement the traditional delivery of information provided by health care professionals today. However, there is still a proportion of people with diabetes who are not benefiting from these channels because they do not use social media channels, they are not interested in belonging to a group discussing their disease (or displaying their health condition by joining the group), or they simply do not know about the existence of such groups.

Analyzing users' engagement with social media posts is a way for patient associations, healthcare authorities, and other stakeholders to understand the voice of the patient [17] and to know what people are interested in at a specific moment. However, measuring social media engagement as a way of understanding users' interests could soon face new challenges. Administrators of some popular channels currently believe that showing engagement metrics could be limiting an increase in the volume of posts [34]. In fact, at the end of 2019, Instagram started to hide the number of likes displayed underneath posts in some countries [34]. If these strategies are adopted by more social media channels, alternative approaches to understanding diabetes patients' interests on these channels would be needed. Alternative strategies to listening to social media users' interests might include the use of automatic topic classification based on natural language processing or other artificial intelligence techniques, which could help to identify the most popularly discussed or searched themes. The use of sentiment analyses could also help us to understand which of these topics are linked with a more positive or a more negative sentiment.

Limitations

This study refers to diabetes social media groups led by a national patient association. Although the channels are open, only social media administrators within the patient association are able to post. Users are only able to respond to these posts. The type of content posted by this organization might differ from that posted in other diabetes social media groups.

Conclusion

Diabetes social media users seem to be least engaged in post content that a priori could be linked to greater empowerment: research and innovation on diabetes, and health education. Diabetes social media groups, public health authorities, and other stakeholders interested in sharing research and innovation content and in promoting health education that engages social media users should consider including videos and emoji in their posts, and preferably publish on popular and visual-based channels, such as Facebook and Instagram, to increase user engagement.

Acknowledgments

We sincerely thank the Norwegian Diabetes Association (Diabetesforbundet) for their engagement in this study. This study was funded by Northern Norway Regional Health Authority (Helse Nord RHF) grant number HNF1370-17. ED receives funding and is supported by the V Plan Propio de Investigación de la Universidad de Sevilla, Spain.

Authors' Contributions

EG, RW, and EÅ designed the study. EG and ED extracted the data. EG, DL, and PEH classified the posts. EG and DL analyzed the data. EG took the lead in writing the manuscript and all authors provided critical feedback. All coauthors have revised the manuscript and approved its final version.

Conflicts of Interest

None declared.

References

1. Stellefson M, Paige S, Apperson A, Spratt S. Social Media Content Analysis of Public Diabetes Facebook Groups. J Diabetes Sci Technol 2019 Mar 31;13(3):428-438. [doi: 10.1177/1932296819839099]
2. Tenderich A, Tenderich B, Barton T, Richards SE. What Are PWDs (People With Diabetes) Doing Online? A Netnographic Analysis. J Diabetes Sci Technol 2019;13(2):187-197. [doi: 10.1177/1932296818813192] [Medline: 30477335]
3. Årsand E, Bradway M, Gabarron E. What Are Diabetes Patients Versus Health Care Personnel Discussing on Social Media? J Diabetes Sci Technol 2019 Jan 13;13(2):198-205. [doi: 10.1177/1932296818821649]
4. Gabarron E, Bradway M, Årsand E. What are diabetes patients discussing on social media? Int J Integr Care 2016 Nov 09;16(5):14. [doi: 10.5334/ijic.2600]
5. Mattson MJ, Hall JG. Linking health communication with social support. In: Health as communication nexus : a service-learning approach. United States: Dubuque, IA : Kendall Hunt Publishing Company; 2011.
6. Hajek A, König H. The association between use of online social networks sites and perceived social isolation among individuals in the second half of life: results based on a nationally representative sample in Germany. BMC Public Health 2019 Jan 9;19(1):40. [doi: 10.1186/s12889-018-6369-6]
7. Wong D, Amon KL, Keep M. Desire to Belong Affects Instagram Behavior and Perceived Social Support. Cyberpsychology, Behavior, and Social Networking 2019 Jul;22(7):465-471. [doi: 10.1089/cyber.2018.0533]
8. Alzahrani A, Alanzi T. Social Media Use By People With Diabetes In Saudi Arabia: A Survey About Purposes, Benefits And Risks. DMSO 2019 Nov;12:2363-2372. [doi: 10.2147/dmso.s208141]
9. Kuykendall L, Tay L, Ng V. Leisure engagement and subjective well-being: A meta-analysis. Psychological Bulletin 2015;141(2):364-403. [doi: 10.1037/a0038508]
10. Huang C. Time Spent on Social Network Sites and Psychological Well-Being: A Meta-Analysis. Cyberpsychology, Behavior, and Social Networking 2017 Jun;20(6):346-354. [doi: 10.1089/cyber.2016.0758]
11. Litchman ML, Edelman LS, Donaldson GW. Effect of Diabetes Online Community Engagement on Health Indicators: Cross-Sectional Study. JMIR Diabetes 2018 Apr 24;3(2):e8. [doi: 10.2196/diabetes.8603]
12. Benetoli A, Chen T, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Education and Counseling 2018 Mar;101(3):439-444. [doi: 10.1016/j.pec.2017.08.015]
13. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: The benefits and challenges. Health Informatics J 2014 Feb 18;20(1):50-58. [doi: 10.1177/1460458213476969]
14. Gavrila V, Garrity A, Hirschfeld E, Edwards B, Lee JM. Peer Support Through a Diabetes Social Media Community. J Diabetes Sci Technol 2019 Jan 02;13(3):493-497. [doi: 10.1177/1932296818818828]
15. Benetoli A, Chen TF, Aslani P. Consumer perceptions of using social media for health purposes: Benefits and drawbacks. Health Informatics J 2018 Sep 07;25(4):1661-1674. [doi: 10.1177/1460458218796664]
16. Gabarron E, Årsand E, Wynn R. Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review. J Med Internet Res 2018 Aug 10;20(8):e10303. [doi: 10.2196/10303]
17. Rozenblum R, Greaves F, Bates DW. The role of social media around patient experience and engagement. BMJ Qual Saf 2017 Apr 20;26(10):845-848. [doi: 10.1136/bmjqs-2017-006457]
18. Petrovski G, Zivkovic M. Are We Ready to Treat Our Diabetes Patients Using Social Media? Yes, We Are. J Diabetes Sci Technol 2018 Aug 22;13(2):171-175. [doi: 10.1177/1932296818795441]
19. Elnaggar A, Ta Park V, Lee SJ, Bender M, Siegmund LA, Park LG. Patients' Use of Social Media for Diabetes Self-Care: Systematic Review. J Med Internet Res 2020 Apr 24;22(4):e14209. [doi: 10.2196/14209]
20. Gabarron E, Dorronzoro E, Bradway M, Rivera-Romero O, Wynn R, Årsand E. Preferences and interests of diabetes social media users regarding a health-promotion intervention. PPA 2018 Nov;Volume 12:2499-2506. [doi: 10.2147/ppa.s184369]

21. Kaur W, Balakrishnan V, Rana O, Sinniah A. Liking, sharing, commenting and reacting on Facebook: User behaviors' impact on sentiment intensity. Telematics and Informatics 2019 Jun;39:25-36. [doi: 10.1016/j.tele.2018.12.005]
22. Wohn DY, Carr CT, Hayes RA. How Affective Is a "Like"?: The Effect of Paralinguistic Digital Affordances on Perceived Social Support. Cyberpsychology, Behavior, and Social Networking 2016 Sep;19(9):562-566. [doi: 10.1089/cyber.2016.0162]
23. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online. J Med Internet Res 2013 Nov 01;15(11):e239. [doi: 10.2196/jmir.2721]
24. Bhattacharya S, Srinivasan P, Polgreen P. Social media engagement analysis of U.S. Federal health agencies on Facebook. BMC Med Inform Decis Mak 2017 Apr 21;17(1):49. [doi: 10.1186/s12911-017-0447-z]
25. Rus HM, Cameron LD. Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages. ann. behav. med 2016 Apr 8;50(5):678-689. [doi: 10.1007/s12160-016-9793-9]
26. Gabarron E, Bradway M, Fernandez-Luque L, Chomutare T, Hansen AH, Wynn R, et al. Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. BMC Health Serv Res 2018 Jun 5;18(1):414. [doi: 10.1186/s12913-018-3178-7]
27. Diabetesforbundet. The Norwegian Diabetes Association. Diabetesforbundet. URL: https://www.diabetes.no [accessed 2020-09-20]
28. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
29. Baumeister RD. Self-Presentation Theory: Self-Construction and Audience Pleasing. In: Theories of Group Behavior. New York, NY: Springer; 1987:71-87.
30. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. Proc Natl Acad Sci USA 2015 Jan 12;112(4):1036-1040. [doi: 10.1073/pnas.1418680112]
31. Harris JK, Mart A, Moreland-Russell S, Caburnay CA. Diabetes Topics Associated With Engagement on Twitter. Prev. Chronic Dis 2015 May 07;12:E62. [doi: 10.5888/pcd12.140402]
32. Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and Public Health Communication. PLoS ONE 2016 Sep 15;11(9):e0162765. [doi: 10.1371/journal.pone.0162765]
33. Gabarron E, Dorronzoro E, Rivera-Romero O, Wynn R. Diabetes on Twitter: A Sentiment Analysis. J Diabetes Sci Technol 2018 Nov 19;13(3):439-444. [doi: 10.1177/1932296818811679]
34. Rodriguez S. Facebook has a theory that hiding 'likes' will increase post volume, and Instagram is testing that theory. CNBC. 2019 Dec 08. URL: https://www.cnbc.com/2019/12/06/instagram-hiding-likes-could-increase-post-volume.html [accessed 2020-09-20]

Abbreviations

API: application programming interface

Edited by G Eysenbach; submitted 10.06.20; peer-reviewed by W Kaur, F Sanz; comments to author 17.07.20; revised version received 22.07.20; accepted 02.09.20; published 29.09.20.

Please cite as:
Gabarron E, Larbi D, Dorronzoro E, Hasvold PE, Wynn R, Årsand E
Factors Engaging Users of Diabetes Social Media Channels on Facebook, Twitter, and Instagram: Observational Study
J Med Internet Res 2020;22(9):e21204
URL: http://www.jmir.org/2020/9/e21204/
doi:10.2196/21204
PMID:32990632

Original Paper

Compensatory Social Networking Site Use, Family Support, and Depression Among College Freshman: Three-Wave Panel Study

Mingjie Zhou[1,2], BA, MA, PhD; Fugui Li[1,2], BA, MSc; Yanhong Wang[3], BA, MSc; Shuang Chen[4], BA, PhD; Kexin Wang[5], BA, MSc, PhD

[1]CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

[3]Mental Health Counseling Center, Yang-En University, Quanzhou, China

[4]School of Media, University of Chinese Academy of Social Sciences, Beijing, China

[5]College of Media and International Culture, Zhejiang University, Hangzhou, China

Corresponding Author:
Kexin Wang, BA, MSc, PhD
College of Media and International Culture
Zhejiang University
Main Teaching Building, 4th Fl.
Hangzhou, 310007
China
Phone: 86 18611897400
Fax: 86 10 64872070
Email: wangkexin0809@zju.edu.cn

Abstract

Background: Freshmen were found to use social networking sites (SNS) as a useful medium to effectively adjust to college life, which hints at a tendency to resort to SNS for social compensation. However, the compensatory use of SNS is usually problematic.

Objective: This study explores why a subgroup of freshmen developed depressive symptoms while socially adjusting to college by investigating the antecedent role of introversion, the explanatory role of compensatory use of SNS, and the protective role of perceived family support. The study is among the first to point out the relevance of the compensatory use of SNS in explaining the indirect association between introversion and depression with a longitudinal design.

Methods: A 3-wave panel sample of freshmen (N=1137) is used to examine the moderated mediation model.

Results: We found that introversion at Wave 1 positively predicted compensatory use of SNS at Wave 2 and subsequently increased depression at Wave 3 (unstandardized B=0.07, SE 0.02, *P*<.001, 95% CI 0.04-0.10; unstandardized B=0.09, SE 0.01, *P*<.001, 95% CI 0.06-0.12). The moderated mediation model further examined the buffering role of perceived family support within the link between introversion and compensatory SNS use (index=0.0031, SE 0.0015, 95% CI 0.0003-0.0062). Unexpectedly, we found that family support in Wave 1 decreased compensatory SNS use for less introverted freshmen in Wave 2 and further decreased depression in Wave 3.

Conclusions: Unexpectedly, our findings uncover an enhancing effect, rather than a buffering effect, of family support by embedding its effect within the relationship between introversion and compensatory SNS use. Appreciating the differences in the casual pathways for freshmen with different levels of introversion clarifies how SNS affect young adults' lives.

(J Med Internet Res 2020;22(9):e18458) doi:10.2196/18458

KEYWORDS

freshmen; introversion; compensatory use of SNS; depression; family support; social media

Introduction

Freshmen Use of Social Networking Sites and Depression

Matriculating at a university, a process that marks the transition into emerging adulthood, can be a challenging test for many freshmen to adjust to [1]. Among various psychological symptoms that may appear during maladjustment, depression is given particular notice by researchers. For instance, a recent study discovered that 25.5% of freshmen reported depressive symptoms, with total scores higher than 16 on the Center for Epidemiologic Studies Depression (CES-D) scale; moreover, nearly 6% reported suicidal ideation [2]. In college students, depression is associated with suicidal behavior [3], increased risk of substance use [4], as well as reduced retention [5]. Thus, it is of critical importance to identify risk factors and to unravel influential mechanisms in the adjustment phase.

The use of social networking sites (SNS) is identified as a new potential risk factor for depression, especially among adolescents and emerging adults [6]. Despite the fact that SNS provide a new channel for young people to maintain old connections and build new ones, some studies indicate that the problematic use of SNS may harm adjustment [7]. Compensatory internet use, one of many different kinds of problematic internet uses, is closely related to stressful experiences during the first months of college life. Notably, compensatory internet use is related to compensatory use, which was proposed based on the social compensation hypothesis [8], indicating that people who experience difficulties in offline sociality would end up benefiting from using online interaction platforms (eg, chat rooms). Compensatory internet use, focusing more on a motivational perspective, refers to using a broad range of internet applications as a coping strategy for a negative life situation [9].

More specifically, compensatory internet use is conceptualized by Kardefelt-Winther [9] as a fascination with going online to escape real-life problems or attenuate dysphoric feelings; this coping strategy may cause maladaptive outcomes. For instance, Wang and colleagues [10] found that college students under high stress resort to compensatory use. In addition, Elhai et al [11] indicate that compensatory smartphone use, as one specific type of compensatory internet use, co-occurred with various constructs of depression. Moreover, compensatory internet use can be driven by different life difficulties [9]. Among these difficulties, social problems were found to be predominant in compensatory internet use, leading to the creation of the concept of compensatory SNS use (eg, the use of Facebook) [12]. Essentially, compensatory SNS use specifies using SNS as a means of compensation for personal social inadequacy and thus can be seen as a certain type of compensatory internet use. Following this reasoning, this study sought to investigate the association between freshmen's compensatory SNS use and depression during the stressful transition to university life.

However, it is important to note that the tendency to resort to SNS for social compensation varies among individuals. In general, individuals with low levels of social competence (eg, high in introversion) are more likely to use SNS in a compensatory way. Moreover, various studies have found that the structure of social support is pertinent to the coping strategy [13,14]. Family, along with friends and significant others, is an invaluable social support [15]. According to the social support buffering hypothesis, perceived support has been found to have a buffering effect for negative coping in adverse life situations [16,17]. As such, social support may interact with social traits to impact coping strategy and, in turn, influence mental health. Specifically, we investigated the joint effect of introversion and perceived family support on compensatory SNS use and subsequent depression at a later time among freshmen.

Specifically, the study sought to (1) examine the mediating role of compensatory use of SNS within the association between introversion and depression among freshmen at the very beginning of their adjustment to university life, and (2) test whether family support would protect introverts from resorting to compensatory use and subsequently decrease the risk of developing depression. Notably, the study extends prior research in 3 ways: First, by including media-use risk factors for freshmen, we hope to increase existing knowledge about how freshmen cope with the transition to college. Second, by identifying potential vulnerable groups and protective factors, we present a new, highly targeted intervention for freshmen adjustment. Third, we are the first to use 3-wave data to test our hypotheses on a moderated mediation model for introversion, compensatory SNS use, and depression; this approach responds to a scholarly call for longitudinal research on the antecedents for and outcomes of SNS use.

Introversion and Freshmen's Depressive Symptoms

Scholars believe that core personality is a major concomitant of depression. Introversion (ie, low extroversion), among the Big Five constructs, is especially promising in its association with the phenomenology and the outcome of depression [18]. Various studies have found a positive relationship between introversion and depression. The positive association between introversion and depression might be explained by Eysenck's theory. According to this theory, introverts are more susceptible to punishment [19] or frustrative nonreward [20], which may increase the risk of experiencing a negative mood. In line with this theory, Larsen and Ketelaar [21] found that extroverts reacted more to a positive mood than a negative mood; introverts, however, reacted more to a negative mood than a positive mood. According to Jung's theory [22], introverts are oriented toward internal thoughts and, thus, tend to be ruminative, unsociable, and reserved toward others; this may also explain why more introverted people tend to be more depressed. Empirical evidence from cross-sectional studies supports this positive association. For instance, Saklofske et al [23] found that introversion was positively linked with depression. However, one study using a nationally representative sample found that introversion was related with, but not a significant predictor of, depression [24].

Notably, compared to what was found in cross-sectional studies, the association between introversion and depression in longitudinal studies was weaker. Shull [17] reports a longitudinal association between introversion and depressive symptoms among first-year college students; however, according to a

meta-analysis study, the association between introversion and depression across studies mostly remains observed but seems markedly attenuated when controlling for baseline depression levels.

Taken together, since freshmen's negative moods (ie, a combination of anxiety, tension, depression, anger, confusion, fatigue, and a lack of vigor) increase significantly over time [25], it is likely riskier for more introverted freshmen to develop a depressive mood. Empirical studies found that freshmen's social self-efficacy and self-disclosure were protective predictors of depression [26]; rumination, on the other hand, was one reason why more introverted freshmen became more depressed than extroverts during the transitional phase into university [27]. Therefore, based on Eysenck's theory, Jung's theory, and the longitudinal nature of our study, we assumed that introversion would positively—but weakly—predict depressive freshman mood over time (hypothesis 1).

Compensatory SNS Use as a Mediator

Compensatory SNS use, or the tendency to "go online to escape real-life issues" [9], has been identified as a risk factor for young people's well-being. For instance, Weidman et al [28] suggest that using the internet as an alternative to face-to-face communication reduces well-being. In addition, a few studies indicate a positive association between SNS addiction and depression, which shares an interrelated psychological process with compensatory SNS use [12] and depression [29]. This negative impact of compensatory SNS use on young people's mental health can be explained by Compensatory Internet Use Theory [9], which argues that "the locus of the problem is a reaction by the individual to his negative life situation, facilitated by an internet application." For instance, when young people encounter situations where social stimulation is lacking, they prefer to turn to an internet application (ie, to SNS) that makes socializing accessible instead of making an effort to socialize with the people around them. This escapist coping strategy may have a short-term positive effect in that it can help users get their desired reward of sociability. However, it leads to detrimental effects in the long run: Users may come to rely solely on the internet for socializing [12].

Compensatory SNS use may be particularly salient among freshmen during their college transition. On the one hand, freshmen are unique in that they experience a drastic disruption in their social networks as their families and old friends are, to some extent, out of reach when they attend college. At the same time, they are new to college and have not made significant social ties. On the other hand, the popularity of SNS and smartphones makes it unprecedentedly effortless for freshmen to compensate for sociality online (eg, by turning to an old friend or making new friends online) in contrast to socializing with a potential friend offline. As a result, it is of great importance for scholars to look into this social adjustment phase by focusing on freshmen's compensatory SNS use. However, freshmen vary in their ability to negotiate college life; previous studies have identified several fundamental traits that may contribute to the success or failure of negotiating college life [25].

Because of the centrality of introversion in determining responses to mood, frustration, and social issues, it is conceivable that freshmen with different levels of introversion might react differently to compensatory use and subsequently show diverse levels of vulnerability to depression. In line with this reasoning, Peter et al [30] argue that introverts, who may have difficulty building friendships in person, are more prone to use online contacts as a substitute for an offline social network. Several studies have examined this proposed association between introversion and compensatory use of the internet. For example, Zywiza and Danowski [31] show that a subset of introverted Facebook users strived to make themselves look more popular through online activities. In addition, as suggested above, Goby [32] found that, compared to extroverts, introverts are more likely to use the internet to expand their social network. Therefore, based on social compensation theory and empirical findings, we assumed that compensatory SNS use would mediate the association between introversion and depression (hypothesis 2).

Family Support as a Buffer

Social support was found to moderate the relationship between adverse situations and well-being. More precisely, it can act as a buffer to alleviate the negative influence of stress on depression [33]. This buffer depends on both the size and the structure of the social network [13]. Zimet and colleagues [15] conceptualized the structure of a social network by considering the sources of support, namely, family, friends, and significant others, as distinct subgroups. How and when students receive support from these groups, therefore, becomes important in preventing depression. In addition, Kenny [34] found that the stability and value of family ties positively affect freshmen's social well-being. As noted above, freshmen face the challenge of "leaving home and separating from families and friends" [35]. Moreover, although separation does not necessarily mean being cut-off, low family support was found to be responsible for depression among college students; in particular, it notably interacted with their experiences of stress [36].

Buffering hypothesis indicates that social support can protect one from psychological suffering, but the effects are relatively unimportant for those with low levels of stress [37]. For social adjustment, more introverted freshmen experience higher interpersonal stress [38] and inadequacy and thus might be more prone to use SNS for compensation than less introverted freshmen. Based on the buffering hypothesis, more introverted freshmen tend to benefit from family support than less introverted freshmen, as buffering effects are more effective for those with higher stress. In line with this reasoning, Anschuetz [39] found that social support, including family support, has a buffering effect for highly introverted freshmen to improve their social adjustment.

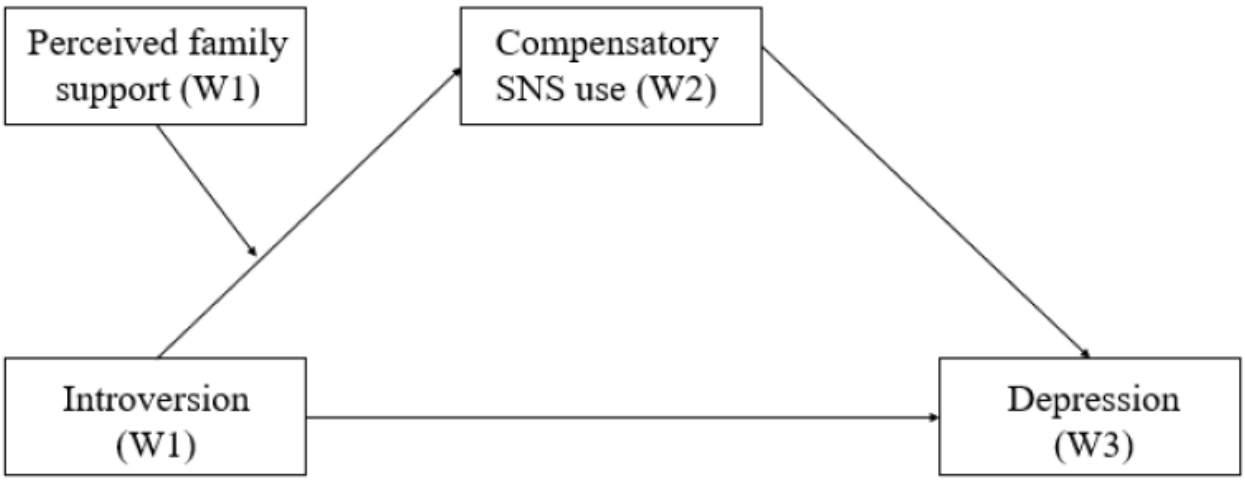
Specifically, social support can help reduce maladaptive coping for vulnerable people [40]. When stressed people perceive adequate instrumental or emotional support, they are less likely to cope with stress adversely. For instance, a recent study reported that social support significantly buffers the association between stress- and coping- motivated alcohol use among college students [41]. As reasoned previously, compensatory

internet use generally, and compensatory SNS use specifically, is not a very healthy coping strategy [16], and introversion is positively related to potentially maladaptive compensatory use. Given that social support can reduce such maladaptive coping, we expected family support to have a moderating effect on the relationship between introversion and compensatory SNS by weakening this link. In other words, more introverted freshmen will be less likely to use SNS for a compensatory motive when they have more family support than when they have less family support.

This moderating effect can be explained by the bidirectional theory, which defines coping as constantly changing cognitive and behavioral efforts to balance internal demands and external support [42]. According to this theory, when an individual's current demands exceed their appraised resources, they will cope in negative ways [43]. Compensatory SNS use is mostly conceptualized as a negative coping strategy [9]. As freshmen negotiating college social life, they would also grapple with both external and internal demands, especially for those high in introversion because they may experience a lower level of social competence than those low in introversion. In this sense, family support is an external resource, which may alleviate an introvert's negative coping strategy by reducing compensatory SNS use. Based on such theories and empirical studies, we theorized that family support would moderate the relationship between introversion and compensatory SNS use and subsequently influence depression; specifically, the relationship between introversion and depression will be weaker when family support is high (hypothesis 3).

This study explores why a subgroup of freshmen developed depressive symptoms while socially adjusting to college by investigating the antecedent role of introversion, the explanatory role of compensatory use of SNS, and the protective role of perceived family support. Figure 1 depicts our proposed model.

Figure 1. Proposed model of the relationship between introversion (W1) and Depression (W3), with compensatory use of SNSs (W2) as a mediator and family support (W1) as a moderator on this mediating mechanism.



Methods

Sample and Procedures

We conducted a 3-wave panel study with an interval of 1 month between each wave, which immediately began after freshmen had registered at a large university (the country has been deleted for peer review) in September 2017. Previous studies of freshmen adjustment have focused on the first 3 months because maladaptive behavior would already be revealed by then [44], and also because the adjustment during this phase can predict further performance.

Trained researchers distributed questionnaires during school hours. Participants were informed that the survey was confidential and that their responses would only be used for research purposes. The researchers obtained informed consent from each of the participants and approval from the host university.

In the first wave, out of 1428 freshmen, 1350 (94.54%) completed the questionnaire. In the second wave, of the 1350 freshmen who had responded in the first wave, 1270 (94.07%) freshmen completed the questionnaire again. In the third wave, of the 1270 freshmen who had participated in the first two waves, 1137 (89.53%) responded to the questionnaire. The final sample consisted of 1137 freshmen, 62.27% (708/1137) of which were women and 37.73% (429/1137) of which were men. The mean age was 18.76 years (SD 0.86). We used a multivariate analysis of variance (MANOVA) to test the potential attribution effects. Using Pillai trace, a MANOVA with all relevant Wave 1 variables showed no significant differences between those who participated at Wave 1 (N=1350) and those who also participated at Wave 3 [N=1137; V=0.01, F(1104)=1.55, P=.12, hp^2 = 0.01].

Measures

Control Variables

Participants reported their age (in years) and gender (1=*male*, 2=*female*). Family income was measured by asking participants to rate their family income level in comparison to the majority of families with whom they were acquainted, using the following scale: *extremely lower than others* (=1), *a lot lower than others* (=2), *a little bit lower than others* (=3), *average* (=4), *a little*

bit higher than others (=5), *a lot higher than others* (=6), and *extremely higher than others* (=7). In addition, participants indicated whether they were an only child (*yes*=1, *no=2*) and whether their family lived in the same city as their university (*yes*=1, *no*=2) in the first wave. We also controlled for baseline depression as a covariate in the moderated mediation model.

Introversion (Wave 1)

We used the subscale of extroversion from the 10-item Big Five Inventory (BFI-10) [45] to construct a measure of introversion. On a 7-point scale *(disagree strongly*=1 to *agree strongly*=7), participants rated the extent to which they agreed with the following descriptions: "I see myself as someone outgoing and sociable" and "I see myself as someone reserved." An average score was calculated after recoding the reversed item. Higher scores represented introversion. The internal consistency was 0.75.

Compensatory SNS Use (Wave 2)

We extracted the subscale of compensatory use of Facebook from the Psycho-Social Aspect of Facebook Use (PSAFU) scale [12]. By replacing "Facebook" with "SNS," we revised the scale to be more general. The resulting 8-item subscale includes items such as "I have more fun socializing on SNS than in real life" and "I find it easier to communicate with people on SNS than in face-to-face, real settings." Participants ranked their agreement with these items from *strongly disagree* (=1) to *strongly agree* (=5). The mean value of these responses was calculated as a new variable. The internal consistency of the scale was 0.80.

Perceived Family Support (Wave 1)

The subscale of the multidimensional scale of perceived social support (MSPSS) developed by Zimet et al [15] was used to measure support from family. Participants rated their agreement with 4 items (eg, "I get emotional help and support from my family" and "my family really tries to help me") from *strongly disagree* (=1) to *strongly agree* (=5). An average score was created to represent perceived family support. The internal consistency of the scale was 0.86.

Depression (Wave 1 and Wave 3)

The patient health questionnaire (PHQ-9) [46]) was used to assess the severity of depression. Participants rated the frequency of described symptoms (eg, "Feeling that you are a failure or have let yourself or your family down" and "having trouble falling or staying asleep, or sleeping too much") from *not at all* (=0) to *nearly every day* (=3). A mean score was calculated as an indicator of depression. The internal consistency of the scale was 0.81 in Wave 1 and 0.84 in Wave 3.

Analytical Strategy

We used the PROCESS macro (model 7) for SPSS (version 21.0, IBM Corp) with bootstrapping (95% CI, 1000 samples) to analyze the data [47]. The model included the following control variables from Wave 1: gender, age, siblings, relative family income, family location, and baseline depression. Table 1 presents the mean, standard deviation, and zero-order correlations among variables.

Table 1. Descriptive statistics (N=1137).

Variables[a]	Mean	SD	1	2	3	4	5	6	7	8	9
1.Gender (W1[b])	—[c]	—	1								
2. Only child (W1[b])	—	—	.23[d]	1							
3. Family location (W1[b])	—	—	-.03	.01	1						
4. Relative family income (W1[b])	3.57	0.80	.03	-.07[e]	-.05	1					
5. Age (W1[b])	18.76	0.86	-.03	.13[c]	-.02	-.09[c]	1				
6. Depression (W1[b])	0.62	0.39	.07[d]	.06[d]	-.03	-.08[c]	.02	1			
7. Introversion (W1[b])	3.59	1.37	.03	.05	-.07[d]	-.05	-.04	.23[c]	1		
8. Perceived family support (W1[b])	5.29	1.20	.02	-.06[d]	.01	.12[c]	.03	-.28[c]	-.14[c]	1	
9. Compensatory SNS[f] use (W2[g])	2.61	0.73	-.17[c]	-.08[c]	-.02	-.02	-.04	.21[c]	.16[c]	-.17[c]	1
10. Depression (W3[h])	0.68	0.39	.02	.05	.01	-.05	-.04	.51[c]	.11[c]	-.25[c]	.24[c]

[a] Numbered row headings correspond with numbered column headings.

[b] W1: Wave 1.

[c]—: Not applicable.

[d]*P*<.01.

[e]*P*<.05.

[f]SNS: social networking sites.

[g]W2: Wave 2.

[h]W3: Wave 3.

Results

Introversion and Depression (Hypothesis 1)

Hypothesis 1 predicted a positive association between introversion (Wave 1) and depression (Wave 3) across time. In contrast with the prediction of hypothesis 1, the results showed that introversion in the first wave was not significantly associated with freshmen's depression in the third wave. Thus, hypothesis 1 was not supported.

Mediating Role of Compensatory SNS Use (Hypothesis 2)

Hypothesis 2 predicted that compensatory SNS use (Wave 2) would mediate the association between introversion (Wave 1) and depression (Wave 3). More specifically, hypothesis 2 predicted that introverted freshmen would resort to compensatory SNS use more frequently and thus have a higher risk of developing depression compared with extroverted freshmen who may not as frequently resort to compensatory use. In line with hypothesis 2, the results revealed that the association between introversion in the first wave and depression in the third wave was mediated by compensatory SNS use in the second wave. This indicated that compensatory SNS use in the first months explained why freshmen with different baseline levels of introversion gradually developed different levels of depression in the following 2 months. Therefore, hypothesis 2 was supported.

Perceived Family Support as a Moderator in the Mediating Model (Hypothesis 3)

Hypothesis 3 further posited that perceived family support (Wave 1) would moderate the association between introversion (Wave 1) and compensatory SNS use (Wave 2). More specifically, hypothesis 3 aimed to test whether family support weakens the positive link between introversion and compensatory SNS use by decreasing more introverted freshmen's SNS use.

Roughly in line with hypothesis 3, the moderated mediation test revealed that perceived family support (Wave 1) was a significant moderator of the relationship between introversion (Wave 1) and compensatory SNS use (Wave 2) (index=0.003, SE 0.001, 95% CI 0.0003-0.0062). To interpret the interaction effect, 3 simple slope tests were conducted. When family support was low (ie, 1 SD below the mean), the influence of introversion on compensatory SNS use was not significant (unstandardized *B*=0.003, SE 0.002, , *P*=.18, 95% CI 0.001-0.007); When the family support was medium, the influence of introversion on compensatory SNS use was significantly positive (*B*=0.007, SE 0. 002, *P*<.001, 95% CI 0.004-0.010). When the family support was high (ie, 1 SD above the mean), the influence of introversion on compensatory SNS use was also significantly positive (*B* =0.010, SE 0.002, *P*<.001, 95% CI 0.006-0.014). Thus, family support can alleviate the impact of introversion on compensatory SNS use only when family support is at medium and higher levels. However, in contrast to the prediction of hypothesis 3, the beneficiaries of family support are less introverted freshmen rather than more introverted freshmen. As depicted in Figure 2, the protective effect of family support on the association between compensatory SNS use (Wave 1) and depression (Wave 3) was stronger for less introverted freshman than for those who were more introverted. As such, hypothesis 3 was not supported.

Table 2, Table 3, and Figure 3 show the results of the moderated mediation analysis. We also conducted the regression analysis with baseline depression as the only control variable. Hypothesis 1 was also not supported (B=-0.01, SE .01, *P*=.25, 95% CI -0.02-0.01), suggesting that introversion at Wave 1 was not a significant predictor of depression at Wave 3. Hypothesis 2 was supported (B=0.09, SE 0.01, *P*<.001, 95% CI 0.06-0.11), suggesting compensatory SNS use at Wave 2 mediated the relationship between introversion at Wave 1 and depression at Wave 3. Moreover, the moderated mediation analysis was also supported (index=0.003, SE 0.001, 95% CI 0.0004-0.0063), indicating that perceived family support at Wave 1 moderated the relationship between introversion at Wave 1 and compensatory SNS use at Wave 2. However, in contrast to hypothesis 3, the results revealed that as perceived family support at Wave 1 increased, a decrease in introversion in Wave 1 was related to a decrease in depression in Wave 3. These focal results were inconsistent with the results reported in Table 2.

Figure 2. The moderating effect of perceived family support (W1) in the relationship between introversion (W1) and compensatory use of social networking sites (SNS; W2).

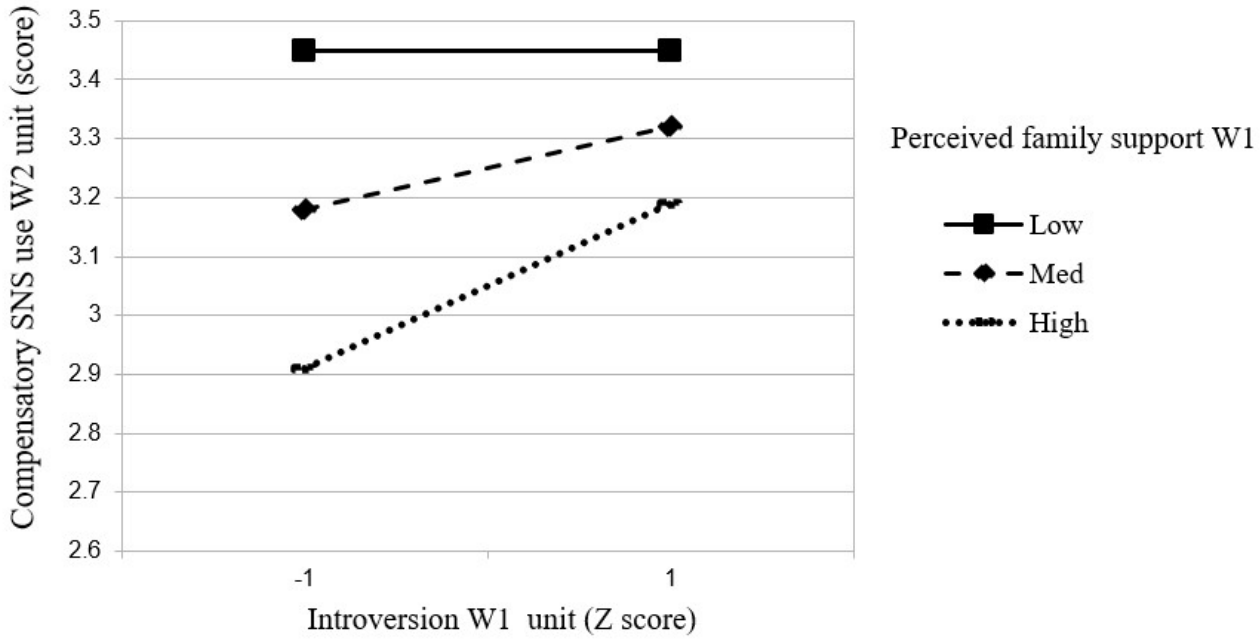


Table 2. Results of moderated mediation analysis (N=1137).

Predictors	Compensatory SNS[a] use (W2[b])				Depression (W3[c])			
	B[d]	SE	LLCI[e]	ULCI[f]	*B*[d]	SE	LLCI[e]	ULCI[f]
Constant	3.25[g]	0.49	2.28	4.22	0.50	0.24	0.03	0.98
Gender (W1[h])	-0.27[g]	0.04	-0.36	-0.18	0.01	0.02	-0.03	0.05
Age (W1[h])	-0.02	0.02	-0.07	0.02	-0.02	0.01	-0.04	0.00
Only child (W1[h])	-0.10[j]	0.05	-0.19	-0.01	0.03	0.02	-0.01	0.07
Relative income (W1[h])	0.01	0.03	-0.04	0.06	0.01	0.01	-0.01	0.05
Family location (W1[h])	-0.01	0.03	-0.07	0.05	0.02	0.01	-0.01	0.05
Depression (W1[h])	0.34[g]	0.06	0.23	0.45	0.48[g]	0.03	0.43	0.54
Introversion (W1[h])	0.07[g]	0.02	0.04	0.10	-0.01	0.01	-0.02	0.01
Perceived family support (W1[h])	-0.20[g]	0.06	-0.32	-0.09	N/A[i]	N/A	N/A	N/A
Introversion (W1[h]) * Family support (W1[h])	0.03[j]	0.01	0.01	0.06	N/A	N/A	N/A	N/A
Compensatory SNS use (W2[b])	N/A	N/A	N/A	N/A	0.09[e]	0.01	0.06	0.12
R^2	0.11				0.28			
ΔR^2	0.11[g]				0.17[g]			

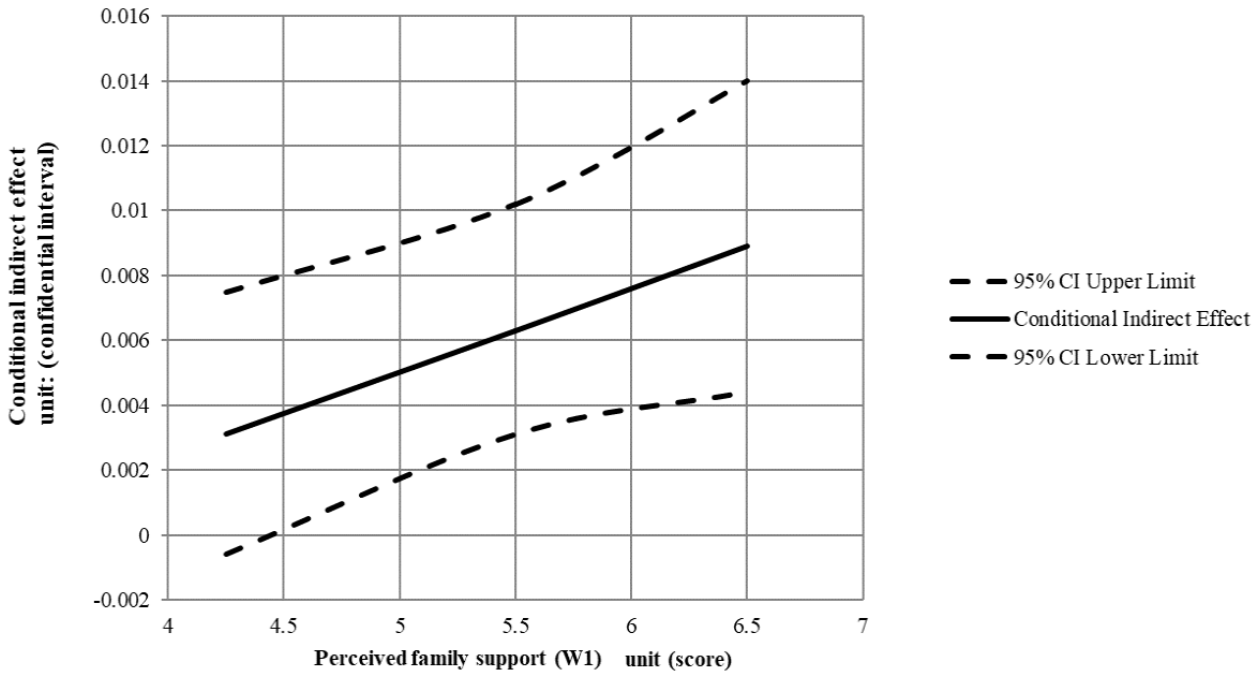
[a]SNS: social networking sites.
[b]W2: Wave 2.
[c]W3: Wave 3.
[d]*B*: unstandardisd coefficient.
[e]LLCI: lower level confidential interval.
[f]ULCI: upper level confidential interval.
[g]*P*<.001.
[h]W1: Wave 1.
[i]N/A: not applicable.
[j]*P*<.05.

Table 3. Conditional indirect effects of introversion (Wave 1) on depression (Wave 3) at the value of perceived family support (Wave 1) through compensatory social networking sites use (Wave 2).

Family support	Effect	Boot SE	Boot LLCI[a]	Boot ULCI[b]
-1 SD	0.0031	0.0021	-0.0008	0.0077
Mean	0.0063	0.0018	0.0030	0.0103
+1 SD	0.0089	0.0025	0.0044	0.0139

[a]LLCI: lower level confidential interval.
[b]ULCI: upper level confidential interval.

Figure 3. The conditional indirect effect of introversion (W1) on depression (W3) at values of the perceived family support (W1) through compensatory use of social networking sites (SNS; W2).



Discussion

Principal Findings

Previous studies have documented that freshmen vary in their capacity to negotiate college social life [48] and that depression rises significantly during this adjustment phase [49]. This longitudinal study sought to increase our knowledge of the association between introversion and depression by investigating the explanatory role of compensatory SNS use and the protective role of perceived family support during the initial phase of transition. We found that introversion at Wave 1 positively predicted compensatory use of SNS at Wave 2 and subsequently increased depression at Wave 3. The moderated mediation model further examined the buffering role of perceived family support within the link between introversion and compensatory SNS use. Unexpectedly, we found that family support in Wave 1 decreased compensatory SNS use for less introverted freshmen in Wave 2 and further decreased depression in Wave 3. This finding indicated that family support served as an enhancing role for freshmen with low introversion rather than a buffering role for freshmen with high introversion.

Hypothesized Model

In contrast to the prediction of hypothesis 1, we found that introversion was not significantly associated with depression across time. This finding differs from previous research that found support for a positive association between introversion and depression among freshmen [27]. However, our finding is in line with a recent study that also found no significant association between introversion and depression among freshmen in nursing [50]. In addition, another study based on a national representative sample also found that introversion was related to, but not a significant predictor of, depression. More specifically, "introversion may be better viewed as reflective of shared variance with depression" [24]. Furthermore, Cheng and Furnham [51] argued that although introversion was often a direct predictor of happiness, it was never a direct predictor of depression. Together with the finding in hypothesis 2, we provide more evidence on the indirect association between introversion and depression.

In line with hypothesis 2, we found that compensatory SNS use was a significant mediator within the association between introversion and depression, suggesting that compensatory use can explain how freshmen with different levels of introversion developed different levels of depression in their first 3 months of university (after controlling the baseline depression). The positive association between introversion and compensatory SNS use can be explained by the social compensation hypothesis, which posits that individuals with difficulty managing social life are more likely to use online interaction as a substitute [52]. The positive association between compensatory SNS use and depression can be explained by research on escapism. Escapism is a coping strategy that can aggravate current and future depression because it exacerbates people's vulnerability to stressful events by making them feel increasingly helpless, inadequate, and nervous [53]. It is possible that freshmen who use SNS to compensate for socializing become more depressed because they feel increasingly inadequate in sociality. Taken together, the confirmed mediation pathway provides a possible causal mechanism for understanding, from a media-use perspective, how a subset of freshmen become more depressed.

Partially in line with hypothesis 3, we found that perceived family support (Wave 1) was indeed a moderator within the association between introversion (Wave 1) and compensatory

SNS use (Wave 2), indicating that family support interacted with introversion to affect freshmen's compensatory SNS use. However, this moderating effect contradicted our hypothesis for a buffering effect. As shown in Figure 2, under medium and high family support conditions, compensatory SNS use increased with introversion but did not exceed the low family support condition. However, when family support is relatively adequate, people low in introversion reported less compensatory SNS use. As people low in introversion are more capable of coping with stressful life events and thus have better mental health, we would argue that family support serves as an enhancing role for people who are low in introversion.

The moderated mediation effect may be explained by differences in orientation preferences between more introverted people and less introverted people. Less introverted people are interpersonally oriented while more introverted people have intrapersonal orientations; this suggests that when external family support is available, people differing in introversion vary in gaining the support. Specifically, less introverted people are more likely to benefit from social support. For instance, Kushwaha [54] found that extroverts seek significantly more guidance and support from others when coping with illness. In addition, Zell et al [55] suggested that extroverts have higher levels of intimacy and trust in their offline social ties than introverts, which may translate into greater benefits and more effective support when needed. Accordingly, when family support is relatively abundant, less introverted freshmen are less likely to perceive themselves as socially inadequate and are thus less likely to use SNS as compensation than more introverted freshmen. However, when external family support is low, freshmen show no difference in using SNS for social compensation, as it seems available and effortless. This enhancing effect was also found in resilience. Although resilience has been frequently theorized as a buffer for more vulnerable groups when facing adverse events [56], one study found that resilience further facilitates well-being among less vulnerable people [57]. Future study may be needed to investigate the boundary condition and the mechanism for both buffering and enhancing effects.

Implications

This study makes 3 key contributions that may guide future research. First, the results detected compensatory SNS use as a risk factor for depression among freshmen during their adjustment to college—this conflicts with the findings of a previous study that theorizes compensatory use as only the consequence, not the cause, of other psychological problems [9]. Therefore, the results of this study suggest a need for further investigation into the negative effect of compensatory SNS use in other stressful life events among other subgroups. More specifically, it may be highly relevant to uncover why and how certain types of SNS use meant to make people feel better end up being harmful. Second, the results supported the hypothesized mediating role of the compensatory use of SNS within the association between introversion and depression, which clarifies how core personality (introversion) predicts well-being (depression). Previous studies posit that extroversion is a direct predictor of happiness but not depression [51]. Our findings may thus increase understandings of the mechanism by which introversion may indirectly predict depression. Third, our findings reveal that family support has an enhancing effect for freshmen with low introversion to avoid using SNS out of a compensatory motive in the first 3 months of adjustment; this indicates the need for further investigation into how family support functions for vulnerable youth using negative coping strategies, especially during key developmental phases.

However, some questions still remain unanswered. First, if family support plays a role in preventing compensatory SNS use and thus decreases depression over time, it may be highly relevant to investigate whether other sources of social supports (eg, friendship) also reduce maladaptive SNS use and, in turn, decrease depression. In addition, it remains unclear why family support did not prevent depression in more introverted freshman. It could be that such freshman resorted to compensatory use for self-affirmation or recognition from sources other than social supports; that is, social contact itself may not have satisfied this group's inner need for validation. We hope this study can spark future researchers to take up these questions.

Notably, in practical terms, the study revealed one possible mechanism through which freshmen become more depressed at the very beginning of their enrollment: More introverted freshmen are likely to resort to compensatory SNS use, which increases depression over time. When family support is relatively sufficient, freshmen low in introversion reported less compensatory SNS use and lower levels of depression than those high in introversion. This may help university administrators and families gain insight into prediction and targeted intervention for vulnerable freshmen.

Limitations

When interpreting our findings, 2 limitations have to be considered. First, the measurement of compensatory SNS use adopted in this study does not differentiate between various kinds of SNS. However, the type of SNS may influence the psychological outcomes of compensatory use. For instance, WeChat, the most prevalent SNS in China, is a relatively closed platform on which freshmen have more acquaintances or even intimate contacts than in other platforms; in contrast, QQ, which is especially popular among young adults, involves a larger percentage of strangers and superficial contacts. Consequently, seeking social compensation through WeChat may not aggravate depressive symptoms as much as seeking compensation through QQ. Further research should consider the differences between these platforms, perhaps by applying a more nuanced measurement of compensatory use.

Second, although our findings provide insight into the potential casual mechanism of how freshmen become more depressed by investigating the explanatory value of compensatory SNS use, the study remained focused on media and sociality. Other moderators (eg, perceived friend supports) and mediators (eg, self-reaffirmation) should be examined in future studies to provide a more comprehensive picture of how freshmen adjust to academic achievement and other aspects of college [58].

Conclusion

In today's highly mediatized environment, young adults have access to an ever-accumulating set of SNS. Socially vulnerable

freshmen may use online contacts as a substitute for face-to-face interactions, especially when their transition to college disrupts their social networks. The current study is among the first to point out the relevance of the compensatory use of SNS in explaining the indirect association between introversion and depression with a longitudinal design. Our findings uncover an enhancing effect of family support by embedding its effect within the relationship between introversion and compensatory SNS use. Together, appreciating the differences in the casual pathways for freshmen with different levels of introversion clarifies how SNS affect young adults' lives.

Conflicts of Interest

None declared.

References

1. Dyson R, Renk K. Freshmen adaptation to university life: depressive symptoms, stress, and coping. J Clin Psychol 2006 Oct;62(10):1231-1244. [doi: 10.1002/jclp.20295] [Medline: 16810671]
2. Supartini A, Honda T. The impact of sleep timing, sleep duration, and sleep quality on depressive symptoms and suicidal ideation amongst Japanese freshmen: the EQUSITE study. Sleep disorders 2016(3):1-10. [doi: 10.1155/2016/8737654]
3. Furr SR, Westefeld JS, McConnell GN, Jenkins JM. Suicide and depression among college students: A decade later. Professional Psychology: Research and Practice 2001;32(1):97-100. [doi: 10.1037/0735-7028.32.1.97]
4. Pedersen DE. Gender differences in college binge drinking: Examining the role of depression and school stress. The Social Science Journal 2019 Dec 09;50(4):521-529. [doi: 10.1016/j.soscij.2013.03.003]
5. Eisenberg D, Golberstein E, Hunt J. Mental health and academic success in college. B E J Econom Anal Policy 2009;9(1):1935-1682. [doi: 10.2202/1935-1682.2191]
6. Kross E, Verduyn P, Demiralp E, Park J, Lee DS, Lin N, et al. Facebook use predicts declines in subjective well-being in young adults. PLoS One 2013 Aug;8(8):e69841 [FREE Full text] [doi: 10.1371/journal.pone.0069841] [Medline: 23967061]
7. Wohn DY, LaRose R. Effects of loneliness and differential usage of Facebook on college adjustment of first-year students. Computers & Education 2014 Jul;76:158-167. [doi: 10.1016/j.compedu.2014.03.018]
8. Valkenburg PM, Schouten AP, Peter J. Adolescents' identity experiments on the internet. New Media & Society 2016 Jun 30;7(3):383-402. [doi: 10.1177/1461444805052282]
9. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. Computers in Human Behavior 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
10. Wang J, Wang H, Gaskin J, Wang L. The role of stress and motivation in problematic smartphone use among college students. Computers in Human Behavior 2015 Dec;53:181-188. [doi: 10.1016/j.chb.2015.07.005]
11. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. Journal of Affective Disorders 2017 Jan;207:251-259. [doi: 10.1016/j.jad.2016.08.030]
12. Bodroža B, Jovanović T. Validation of the new scale for measuring behaviors of Facebook users: Psycho-Social Aspects of Facebook Use (PSAFU). Computers in Human Behavior 2016 Jan;54:425-435. [doi: 10.1016/j.chb.2015.07.032]
13. Rosenquist JN, Fowler JH, Christakis NA. Social network determinants of depression. Mol Psychiatry 2011 Mar;16(3):273-281 [FREE Full text] [doi: 10.1038/mp.2010.13] [Medline: 20231839]
14. Thoits PA. Social support as coping assistance. Journal of Consulting and Clinical Psychology 1986;54(4):416-423 [FREE Full text] [doi: 10.1037//0022-006x.54.4.416]
15. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The Multidimensional Scale of Perceived Social Support. Journal of Personality Assessment 1988 Mar;52(1):30-41. [doi: 10.1207/s15327752jpa5201_2]
16. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychological Bulletin 1985;98(2):310-357. [doi: 10.1037/0033-2909.98.2.310]
17. Manne S, Ostroff J, Sherman M, Glassman M, Ross S, Goldstein L, et al. Buffering Effects of Family and Friend Support on Associations between Partner Unsupportive Behaviors and Coping among Women with Breast Cancer. Journal of Social and Personal Relationships 2016 Jun 29;20(6):771-792. [doi: 10.1177/0265407503206004]
18. Janowsky DS. Introversion and extroversion: Implications for depression and suicidality. Curr Psychiatry Rep 2001 Nov;3(6):444-450. [doi: 10.1007/s11920-001-0037-7]
19. Eysenck H. The biological basis of personality. Charles C. Thomas.MZ(1990) Biological dimensions of personality. isbn 1990:0898624304. [doi: 10.4324/9781351305280]
20. Gray J. The psychophysiological nature of introversion-extraversion: A modification of Eysenck's theory. In: Biological bases of individual behavior. San Diego: CA: Academic Press; 1972:182-205.
21. Larsen RJ, Ketelaar T. Extraversion, neuroticism and susceptibility to positive and negative mood induction procedures. Personality and Individual Differences 1989 Jan;10(12):1221-1228. [doi: 10.1016/0191-8869(89)90233-x]
22. Jung C. The collected works of CG Jung, Vol 6: Psychological types. Princeton, NJ (originally published 1921): Princeton University Press; 1971:0691097704.

23. Saklofske D, Kelly I, Janzen B. Neuroticism, depression, and depression proneness. Personality and Individual Differences 1995 Jan;18(1):27-31. [doi: 10.1016/0191-8869(94)00128-f]
24. Cox BJ, McWilliams LA, Enns MW, Clara IP. Broad and specific personality dimensions associated with major depression in a nationally representative sample. Comprehensive Psychiatry 2004 Jul;45(4):246-253. [doi: 10.1016/j.comppsych.2004.03.002]
25. Pritchard ME, Wilson GS, Yamnitz B. What Predicts Adjustment Among College Students? A Longitudinal Panel Study. Journal of American College Health 2007 Jan;56(1):15-22. [doi: 10.3200/jach.56.1.15-22]
26. Wei M, Russell DW, Zakalik RA. Adult Attachment, Social Self-Efficacy, Self-Disclosure, Loneliness, and Subsequent Depression for Freshman College Students: A Longitudinal Study. Journal of Counseling Psychology 2005 Oct;52(4):602-614. [doi: 10.1037/0022-0167.52.4.602]
27. Shull A. Rumination Mediates the Impact of Personality on the Development of Depression During the Transition to College Doctoral dissertation, University of Michigan. 2014. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/107778/ashull.pdf?sequence=1 [accessed 2019-08-30]
28. Weidman AC, Fernandez KC, Levinson CA, Augustine AA, Larsen RJ, Rodebaugh TL. Compensatory internet use among individuals higher in social anxiety and its implications for well-being. Personality and Individual Differences 2012 Aug;53(3):191-195. [doi: 10.1016/j.paid.2012.03.003]
29. Banjanin N, Banjanin N, Dimitrijevic I, Pantic I. Relationship between internet use and depression: Focus on physiological mood oscillations, social networking and online addictive behavior. Computers in Human Behavior 2015 Feb;43:308-312. [doi: 10.1016/j.chb.2014.11.013]
30. Peter J, Valkenburg PM, Schouten AP. Developing a Model of Adolescent Friendship Formation on the Internet. CyberPsychology & Behavior 2005 Oct;8(5):423-430. [doi: 10.1089/cpb.2005.8.423]
31. Zywica J, Danowski J. The faces of Facebookers: Investigating social enhancement and social compensation hypotheses; predicting Facebook? and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with semantic networks. J Comput Mediat Commun 2008;14(1):1-34. [doi: 10.1111/j.1083-6101.2008.01429.x]
32. Goby VP. Personality and Online/Offline Choices: MBTI Profiles and Favored Communication Modes in a Singapore Study. CyberPsychology & Behavior 2006 Feb;9(1):5-13. [doi: 10.1089/cpb.2006.9.5]
33. Wang X, Cai L, Qian J, Peng J. Social support moderates stress effects on depression. Int J Ment Health Syst 2014;8(1):41. [doi: 10.1186/1752-4458-8-41]
34. Kenny M. Family ties and leaving home for college: Recent findings and implications. Journal of College Student Personnel 1987;28(5):438-442.
35. Margolis G. Moving away: Perspectives on counseling anxious freshmen. Adolescence 1981;16(63):633-640.
36. Rich A, Bonner R. Concurrent validity of a stress-vulnerability model of suicidal ideation and behavior: a follow-up study. Suicide Life Threat Behav 1987;17(4):a. [doi: 10.1111/j.1943-278x.1987.tb00067.x]
37. Cohen S, Hoberman HM. Positive Events and Social Supports as Buffers of Life Change Stress1. J Appl Social Pyschol 1983 Apr;13(2):99-125. [doi: 10.1111/j.1559-1816.1983.tb02325.x]
38. Klimstra TA, Noftle EE, Luyckx K, Goossens L, Robins RW. Personality development and adjustment in college: A multifaceted, cross-national view. Journal of Personality and Social Psychology 2018 Aug;115(2):338-361. [doi: 10.1037/pspp0000205]
39. Anschuetz H. Coping with college transition: The effects of trait vulnerability and social support. URL: https://search.proquest.com/docview/305383309?pq-origsite=gscholar&fromopenview=true [accessed 2019-08-30]
40. Thoits PA. Mechanisms Linking Social Ties and Support to Physical and Mental Health. J Health Soc Behav 2011 Jun 14;52(2):145-161. [doi: 10.1177/0022146510395592]
41. Pittman DM, Quayson AA, Rush CR, Minges ML. Revisiting resilience: Examining the relationships between stress, social support, and drinking behavior among black college students with parental substance use disorder histories. Journal of Ethnicity in Substance Abuse 2019 Dec 26:1-22. [doi: 10.1080/15332640.2019.1707142]
42. Lazarus R, Folkman S. Stress, appraisal, and coping. New York: Springer; 1984:0826141927.
43. Castellani B, Wedgeworth R, Wootton E, Rugle L. A bi-directional theory of addiction: Examining coping and the factors related to substance relapse. Addictive Behaviors 1997 Jan;22(1):139-144. [doi: 10.1016/s0306-4603(96)00026-3]
44. Martin KA, Leary MR. Self-presentational determinants of health risk behavior among college freshmen. Psychology & Health 2007 Dec 19;16(1):17-27. [doi: 10.1080/08870440108405487]
45. Rammstedt B, John OP. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 2007 Feb;41(1):203-212. [doi: 10.1016/j.jrp.2006.02.001]
46. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
47. Hayes A. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York: The Guildford Press; 2018:A.
48. Yang C, Lee Y. Interactants and activities on Facebook, Instagram, and Twitter: Associations between social media use and social adjustment to college. Applied Developmental Science 2018 Apr 04;24(1):62-78. [doi: 10.1080/10888691.2018.1440233]

49. Kötter T, Tautphäus Y, Obst KU, Voltmer E, Scherer M. Health-promoting factors in the freshman year of medical school: a longitudinal study. Med Educ 2016 May 11;50(6):646-656. [doi: 10.1111/medu.12987]
50. Iorga M, Muraru D, Soponaru C, Petrariu F. Factors influencing the level of depression among freshman nursing students. The Medical-Surgical Journal 2017;121(4):779-786.
51. Cheng H, Furnham A. Personality, self-esteem, and demographic predictions of happiness and depression. Personality and Individual Differences 2003 Apr;34(6):921-942. [doi: 10.1016/s0191-8869(02)00078-8]
52. Valkenburg P, Peter J. Online communication and adolescent well-being: Testing the stimulation versus the displacement hypothesis. J Comput Mediat Commun 2007;12(4):1169-1182. [doi: 10.1111/j.1083-6101.2007.00368.x]
53. Rohde P, Lewinsohn PM, Tilson M, Seeley JR. Dimensionality of coping and its relation to depression. Journal of Personality and Social Psychology 1990;58(3):499-511. [doi: 10.1037/0022-3514.58.3.499]
54. Govind SK. A study of seeking guidance and support coping strategy of cancer patients. Int. J. Psychol. Couns 2014 May 31;6(5):66-73. [doi: 10.5897/ijpc2013.0199]
55. Zell D, McGrath C, Vance C. Examining the interaction of extroversion and network structure in the formation of effective informal support networks. Journal of Behavioral and Applied Management 2014;15(2):59-81.
56. Manning LK, Carr DC, Kail BL. Do Higher Levels of Resilience Buffer the Deleterious Impact of Chronic Illness on Disability in Later Life? GERONT 2014 Jul 25;56(3):514-524. [doi: 10.1093/geront/gnu068]
57. Zhang J, Zhang J, Zhou M, Yu NX. Neighborhood Characteristics and Older Adults' Well-Being: The Roles of Sense of Community and Personal Resilience. Soc Indic Res 2017 Apr 12;137(3):949-963. [doi: 10.1007/s11205-017-1626-0]
58. Baker RW, Siryk B. Measuring adjustment to college. Journal of Counseling Psychology 1984;31(2):179-189. [doi: 10.1037/0022-0167.31.2.179]

Abbreviations

MANOVA: multivariate analysis of variance
SNS: social networking sites

Edited by G Eysenbach; submitted 27.02.20; peer-reviewed by B Bodroža, Q Gao; comments to author 06.04.20; revised version received 27.05.20; accepted 11.08.20; published 02.09.20.

Please cite as:
Zhou M, Li F, Wang Y, Chen S, Wang K
Compensatory Social Networking Site Use, Family Support, and Depression Among College Freshman: Three-Wave Panel Study
J Med Internet Res 2020;22(9):e18458
URL: https://www.jmir.org/2020/9/e18458
doi:10.2196/18458
PMID:32795999

Original Paper

Real-World Evidence on the Effect of Missing an Oral Contraceptive Dose: Analysis of Internet Search Engine Queries

Irit Hochberg[1,2], MD, PhD; Sharon Orshalimy[3], MD; Elad Yom-Tov[4,5], PhD

[1]Institute of Endocrinology, Diabetes and Metabolism, Rambam Health Care Campus, Haifa, Israel

[2]Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel

[3]School of Public Health, Ben Gurion University, Beer Sheva, Israel

[4]Microsoft Research, Herzeliya, Israel

[5]Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
13 Shenkar st
Herzeliya, 46733
Israel
Phone: 972 747111359
Email: eladyt@microsoft.com

Abstract

Background: Oral contraceptives (OCs) are a unique chronic medication with which a memory slip may result in a threat that could change a person's life course. Subjective concerns of missed OC doses among women have been addressed infrequently. Anonymized queries to internet search engines provide unique access to concerns and information gaps faced by a large number of internet users.

Objective: We aimed to quantitate the frequency of queries by women seeking information in an internet search engine, after missing one or more doses of an OC; their further queries on emergency contraception, abortion, and miscarriage; and their rate of reporting a pregnancy timed to the cycle of missing an OC.

Methods: We extracted all English-language queries submitted to Bing in the United States during 2018, which mentioned a missed OC and subsequent queries of the same users on miscarriage, abortion, emergency contraceptives, and week of pregnancy.

Results: We identified 26,395 Bing users in the United States who queried about missing OC pills and the fraction that further queried about miscarriage, abortion, emergency contraceptive, and week of pregnancy. Users under the age of 30 years who asked about forgetting an OC dose were more likely to ask about abortion (1.5 times) and emergency contraception (1.7 times) (P<.001 for both), while users at ages of 30-34 years were more likely to query about pregnancy (2.1 times) and miscarriage (5.4 times) (P<.001 for both).

Conclusions: Our data indicate that many women missing a dose of OC might not have received sufficient information from their health care providers or chose to obtain it online. Queries about abortion and miscarriage peaking in the subsequent days indicate a common worry of possible pregnancy. These results reinforce the importance of providing comprehensive written information on missed pills when prescribing an OC.

(J Med Internet Res 2020;22(9):e20632) doi:10.2196/20632

KEYWORDS

search engines; birth control; abortion; miscarriage

Introduction

The World Health Organization (WHO) recognized contraceptives and family planning methods as key elements for fulfilling human rights and promoting women's autonomy and wellbeing [1]. Therefore, a wide range of contraception options are included in their Model List of Essential Medicines. Contraceptive access, choice, and knowledge as well as correct contraceptive use are key elements in increasing compliance to and efficacy of contraception and decreasing the number of unintended pregnancies [2,3].

A main component of the WHO guidelines for contraceptive use refers to the quality of contraceptive counseling and delivering correct information before choosing a method of contraception [4]. Misuse of contraceptives and discontinuation are associated with a lack of knowledge about contraceptives, individual lifestyles, social structure, age, education, class, ethnicity, and race [5-8]. Oral contraceptives (OCs) are the most commonly used contraception in the United States [9], and optimal use over a year prevents 99.5% pregnancies, while nonoptimal use has a failure rate of >7% over a year [10]. Age is an important demographic characteristic for OC misuse, as young women and adolescents report forgetting a pill more often than others [5,11-13] and have a much higher rate of failure of oral contraception compared to older women [14]. Multiple studies have shown that approximately 1 of 2 young users misses 1 or more OC pills each month [15-18]. It is important to give clear instructions on the appropriate measures to prevent pregnancy after a missed dose while providing contraception counseling to women, according to their lifestyles, sexual partner status, and health condition [5,19].

OCs are a unique chronic medication where a slip of memory may be experienced as a threat that could change the life course of a person. Subjective experiences and concerns in women missing doses of OCs have been addressed infrequently, but there is evidence that this event leads to stress and affects women's well-being and ability to function at work [20].

Past research shows that young people frequently prefer to search for sexual health advice online, over turning to a health care provider they might have seen in person months before, due to the accessibility and privacy afforded by online search [21,22]. Anonymized queries to internet search engines provide a unique access to the incidence of concerns and information gaps in a large number of internet users [23-25]. Further queries by the same users can shed light on their attitudes, behaviors, and health consequences in the period following the query [26]. In this study, we aimed to quantitate the frequency of queries by women seeking information in an internet search engine after missing 1 or more doses of an OC and their further queries on emergency contraception, abortion, and pregnancy.

Methods

We extracted all English-language search queries submitted to Bing by users in the United States during 2018. Bing's market share in the United States was estimated at 25% [27]. It is estimated that Bing users are a representative sample of the US population [28]. Data for each search query included an anonymous user identifier, time and date of the search, and search text. User gender and age groups, reported when users registered to Bing, were available for a subset of the users.

The text of the queries was used to filter the queries into 1 of 5 classes (Textbox 1).

Using the pregnancy queries, we calculated the first day of the last menstrual period (LMP) for women who reported pregnancy by subtracting the number of weeks reported in their queries from the date of the query.

The likelihood of pregnancy was calculated as the percentage of women who queried for a missed OC dose and later queried for a week of pregnancy. The empirical likelihood to query about each week of pregnancy was calculated as the fraction of queries, which mentioned a specific week of pregnancy. To compensate for the finite data period, each user who reported a missed dose was assigned a weight relative to the likelihood for querying about pregnancy, relative to the date of reporting about the missed dose. For example, if a user asked about a missed dose in January, they would be given a weight of 1 since the full term of pregnancy was within the data period. Conversely, a user who queried during December was given a low weight, since they could only query for the first few weeks of pregnancy.

Statistical analysis was conducted using MATLAB 9.7 with the statistical toolbox version 11.6. This study was approved by the Behavioral Sciences Research Ethics Committee of the Technion.

Textbox 1. Classes of search queries of users.

1. Missed OC queries: queries which contained the words "miss," "skip," or "forgot" (and their variations, eg, forgotten) and the name of an OC (both brand and generic), or the phrase "minipill," "birth control," or "contraceptive."
2. Miscarriage queries: queries which contained the words "after miscarriage," "post miscarriage," or "I had a miscarriage."
3. Abortion queries: queries which contained the word "abortion," excluding queries referring to specific legislation, abortion debates, or celebrities who had an abortion.
4. Emergency contraceptives: queries which contained the words "plan b" or "morning after pill," excluding queries referring to specific legislation, abortion debates, or celebrities who had an abortion.
5. Pregnancy queries: queries which contained the words "week" and "pregnancy" (eg, "what to expect on week 21 of pregnancy"). Past studies have shown that such queries have a high specificity for actual pregnancy [29].

Results

Accuracy of Identification

We validated the identification of relevant queries by manually inspecting the 20 most common queries of each of the 5 classes for whether they could be interpreted as relevant to the topic. For example, queries regarding general news on OC use were deemed irrelevant to the research topic and were excluded from the study. Two of the authors independently reviewed these queries and marked them for relevancy. Kappa agreement between the labelers was, on average, 0.866 (n=5). Considering only the queries that both labelers agreed to be relevant, 75% of missed OC queries; 100% of pregnancy, miscarriage, and

abortion queries; and 80% of emergency OC queries were correctly identified.

Missed OC Queries

In 2018, 26,395 Bing users in the United States queried about missing OC pills. Of these, 60.8% did not mention the type of OC they missed, 20.9% mentioned a combined OC brand, and 21.7% mentioned a progestin-only OC brand or minipill (a total number of users is greater than 100% since a minority of users mentioned multiple types). Of users who mentioned the number of missed doses, 21.4% mentioned forgetting 1 dose; 6.3% forgot 2 doses; 3.4% forgot 3 doses; and 1.8% forgot 4 or more doses. Other users did not mention the specific number of doses they forgot, and their queries did not directly indicate this information.

There was a weak positive correlation (ρ=0.07, P<.001) between user age and the number of missing doses.

Figure 1 shows the percentage of users who asked about a missing OC dose and then queried about abortion, miscarriage, emergency contraception, or a week of pregnancy (the latter indicating that they most probably were pregnant at a later time). Furthermore, it shows how many of the users who query for the first 3 classes followed with queries for a week of pregnancy.

The majority of the queries for these 3 classes occur within a few days of the missed OC dose query (Figure 2). The distribution of queries that were made within 24 hours after a query about missing an OC was as follows: 7% of miscarriage queries, 20% of abortion queries, and 37% of emergency contraceptive queries. The median query time after a query about missing an OC was 35 days for miscarriage queries, 28 days for abortion queries, and 11 days for emergency contraceptive queries.

Figure 1. Percentage of users who asked about missed OC doses and then asked about miscarriage, pregnancy (as indicated by a query for the week of pregnancy), abortion, and emergency contraception as well as the percentage of users who asked about missed OC doses, followed by a query about miscarriage, abortion, or emergency contraceptive, before querying about the week of pregnancy.

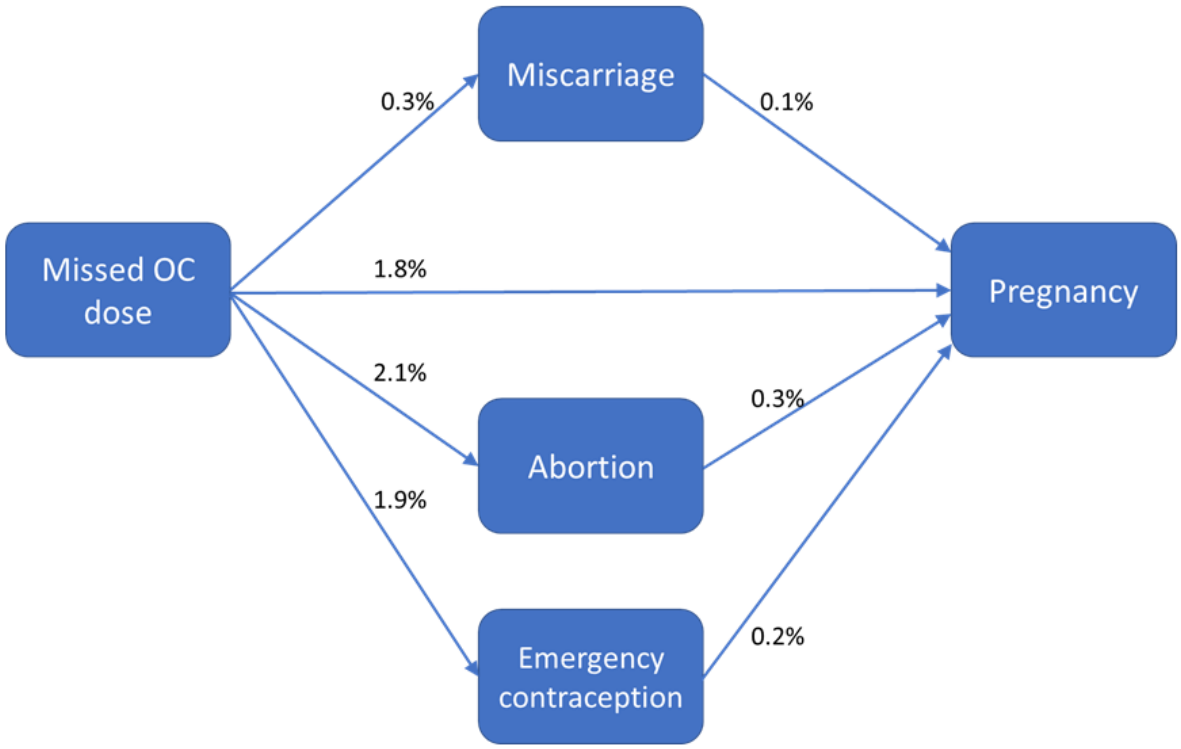
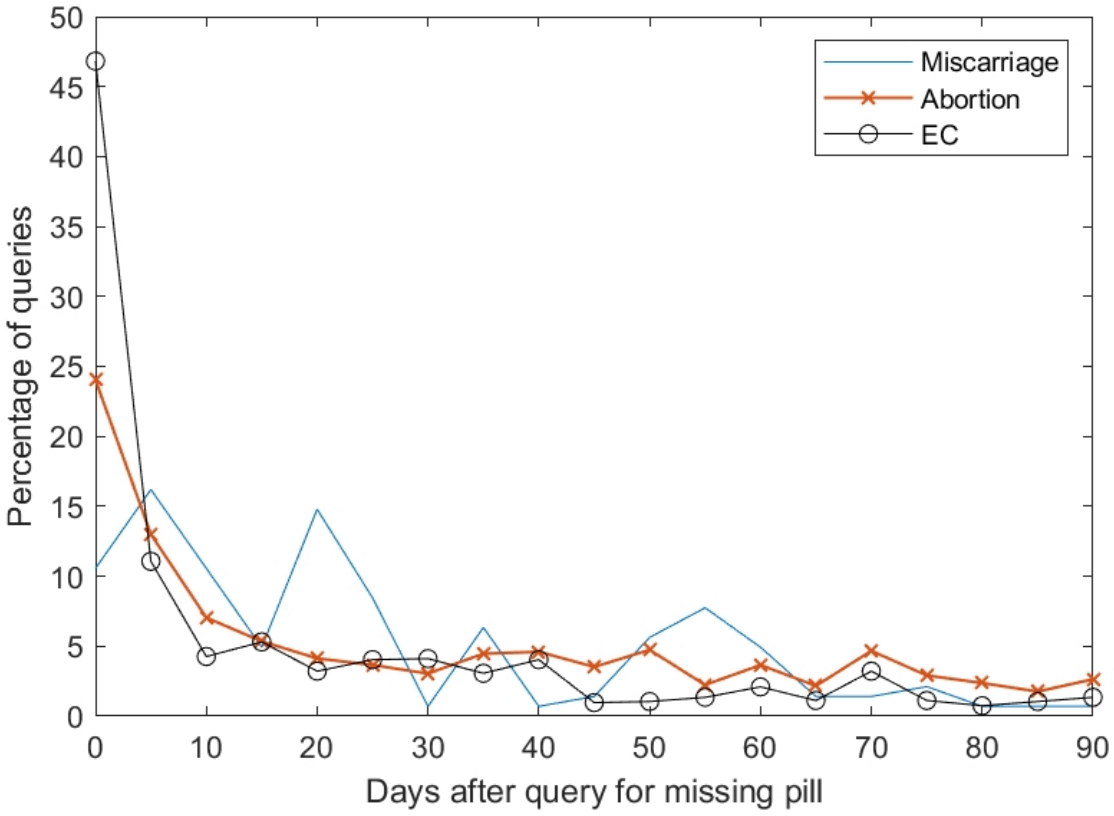


Figure 2. Distribution of the number of days between queries for a missed OC dose and for miscarriage, abortion, and emergency contraceptive.



Unintended Pregnancies Following a Missed OC Query

Each user who reported a missed dose was assigned a weight relative to the likelihood for querying about pregnancy, relative to the date of reporting the missed dose (see Methods). The weighted percentage of women who queried for pregnancy week following a query for a missed pill was 4.7%. Only 19% of these queries could be timed to the cycle where the same women had queried about missing a pill. We calculated the inferred LMP from each of these queries. Among the 3130 women who queried for pregnancy at different weeks of gestation, there was a good consistency between the calculated LMP for recurrent queries, and the median number of days between the inferred LMP was 4 days.

Some women mentioned in their query that they missed a placebo pill. For those, the rate of pregnancy was 0.2% versus 4.5% for the general population. The weighted rate of pregnancy for women who mentioned taking a minipill was 8.7% versus 4.7% for other medications.

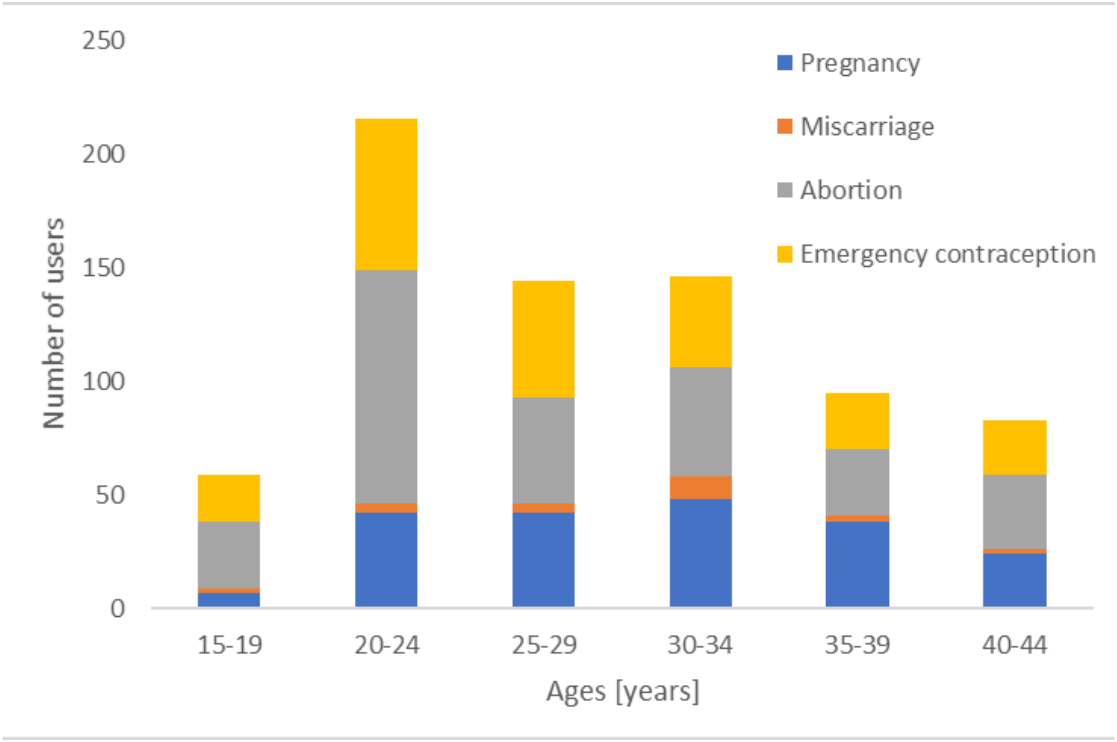
From queries on the week of pregnancy, we calculated the first day of the last cycle. The average day of conception was 14.4 days (SD 7.2) after the first day of the cycle.

The weighted percentage of females who queried for pregnancy following a query for an unspecified number of missed pills was 4.7%. The weighted pregnancy rates of females who queried for 2 or more missing doses was 5.1% (ie, 8.5% higher).

Age Distribution of Users Searching for Information on Missed OC

The average age of users who queried for a missing dose was 32 (SD 12) years (mode=20 years). Figure 3 shows the distribution by age of users who asked about missed OC dose and then each of the other query types. Figure 3 demonstrates that users under the age of 30 years are more likely to ask about abortion (1.5 times) and emergency contraception (1.7 times) (chi-square $P<.001$ for both), while users at ages of 30-34 years are more likely to query about pregnancy (2.1 times) and miscarriage (5.4 times) (chi-square $P<.001$ for both).

Figure 3. Number of users by age group who asked about missed OC doses and each of the other query categories.



Discussion

The study identified a myriad of women who queried missing an OC pill. Young women are more likely to miss a dose or multiple doses and to search for information on emergency contraceptives and abortion. Following a missed OC query, abortion and miscarriage queries are common, peaking in the first few days after the missed OC.

A small but not insignificant percentage of women querying for information after missing an OC dose eventually get pregnant in the month of query. As expected, the likelihood of pregnancy after missing a placebo pill is practically zero, and the likelihood of pregnancy after missing a progestin-only pill is double the risk of missing a dose of combined OC.

A majority of people occasionally forget to take a dose of a chronic medication, and studies have found that about 1 in 2 young women miss a dose of OC regularly [15-18]. Our data indicate that some women missing a dose of OC did not obtain sufficient information from their health care providers on the consequences of missing a pill, and thus, they resort to search the relevant information on the internet. Search for emergency contraception, which peaks in the first few days after querying about missing an OC dose, is an additional sign of inadequate information supplied by the health care provider.

Queries about abortion and miscarriage peak in the first few days, following a query about forgetting an OC dose. Many of these queries are most likely not an indication that the woman has confirmed being pregnant but a sign that she is very worried of being pregnant and thinking ahead about her further options of dealing with a possible unwanted pregnancy.

Our results also show that only one-third of queries about emergency contraception occur within the first 24 hours after the missed OC query. This, together with the fact that the median time between the queries is 11 days, indicates that many women are finding information on emergency contraception too late for it to be useful.

The strength of this study is access to the moment of uncertainty and information-seeking behavior of a large number of women missing a dose of OC. The limitations are absence of information on missed OC doses in other months or further doses in the same month. Pregnancy or its absence cannot be comprehensively confirmed.

The main implication of our results is reinforcement of the importance of providing comprehensive written information and directing patients to reliable information resources on missed pills. Most women are more familiar with what to do when one misses a pill, but they lack the knowledge of what to do when missing more than 1 OC dose [19]. Therefore, it is important to provide instructions for when a woman misses more than 1 OC pill in a package.

As young women are more likely to miss a pill, it may be more effective to offer women in these age groups contraceptive alternatives that do not depend on daily compliance. Health care providers should establish special counseling methods for young women who choose OCs and make sure that the instructions are understood by those young women. All OC users should also be provided with information on emergency contraceptive options and rapid access to an emergency contraception in case it is needed due to missed pills and unprotected intercourse. As young people often turn to internet search engines for health advice, health providers could comply with the WHO guidelines for contraceptive use by not only supplying information during people's visit to the clinic but also adding detailed information on preventing pregnancy in case of missed OC doses to their websites or as part of a mobile app that is given to patients at

the initial contraceptive consultation [21]. In addition, it is important that resources with accurate information be provided by search engines when people search for "missed pill" or any of the terms identified and used for this study. Better information can lead to better planning of contraceptive use when a pill is missed, reducing the risk of pregnancy and preventing stress and anxiety related to the fear of an unwanted pregnancy.

Conflicts of Interest

EYT is an employee of Microsoft, owner of the search engine Bing.

References

1. World Health Organization. Selected practice recommendations for contraceptive use. Second ed. Geneva: World Health Organization; 2004.
2. Johnson S, Pion C, Jennings V. Current methods and attitudes of women towards contraception in Europe and America. Reprod Health 2013 Feb 05;10(1):7 [FREE Full text] [doi: 10.1186/1742-4755-10-7] [Medline: 23384291]
3. Trussell J. Contraceptive failure in the United States. Contraception 2004 Aug;70(2):89-96. [doi: 10.1016/j.contraception.2004.03.009] [Medline: 15288211]
4. Dehlendorf C, Fox E, Sobel L, Borrero S. Patient-Centered Contraceptive Counseling: Evidence to Inform Practice. Curr Obstet Gynecol Rep 2016 Jan 29;5(1):55-63. [doi: 10.1007/s13669-016-0139-1]
5. Rosenberg MJ, Waugh MS, Meehan TE. Use and misuse of oral contraceptives: risk indicators for poor pill taking and discontinuation. Contraception 1995 May;51(5):283-288 [FREE Full text] [doi: 10.1016/0010-7824(95)00074-k] [Medline: 7628201]
6. He K, Dalton VK, Zochowski MK, Hall KS. Women's Contraceptive Preference-Use Mismatch. J Womens Health (Larchmt) 2017 Jun;26(6):692-701 [FREE Full text] [doi: 10.1089/jwh.2016.5807] [Medline: 27710196]
7. Jain AK, Obare F, RamaRao S, Askew I. Reducing unmet need by supporting women with met need. Int Perspect Sex Reprod Health 2013 Sep;39(3):133-141. [doi: 10.1363/3913313] [Medline: 24135045]
8. Ugaz JI, Chatterji M, Gribble JN, Banke K. Is Household Wealth Associated With Use of Long-Acting Reversible and Permanent Methods of Contraception? A Multi-Country Analysis. Glob Health Sci Pract 2016 Mar;4(1):43-54 [FREE Full text] [doi: 10.9745/GHSP-D-15-00234] [Medline: 27016543]
9. Daniels K, Mosher WD, Jones J. Contraceptive methods women have ever used: United States, 1982-2010. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2013.
10. Sundaram A, Vaughan B, Kost K, Bankole A, Finer L, Singh S, et al. Contraceptive Failure in the United States: Estimates from the 2006-2010 National Survey of Family Growth. Perspect Sex Reprod Health 2017 Mar;49(1):7-16 [FREE Full text] [doi: 10.1363/psrh.12017] [Medline: 28245088]
11. Black KI, Gupta S, Rassi A, Kubba A. Why do women experience untimed pregnancies? A review of contraceptive failure rates. Best Pract Res Clin Obstet Gynaecol 2010 Aug;24(4):443-455 [FREE Full text] [doi: 10.1016/j.bpobgyn.2010.02.002] [Medline: 20335073]
12. Free C, Ogden J, Lee R. Young women's contraception use as a contextual and dynamic behaviour: A qualitative study. Psychology & Health 2005 Oct;20(5):673-690. [doi: 10.1080/08870440412331337110]
13. Caetano C, Peers T, Papadopoulos L, Wiggers K, Engler Y, Grant H. Millennials and contraception: why do they forget? An international survey exploring the impact of lifestyles and stress levels on adherence to a daily contraceptive regimen. Eur J Contracept Reprod Health Care 2019 Feb;24(1):30-38. [doi: 10.1080/13625187.2018.1563065] [Medline: 30689459]
14. Winner B, Peipert JF, Zhao Q, Buckel C, Madden T, Allsworth JE, et al. Effectiveness of Long-Acting Reversible Contraception. N Engl J Med 2012 May 24;366(21):1998-2007. [doi: 10.1056/nejmoa1110855]
15. Frost JJ, Darroch JE. Factors associated with contraceptive choice and inconsistent method use, United States, 2004. Perspect Sex Reprod Health 2008 Jun;40(2):94-104 [FREE Full text] [doi: 10.1363/4009408] [Medline: 18577142]
16. Molloy GJ, Graham H, McGuinness H. Adherence to the oral contraceptive pill: a cross-sectional survey of modifiable behavioural determinants. BMC Public Health 2012 Oct 02;12:838 [FREE Full text] [doi: 10.1186/1471-2458-12-838] [Medline: 23031437]
17. Rosenberg MJ, Waugh MS, Burnhill MS. Compliance, counseling and satisfaction with oral contraceptives: a prospective evaluation. Fam Plann Perspect 1998;30(2):89-92, 104 [FREE Full text] [Medline: 9561874]
18. Hooper DJ. Attitudes, awareness, compliance and preferences among hormonal contraception users: a global, cross-sectional, self-administered, online survey. Clin Drug Investig 2010;30(11):749-763 [FREE Full text] [doi: 10.2165/11538900-000000000-00000] [Medline: 20818837]
19. Zapata LB, Steenland MW, Brahmi D, Marchbanks PA, Curtis KM. Patient understanding of oral contraceptive pill instructions related to missed pills: a systematic review. Contraception 2013 May;87(5):674-684 [FREE Full text] [doi: 10.1016/j.contraception.2012.08.026] [Medline: 23040136]

20. Lete I, Doval JL, Pérez-Campos E, Lertxundi R, Correa M, de la Viuda E, et al. Self-described impact of noncompliance among users of a combined hormonal contraceptive method. Contraception 2008 Apr;77(4):276-282 [FREE Full text] [doi: 10.1016/j.contraception.2007.11.009] [Medline: 18342651]
21. Levine D. Using Technology, New Media, and Mobile for Sexual and Reproductive Health. Sex Res Soc Policy 2011 Feb 25;8(1):18-26. [doi: 10.1007/s13178-011-0040-7]
22. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157 [FREE Full text] [doi: 10.1093/her/cyt071] [Medline: 23861481]
23. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
24. Hochberg I, Allon R, Yom-Tov E. Assessment of the Frequency of Online Searches for Symptoms Before Diagnosis: Analysis of Archival Data. J Med Internet Res 2020 Mar 06;22(3):e15065 [FREE Full text] [doi: 10.2196/15065] [Medline: 32141835]
25. Lebwohl B, Yom-Tov E. Symptoms Prompting Interest in Celiac Disease and the Gluten-Free Diet: Analysis of Internet Search Term Data. J Med Internet Res 2019 Apr 08;21(4):e13082 [FREE Full text] [doi: 10.2196/13082] [Medline: 30958273]
26. Yom-Tov E. Crowdsourced Health: How What You Do on the Internet Will Improve Medicine. Boston, MA: MIT Press; 2016.
27. Latest Rankings. Comscore. 2018. URL: https://www.comscore.com/Insights/Rankings [accessed 2020-08-21]
28. Rosenblum S, Yom-Tov E. Seeking Web-Based Information About Attention Deficit Hyperactivity Disorder: Where, What, and When. J Med Internet Res 2017 Apr 21;19(4):e126 [FREE Full text] [doi: 10.2196/jmir.6579] [Medline: 28432038]
29. Fourney A, White RW, Horvitz EJ. Exploring Time-Dependent Concerns about Pregnancy and Childbirth from Search Logs. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2015 Apr Presented at: CHI 15: CHI Conference on Human Factors in Computing Systems; 2015; Seoul, Republic of Korea p. 737-746. [doi: 10.1145/2702123.2702427]

Abbreviations

LMP: last menstrual period
OC: oral contraceptive
WHO: World Health Organization

Edited by G Eysenbach; submitted 24.05.20; peer-reviewed by F Sanz, D Levine; comments to author 18.06.20; revised version received 30.06.20; accepted 26.07.20; published 15.09.20.

Please cite as:
Hochberg I, Orshalimy S, Yom-Tov E
Real-World Evidence on the Effect of Missing an Oral Contraceptive Dose: Analysis of Internet Search Engine Queries
J Med Internet Res 2020;22(9):e20632
URL: http://www.jmir.org/2020/9/e20632/
doi:10.2196/20632
PMID:32930672

Original Paper

Changing Emotions About Fukushima Related to the Fukushima Nuclear Power Station Accident—How Rumors Determined People's Attitudes: Social Media Sentiment Analysis

Shin Hasegawa[1,2], MSc; Teppei Suzuki[3,4], MBA, PhD; Ayako Yagahara[4,5], PhD; Reiko Kanda[2], PhD; Tatsuo Aono[2], PhD; Kazuaki Yajima[2], PhD; Katsuhiko Ogasawara[2,4], MBA, PhD

[1]Graduate School of Health Sciences, Hokkaido University, Sapporo, Japan

[2]Quantum Medical Science Directorate, National Institutes for Quantum and Radiological Science and Technology, Chiba, Japan

[3]Hokkaido University of Education, Iwamizawa, Japan

[4]Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

[5]Department of Radiological Technology, Hokkaido University of Science, Sapporo, Japan

Corresponding Author:
Katsuhiko Ogasawara, MBA, PhD
Faculty of Health Sciences
Hokkaido University
N12-W5, Kita-ku
Sapporo, 060-0812
Japan
Phone: 81 11 706 3409
Email: oga@hs.hokudai.ac.jp

Abstract

Background: Public interest in radiation rose after the Tokyo Electric Power Company (TEPCO) Fukushima Daiichi Nuclear Power Station accident was caused by an earthquake off the Pacific coast of Tohoku on March 11, 2011. Various reports on the accident and radiation were spread by the mass media, and people displayed their emotional reactions, which were thought to be related to information about the Fukushima accident, on Twitter, Facebook, and other social networking sites. Fears about radiation were spread as well, leading to harmful rumors about Fukushima and the refusal to test children for radiation. It is believed that identifying the process by which people emotionally responded to this information, and hence became gripped by an increased aversion to Fukushima, might be useful in risk communication when similar disasters and accidents occur in the future. There are few studies surveying how people feel about radiation in Fukushima and other regions in an unbiased form.

Objective: The purpose of this study is to identify how the feelings of local residents toward radiation changed according to Twitter.

Methods: We used approximately 19 million tweets in Japanese containing the words "radiation" (放射線), "radioactivity" (放射能), and "radioactive substances" (放射性物質) that were posted to Twitter over a 1-year period following the Fukushima nuclear accident. We used regional identifiers contained in tweets (ie, nouns, proper nouns, place names, postal codes, and telephone numbers) to categorize them according to their prefecture, and then analyzed the feelings toward those prefectures from the semantic orientation of the words contained in individual tweets (ie, positive impressions or negative impressions).

Results: Tweets about radiation increased soon after the earthquake and then decreased, and feelings about radiation trended positively. We determined that, on average, tweets associating Fukushima Prefecture with radiation show more positive feelings than those about other prefectures, but have trended negatively over time. We also found that as other tweets have trended positively, only bots and retweets about Fukushima Prefecture have trended negatively.

Conclusions: The number of tweets about radiation has decreased overall, and feelings about radiation have trended positively. However, the fact that tweets about Fukushima Prefecture trended negatively, despite decreasing in percentage, suggests that negative feelings toward Fukushima Prefecture have become more extreme. We found that while the bots and retweets that were not about Fukushima Prefecture gradually trended toward positive feelings, the bots and retweets about Fukushima Prefecture trended toward negative feelings.

(J Med Internet Res 2020;22(9):e18662) doi:10.2196/18662

KEYWORDS

Fukushima nuclear accident; Twitter messaging; radiation; radioactivity; radioactive hazard release; information dissemination; belief in rumors; disaster medicine; infodemiology; infoveillance; infodemic

Introduction

Overview

At 2:46 PM JST (Japan Standard Time) on March 11, 2011, a magnitude-9 earthquake occurred off the Pacific coast of Tohoku—the Great East Japan Earthquake. Its epicenter was off the Sanriku coast, and a tsunami with a run-up height of 14-15 meters followed one hour later, causing a blackout of the Tokyo Electric Power Company (TEPCO) Fukushima Daiichi Nuclear Power Station. A meltdown occurred in reactors 1, 2, and 3 while they were in operation, causing a large quantity of hydrogen to be generated. A hydrogen explosion occurred in reactor 1 on March 12, followed by another explosion in reactor 3 on March 14. Reactors 2 and 4 were damaged, releasing a large quantity of radioactive substances into the environment. This accident was classified as Level 7 according to the International Nuclear and Radiological Event Scale [1], and its effects were evaluated and reported by international organizations [2-4]. They indicated that the accident had increased anxieties about radiation and had given rise to a social phenomenon described as inciting harmful rumors about the disaster area. This has resulted in physical damage from the disaster alongside economic damage from, for example, consumer reluctance to buy agricultural products from the disaster area [5]. In the medical field, anxieties about radiation were reflected in a decrease in the number of computed tomography (CT) scans and other forms of radiographic examination performed on young children in Fukushima Prefecture compared to before the accident [6,7]. One out of every four doctors surveyed reported that parents of young children were refusing to subject them to such examinations due to the risk of radiation [7] and that they had witnessed an aversion to radiation. We believe that people have become gripped by increasingly negative feelings over time concerning Fukushima Prefecture as a disaster area associated with radiation, and this may have influenced behavior such as restrained consumption activities and an aversion to medical radiation.

Background

Immediately following the earthquake, telephone lines were damaged and communication was cut off or limited. Outgoing calls on mobile phones were restricted up to 95%. For packet communications, NTT (Nippon Telegraph and Telephone) Docomo, the predominant mobile phone operator in Japan, imposed a 30% restriction but it was soon lifted. Other carriers did not implement any restrictions [8]. Therefore, social networking services were used as a means to transmit information, and communities to exchange information rapidly formed on Twitter [9,10]. In a survey, Twitter was found to be the most-used form of social media in coping with the disaster over Facebook or Mixi, and it was shown to have an influence on attitudes toward the Fukushima nuclear accident [11]. Successive reports were issued by the mass media on the condition of the Fukushima nuclear station, the city of Fukushima, other regions affected by radioactive substances, and the effects of the radiation itself; social networks not only carried this reported information and the responses to it, but also rapidly spread unreliable information, misinformation, and ugly rumors, thus indicating that social media can cause social unrest and chaos [12]. Ikegami et al [13] proposed a reliable analysis system for tweets about the 2011 earthquake disaster using topic categories according to latent Dirichlet allocation, with 2960 tweets containing the words “earthquake disaster,” “earthquake,” “tsunami,” “radioactivity,” “radioactive substances,” and “Becquerel” as a dataset, and sentiment analysis using a semantic direction dictionary [14]. Wang and Kim [15] showed that behavior in cyberspace and real-world behavior mutually influence one another. Using this as a basis, we believe that people who have come into contact with social anxieties and ugly rumors as spread on social media networks have an increased aversion to Fukushima Prefecture, which may lead to additional harmful rumors and a refusal of radiographic examinations.

Prior Work

It has been shown that 5 years after the accident, groups that relied on the internet as a source of information had significantly higher anxieties about health issues caused by radiation exposure than groups that used other information sources [16]. Mothers of children younger than elementary school age using Twitter and other forms of social media were shown to have a higher degree of risk perception and to actively pursue risk-reducing activities [17]. It is not hard to imagine that high anxiety and risk-reducing activities may lead to a refusal of radiographic examinations of children. Risk communication is valuable in eliminating these social anxieties, but studies have indicated that social media was not fully utilized at the time of the Fukushima accident [18]. Yagahara et al [19] analyzed tweets for 7 days from the day of the earthquake on March 11 until March 17 to examine the changes in interest in radiation among Japanese citizens. Their analysis was based on co-occurrence networks related to radiation as the accident’s situation progressed. Since the analysis was also conducted in relation to regions, it did not identify how the interest subsequently formed people’s attitudes toward Fukushima. Aoki et al [20] surveyed tweeting trends 1 year after the earthquake by dividing regions tweeted about into four zones based on geotags. This analysis only concerned the regions that people were tweeting from and did not analyze the content of the tweets themselves. Therefore, we believe that surveying how people received information associated with the Fukushima nuclear accident and how they reacted to it will be of great significance in establishing a basis for effective risk communication when similar disasters and accidents occur in the future.

Goal of This Study

This study concerns the decrease in CT scans and other radiographic examinations [6,7], the persistent restrained buying of agricultural products by consumers, and harmful rumors [5] regarding Fukushima Prefecture. Its purpose is to comparatively identify how feelings associated with radiation have shifted from Fukushima Prefecture to other regions 1 year after the 2011 earthquake and accident and thereby clarify how such situations are formed. People may have become gripped by more negative feelings over time regarding Fukushima Prefecture as a disaster area associated with radiation, which may have influenced an increased aversion toward Fukushima Prefecture in general.

Methods

Overview

In this study, we selected statements about radiation containing the words "radiation" (放射線), "radioactivity"（放射能), and/or "radioactive substances" （放射性物質） that were posted on Twitter in Japanese between the occurrence of the Great East Japan Earthquake at 12 AM JST on March 11, 2011, until 1 year later (ie, 11:59 PM JST on March 10, 2012) using approximately 19 million tweets. We grouped tweets starting with retweets (RTs) and quote tweets (QTs) to indicate that they were automatic tweets—referred to as *bots*—and RTs that had reposted someone else's tweet into an RT group, leaving approximately 9 million tweets, excluding the RT group, as a target group. Bot accounts were accounts in which the user's ID began or ended with the word *bot*. In order to more accurately sample the original poster's feelings, we deleted anything in the target group that followed an RT, which indicated that someone else's tweet had been reposted, and then processed each of the tweets using semantic orientation, which refers to a binary attribute that shows how a word might generally carry a positive or a negative impression. Takamura et al [14] used the Japanese dictionary *Iwanami Kokugo Jiten* to create semantic orientation values of words by assigning semantic orientation values with actual values from –1 (carries mostly a negative impression) to 1 (carries mostly a positive impression) for 49,002 nouns, 4254 verbs, 665 adjectives, and 1207 adverbs. All words and phrases contained in this study's tweets were scored using the above morphological analysis based on the original form of the words by using their semantic orientation values. In addition, the data used tweets containing any of the following words: "radiation," "radioactivity," or "radioactive substances." The semantic orientation values for these words were evaluated: "radioactivity" had a value of –0.598318, "radiation" had a value of –0.560393, and "radioactive" had a value of –0.178744; there were no instances of "radioactive substances" in Takamura et al's correspondence table. These were originally logged as negative words. The purpose of this study is to analyze the feelings associated with the three words above; in order to exclude these influences, we scored these words as having 0 points. We scored the words that were not included in Takamura et al's semantic orientation values of words [14] in the correspondence table as having 0 points as well, so as to avoid affecting the tweets' semantic value orientations, which we calculated using the following formula:

$$T_{pn} = \sum W_{pn}/W_c$$

where T_{pn} represents the tweet semantic orientation value, W_{pn} represents the word semantic orientation value, and W_c represents the number of words contained in a tweet.

In order to categorize the prefectures to which the tweets related, the words and phrases related to regions contained in tweets (ie, place names, telephone numbers, and postal codes) were sampled and categorized by prefecture. For place names, we morphologically analyzed tweets using the morphological analysis engine MeCab [21] and the mecab-ipadic-NEologd [22] dictionary and searched the following word types for their address character strings with the Yahoo! Geocoding application programming interface [23]: parts of speech = nouns; parts of speech (subcategory 1) = proper nouns; and parts of speech (subcategory 2) = regions. We then identified and categorized the prefectures (Multimedia Appendix 1 contains the scripts to process these procedures).

Sentiment Analysis About Radiation Regarding Fukushima Prefecture and Other Prefectures

The tweets categorized by prefecture were divided into two groups: Fukushima Prefecture and other prefectures. We sampled the average weekly tweets' semantic orientation values for Fukushima Prefecture and other prefectures and then surveyed how the feelings on radiation regarding each region had changed. The dataset used 18,841,755 out of 18,851,259 tweets of words having semantic orientation values. Semantic orientation values were sampled for almost all tweets. The dataset included a total of 18,851,259 tweets and a target group of 9,025,831 tweets, excluding bots and RTs. The respective daily changes in the number of tweets are indicated in Figure 1 by a blue line and a red line. The linear approximation of each is drawn with a dashed line. Figure 2 shows the daily number of tweets at a more granular level (ie, by the minute) for March 11, the day the Great East Japan Earthquake occurred.

Figure 1. The number of tweets per day.

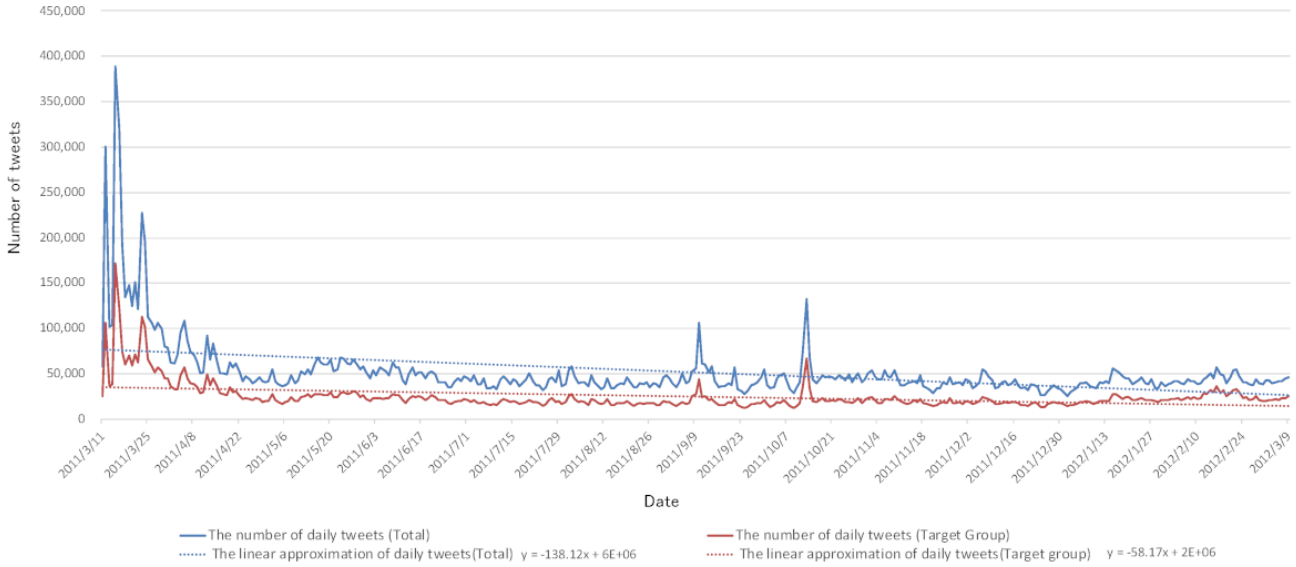
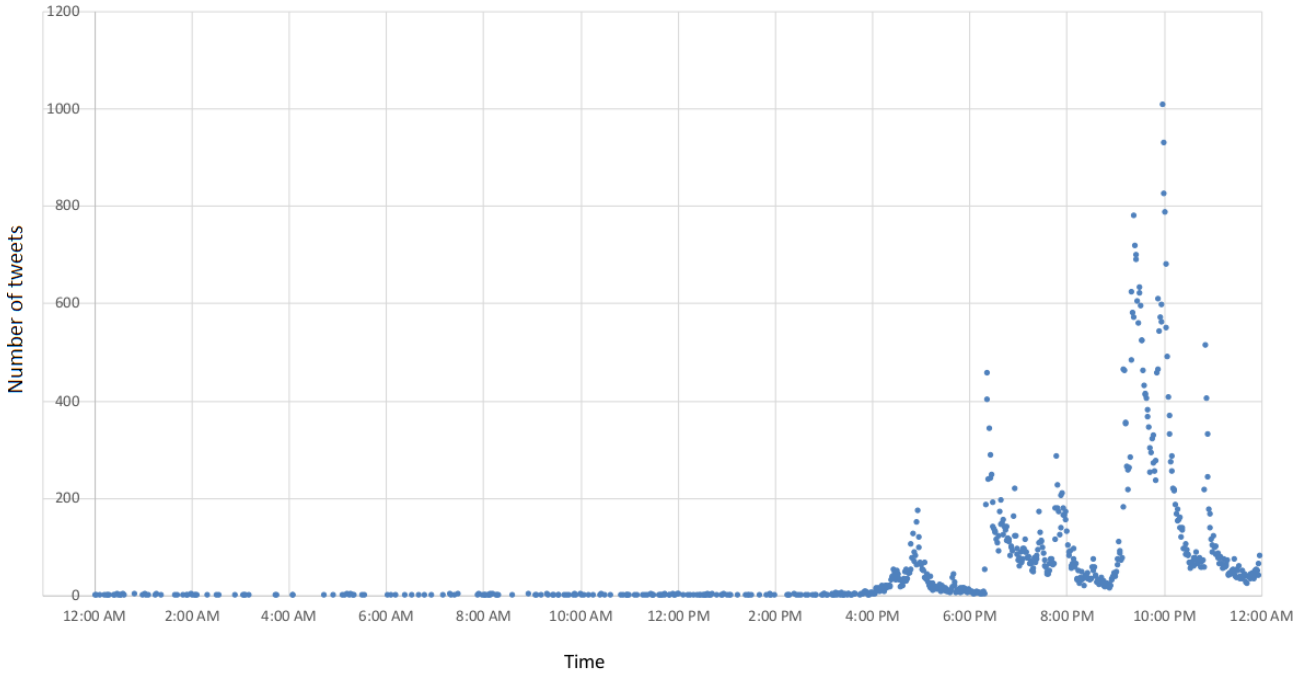


Figure 2. The number of tweets by the minute on March 11, 2011.



Figures 3 and 4, similar to Figure 1, show the respective daily average changes and cumulative changes in the tweets' semantic orientation values. In addition, average changes in the RT group's semantic orientation values are included in Figure 3, and the linear approximation is represented by a dashed line. Figure 5 shows the changes during weekly *F* tests of the average semantic orientation values in order to observe a correlation between the target group and the RT group. A red line is drawn using the significance level α=.05.

Figure 3. Daily average of tweets' semantic orientation values. RT: retweet.

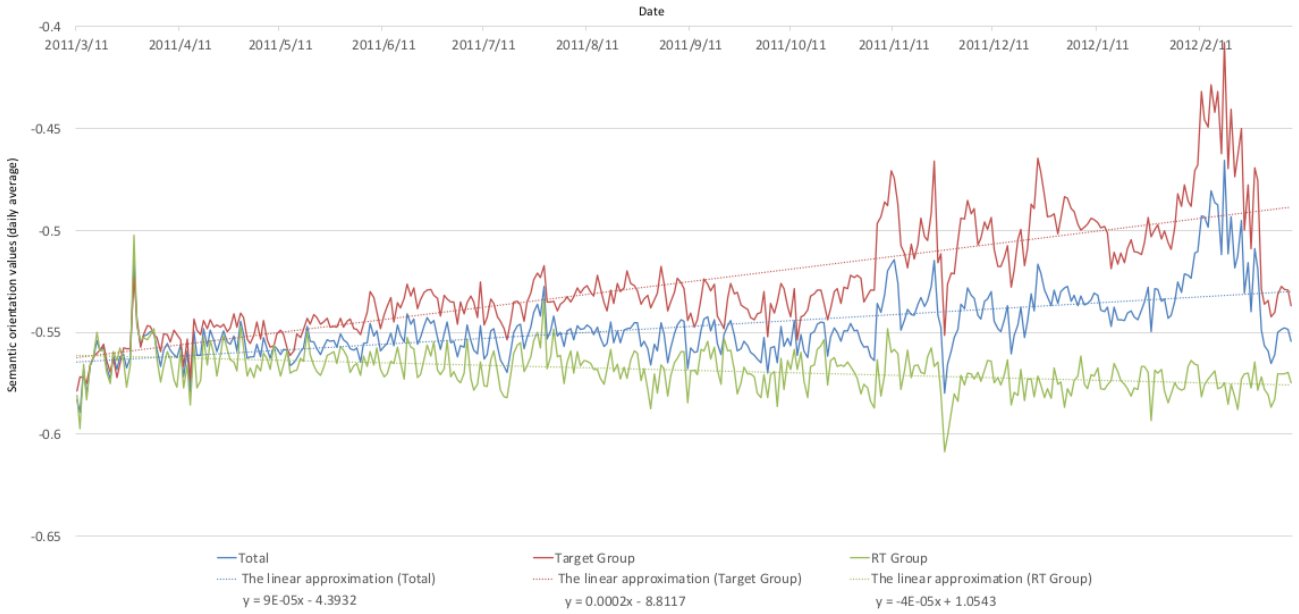


Figure 4. Daily integration of semantic orientation values.

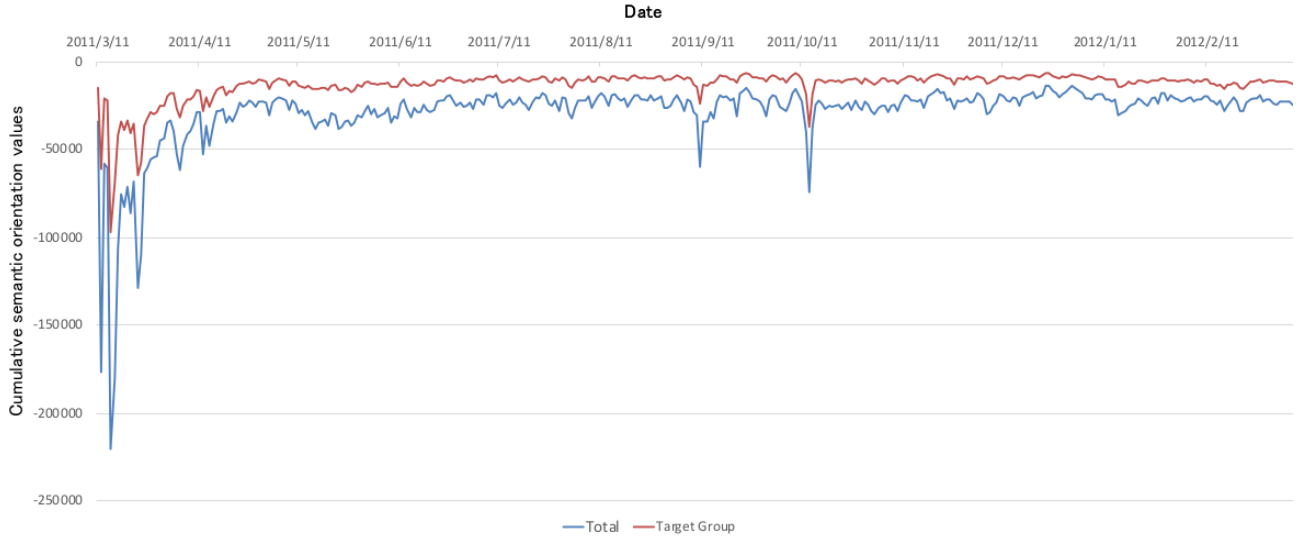
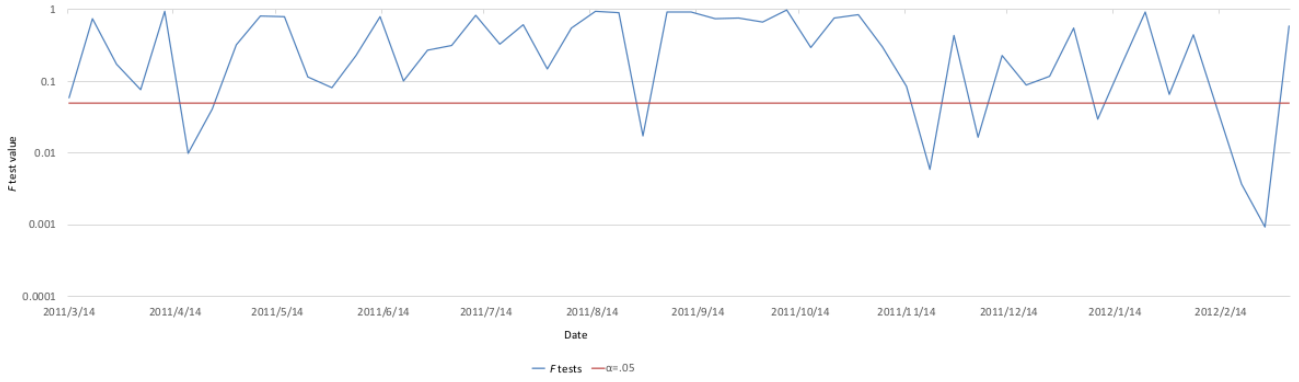


Figure 5. *F* test for target and retweet (RT) groups by week.



The target group was categorized by prefecture. There were 34,233 words representing regional identifiers and 3,004,726 tweets in the target group containing words representing the regional identifiers. We calculated the number of tweets per 1000 people according to the breakdown in each prefecture and its population as of October 1, 2011 [24]. This is represented by shading on the map in Figure 6. Details are shown in Table 1, in which the *Other* row contains words that made it difficult

to identify the prefecture and foreign place names. Words that made it difficult to identify the prefecture include universal words representing addresses, such as “1-chome” and “1-banchi,” for example. Foreign place names frequently included regions that experienced nuclear power accidents in the past, such as Chernobyl and Three Mile Island.

Figure 6. The number of tweets per 1000 people.

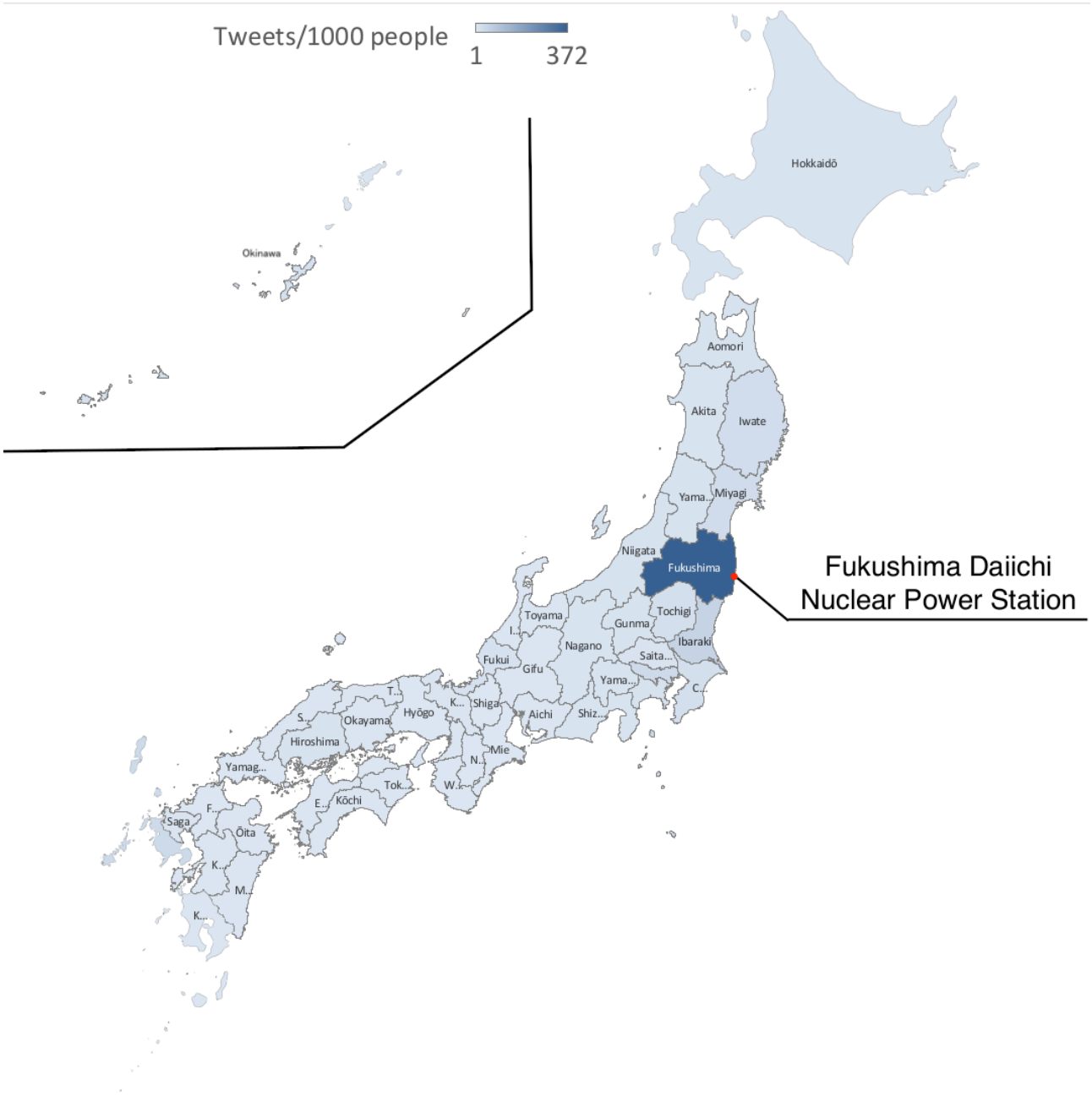


Table 1. Breakdown of words and tweets by prefecture.

Prefecture	Number of words representing the area in the prefecture	Number of tweets	Tweets per 1000 people
Hokkaido	635	51,871	9
Aomori	158	18,181	13
Iwate	545	32,319	25
Miyagi	911	64,714	28
Akita	200	10,473	10
Yamagata	295	13,523	12
Fukushima	1535	741,178	372
Ibaraki	1013	155,482	53
Tochigi	650	41,832	21
Gunnma	714	26,892	13
Saitama	1647	55,702	8
Chiba	1757	129,784	21
Tokyo	2614	441,874	33
Kanagawa	1766	108,510	12
Niigata	380	29,238	12
Toyama	100	4919	5
Ishikawa	121	3461	3
Fukui	137	11,437	14
Yamanashi	313	7881	9
Nagano	663	21,466	10
Gifu	325	8582	4
Shizuoka	615	36,133	10
Aichi	696	21,380	3
Mie	208	2339	1
Shiga	131	3497	2
Kyoto	319	18,232	7
Osaka	677	35,824	4
Hyogo	362	9341	2
Nara	150	3208	2
Wakayama	115	2271	2
Tottori	71	2774	5
Shimane	79	3081	4
Okayama	134	5937	3
Hiroshima	195	35,599	12
Yamaguchi	109	2462	2
Tokushima	75	2113	3
Kagawa	77	2229	2
Ehime	105	3974	3
Kochi	99	2596	3
Fukuoka	326	12,243	2
Saga	83	5757	7
Nagasaki	125	48,477	34

Prefecture	Number of words representing the area in the prefecture	Number of tweets	Tweets per 1000 people
Kumamoto	123	4620	3
Oita	104	3590	3
Miyazaki	106	3284	3
Kagoshima	127	4559	3
Okinawa	165	22,115	16
Other	16,235	1,396,553	N/A[a]

[a]N/A: not applicable; this was not calculated, as the population size for this category is not known.

Figures 7 and 8 show the ratio of the number of tweets between the target group in Fukushima Prefecture and other prefectures as well as the ratio of the average semantic orientation values. Semantic orientation values ranged from –1 to 1; thus, the ratio of average values in Figure 7 is the ratio when 1 is added to the average of the semantic orientation values and, therefore, it results in a value between 0 and 2. Figure 9 shows the changes in the average values of the semantic orientations for Fukushima Prefecture and other prefectures for the target group and RT group; details of the plots are as follows:

1. The solid grey line represents the weekly average of semantic orientation values, excluding bots and RTs, in Fukushima Prefecture.
2. The solid blue line represents the weekly average of semantic orientation values, excluding bots and RTs, outside of Fukushima Prefecture.
3. The solid yellow line represents the weekly average of semantic orientation values of bots and RTs in Fukushima Prefecture.
4. The solid orange line represents the weekly average of semantic orientation values of bots and RTs outside of Fukushima Prefecture.

The linear approximation was drawn with a dashed line for each respective line.

Figure 7. Weekly ratio of the number of tweets for Fukushima and other prefectures. The dotted line represents the linear approximation.

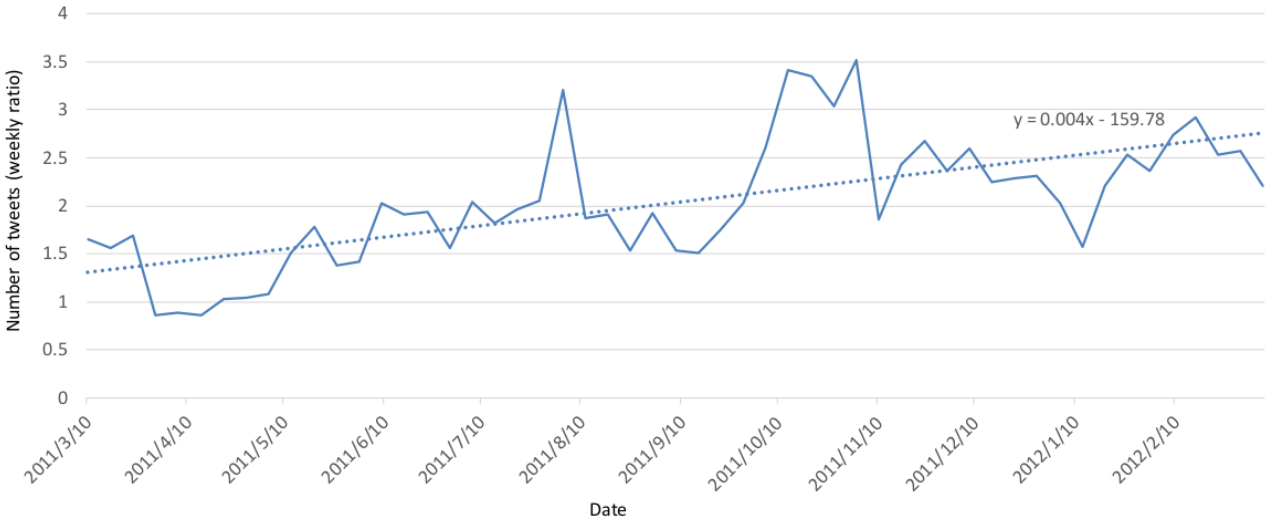


Figure 8. Weekly average ratio of semantic orientation values for Fukushima and other prefectures. The dotted line represents the linear approximation.

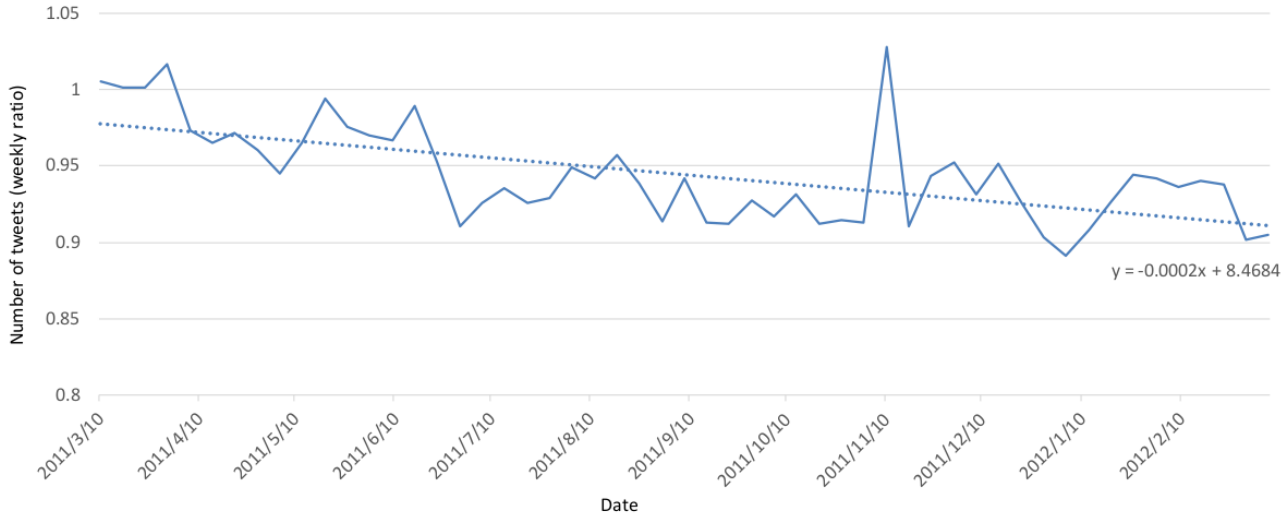
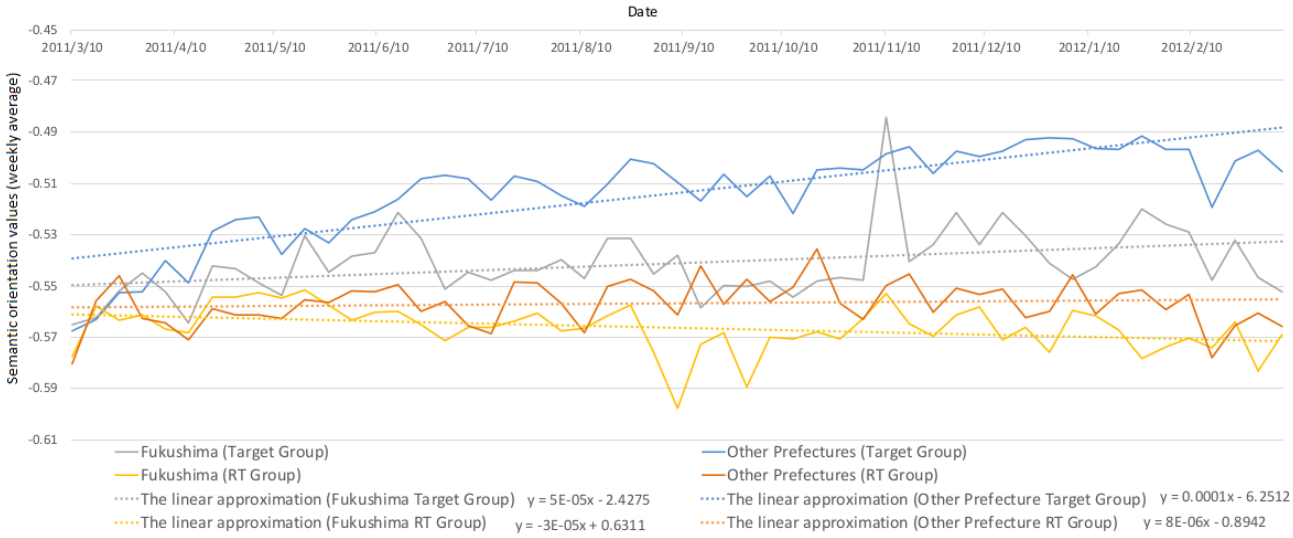


Figure 9. Weekly average of semantic orientation values for Fukushima Prefecture and other prefectures. RT: retweet.



Results

Overview

First, we discuss the characteristics and trends in the overall dataset. After discussing the characteristics for each prefecture, we then compare them with Fukushima Prefecture—as the disaster area and the main subject of harmful rumors—and other prefectures.

Characteristics and Trends in the Overall Dataset

Table 2 shows the table of events in 2011 in chronological order. At 2:46 PM JST on March 11, 2011, the Great East Japan Earthquake occurred, and a resulting tsunami struck various locations about one hour later. This gave rise to concerns about damage to the nuclear power stations on the Pacific coast at around 4 PM. Previously, there had been a few tweets containing the words “radiation,” “radioactivity,” and “radioactive substances,” but we found that the tsunami resulted in a rapid increase in tweets containing these three words (see Figure 2). Similar to tweets that were trending at the time of the 2010 Chilean earthquake as analyzed by Mendoza et al [25], many of the tweets posted immediately after the Great East Japan Earthquake can be found in our dataset. The overall amount of tweeting consisted of approximately 8% (of all tweets in the 1-year period) being posted 1 week after the earthquake and approximately 21% being posted 1 month after the earthquake, and there is a subsequent and gradually decreasing trend for the remaining period. As shown in Figure 1, approximately 300,000 tweets were posted on March 12 when reactor 1 experienced a hydrogen explosion, and approximately 400,000 tweets were posted on March 15 when the damage to reactors 2 and 4 became clear. On March 23, thyroid equivalent dose predictions involving iodine-131 for infants (under 1 year old) using the System for Prediction of Environmental Emergency Dose Information were released by the Cabinet Office’s Nuclear Safety Commission [26], and approximately 230,000 tweets were posted. Thereafter, tweeting hovered around 50,000 per day.

On September 10, the then-Minister of Economy, Trade, and Industry reportedly resigned after a visit to the exclusion zone of the Fukushima nuclear disaster, where he joked to a journalist, “I’ll give you radiation.” Later, a fire broke out at Sendai

Nuclear Power Plant, and the information about these two events spread. On October 14, the number of tweets exceeded 100,000 in response to news that a radium ray source was found under the floor of a private home in Setagaya Ward. As shown in Figure 3, many negative tweets were posted immediately after the Great East Japan Earthquake, and as shown by the approximation curve, they gradually trended positively thereafter. However, we found they were under –0.4 overall, and tweets expressing negative feelings about "radiation," "radioactivity," and "radioactive substances" were still posted. Although there is no difference between the overall dataset (see Figure 3, blue line) and most of the semantic orientation values in the target group immediately after the earthquake, the target group tended to be faster in its acceleration from negative to positive. More negative tweets were posted by bots and RTs, and we found that the sentiments in the bots and RTs tended to be negative as shown by the negative slope of the approximation's straight line. Bots and RTs often serve as a means for information to be spread widely. However, they were found to spread information containing negative sentiments more easily. This suggests that bots and RTs sensitize and lead users and communities that encounter this information to an increased aversion toward certain things.

There were characteristic changes in semantic orientation values close to the dates given below. These have been included, together with how much content was spread on these days. On March 28, a website visualizing the environmental radioactivity levels all over Kanto was launched [27]. Knowledge of the website was rapidly spread as a means of information sharing, together with positive words describing the website as easy to understand, so it tended to score positively at 0.029-0.067 points both before and after its launch, and the same trend was shown in both groups.

On November 11, a large number of more positive tweets than average were posted stating "Sign an emergency petition to save the children of Fukushima" (semantic orientation value: –0.396); thus, the tweets largely trended positively. These tweets were not posted by bots and were not in the form of RTs, and we found that the divergence between the groups was significant in the weeks before and after, as shown in Figure 4.

On November 23, Geiger counter advertisements, with semantic orientation values of about 0.05-0.40, were posted approximately 2-3 times as often compared to the days before and after—November 22 saw 1797 out of 34,173 tweets (5.26%); November 23 saw 3468 out of 33,778 tweets (10.27%); and November 24 saw 1087 out of 40,348 tweets (2.69%)—and this was believed to be influential. Similar trends were confirmed for the December 24 peak of the Geiger counter advertisements both before and after.

On November 26, it was reported that TEPCO had responded on November 24, "Any radioactive substances were not the property of TEPCO. Consequently, TEPCO has no responsibility for decontamination" [28]. However, this information was spread using negative words and, as a result, the semantic orientations in both groups largely trended negatively.

Interestingly, regardless of whether similar lines were previously drawn for the target group and the RT group, from December 7 to February 27, divergence emerged in the form of changes between the two groups. Particularly in the target group, the periods from December 21 to January 15 and from February 7 to February 27 peaked positively. On December 18, a maximum of –0.41 was reached, but no corresponding peak was formed in the RT group, which hovered around –0.58. As with the weekly *F* tests of the target group and RT group as shown in Figure 4, a significant divergence was seen between the groups during this period.

Table 2. Chronological listing of events in 2011.

Date (year/month/day)	Event and details
2011/03/11	An earthquake off the Pacific coast of Tohoku occurred.
2011/03/12	Reactor 1 of Fukushima Daiichi Nuclear Power Station experienced a hydrogen explosion.
2011/03/15	The damage to reactors 2 and 4 of Fukushima Daiichi Nuclear Power Station became clear.
2011/03/23	Thyroid equivalent dose predictions involving iodine-131 for infants (under 1 year old) using the System for Prediction of Environmental Emergency Dose Information were released by the Cabinet Office's Nuclear Safety Commission.
2011/03/28	A website visualizing the environmental radioactivity levels all over Kanto was launched.
2011/09/10	The then-Minister of Economy, Trade, and Industry reportedly resigned after a visit to the exclusion zone of the Fukushima nuclear disaster, where he joked to a journalist, "I'll give you radiation."
	A fire broke out at Sendai Nuclear Power Plant.
2011/10/14	A radium ray source was found under the floor of a private home in Setagaya Ward, Tokyo.
2011/11/11	A large number of tweets were posted stating "Sign an emergency petition to save the children of Fukushima."
2011/11/23	Geiger counter advertisements were posted approximately 2-3 times as often compared to the days before and after.
2011/11/24	At Tokyo District Court, TEPCO responded, "Any radioactive substances were not the property of TEPCO. Consequently, TEPCO has no responsibility for decontamination."
2011/11/26	TEPCO's response was reported in the press.

Characteristics Per Prefecture

As shown in Table 1, there was a spike in tweets about Fukushima Prefecture as a disaster area, which stands out in terms of population ratio. Areas in east Japan close to the disaster area had a larger number of tweets than west Japan and were at high rates in terms of their population ratio. Osaka and Kyoto, which are major cities in west Japan, had high numbers of tweets but were at the same level as other regions in west Japan in terms of the population ratio. This suggests that this bias does not affect the data on Twitter, which has more users in urban areas. In addition, Hiroshima Prefecture and Nagasaki Prefecture, which were devastated in the past by the atomic bomb, stood out in tweeting from other regions in west Japan, suggesting that people may be recalling place names associated with radiation. In the case of Okinawa Prefecture, the US military base located there has been alleged to possess nuclear weapons many times in the past and seemed to be the name of a place brought up in connection to radiation.

Comparison of Fukushima Prefecture and Other Prefectures

As shown in Figure 7, the ratio of the number of tweets about Fukushima Prefecture and other prefectures shows an increasing trend, and the interest in Fukushima Prefecture has risen. Including other prefectures, the number of tweets was at 3 times the highest amount, and the right intercept of the linear approximation was at 2.5 times the highest amount.

The ratio of the average of the semantic orientation values for Fukushima Prefecture and other prefectures fell roughly below the value of 1 when unprecedented periods were excluded. As drawn by collinear approximation, the ratio gradually fell, which may be interpreted as the deepening of negative feelings about Fukushima Prefecture when compared to other regions. The weekly tweet number ratio on November 10 had not increased much compared to the previous week, but the semantic orientations about Fukushima largely trended positively. This is believed to have been influenced by the larger number of more positive tweets than average being spread stating “Sign an emergency petition to save the children of Fukushima” (semantic orientation value: –0.396). These tweets were not posted by bots or in the form of RTs.

As shown in Figures 1 and 3, the overall number of tweets decreased and, as a result, the overall semantic orientation values about radiation trended positively. However, as shown in Figures 7 and 8, the ratio of the number of tweets about Fukushima Prefecture compared to other prefectures increased, and it is thought that the decreasing ratio of semantic orientation values indicates that emotions sharply trended negatively toward Fukushima Prefecture in regard to radiation.

In Figure 9, we found that the RT group had more posts about negative feelings compared to the target group, as in the discussion of Figure 3 above. However, it is possible to confirm the same trend in tweets about regions. Surprisingly, while other tweets trended toward positive feelings, only the collinear approximation of the average trend in the semantic orientation values for the RT group relating to Fukushima Prefecture showed a negative trend, with the tendency to be broadcast with increasing negative feelings over time. This strongly suggests that bots and RTs disperse information that has more negative emotions, and users who come into contact with this information perceive matters relating to the radiation in Fukushima Prefecture with negative emotions, leading to an increased aversion among these people toward Fukushima Prefecture. These results support the hypothesis that people have become gripped by more negative feelings over time regarding Fukushima Prefecture as a disaster area associated with radiation, which may have influenced an increased aversion toward Fukushima in general.

Discussion

Principal Findings

This study was a unidirectional survey of the feelings that nationwide Twitter users had about each prefecture and radiation. It is felt that residents’ feelings toward a particular region, which is subject to harmful rumors, are important when elucidating and ameliorating the processes by which people’s aversions to Fukushima are increasing. The purpose of this study was to identify how information on radiation changed 1 year after the Fukushima nuclear accident with regard to different regions in Japan. We found that immediately after the accident, negative feelings about radiation trended positively over time, but bots and RTs were slow to do so compared to other tweets. We found that tweets associating Fukushima Prefecture with radiation clearly showed more negative feelings than those about other prefectures on average; they further trended as negative and increased in percentage over time. Tweets about radiation decreased overall, and feelings about radiation also trended positively. However, the fact that tweets about Fukushima Prefecture trended negatively while rising in percentage suggests that negative feelings toward Fukushima Prefecture were intensifying. We found that while the bots and RTs that were not about Fukushima Prefecture gradually trended toward positive feelings, the bots and RTs about Fukushima Prefecture trended toward negative feelings. These results point to the possibility that people’s aversions toward Fukushima Prefecture increased as a result of negative feelings that associate Fukushima Prefecture with radiation, as spread by bots and RTs. Signals about risk, such as health risks from radiation, are often amplified by individual and social processes, such as cultural groups and interpersonal networks, that amplify people's responses [29]. This supports the hypothesis that people have become gripped over time by a more negative impression of Fukushima Prefecture as a disaster area associated with radiation, which may have influenced an increased aversion toward Fukushima in general. To confirm this, tracking interaction between a bot and a person at an individual level should be performed as a next step. Additionally, it is well known that confirmation bias is amplified by the use of filter bubbles on social media [30,31]. This effect should be taken into account to analyze the impact of RTs and bots on people’s aversions.

Limitations

Gore et al showed that there can be significant geographic bias in the sentiment expressed in tweets over the same time period

[32]. Padilla et al showed that the sentiment expressed in tweets can be biased based on whether people are local or visiting an area and what other activities they have completed during the course of a day [33]. This study did not take this into account, as it conducted a one-way sentiment analysis of emotions directed toward Fukushima. In the future, in order to identify the process by which people's aversions increase, we want to clarify the feelings that people in a particular region have toward regions that are subject to harmful rumors, such as Fukushima. In a survey by Aoki et al [20], geotagged tweets made up only 0.25% of the target data; therefore, more comprehensive data are required. In addition, since there are deviations in the age composition and region of Twitter users' residences, the users are not necessarily representative of the nation as a whole. In this study, we determined the semantic orientation of tweets according to semantic orientation values toward words. Thus, the correct semantic orientation of tweets is not necessarily representative in terms of the context within tweets or the context based on their relationship to preceding and following tweets. Clearly sarcastic statements like "radioactivity is delicious" have positive semantic orientations due to the word "delicious," so these tweets are judged to have positive semantic orientations. It seems a technique is needed for correctly evaluating the semantic orientation in terms of both the written sentences and their context. It is well known that there is a normalcy bias in the responses of the public figures to a serious event that has occurred suddenly and unexpectedly [34]. In particular, when a severe nuclear accident occurs, affected people may very likely tweet simply to calm their own minds. As a result, their tweets may reflect not their feelings but their wishes. Additionally, in some cases, recall bias has led to an overestimation of the health risks from radiation, and tweets then expressed excessive aversion. In order to analyze the impact of these cognitive biases [35], it is necessary to evaluate the content of the tweet and the network.

Japanese speakers tend to skip words if their meaning is conveyed [36], which is significant since Twitter is limited to 140 characters. "Radioactive iodine" and "radioactive cesium" are frequently used radioactive isotopes that are often indicated simply with "iodine" and "cesium." "Cesium" is not a familiar word in daily life and likely indicates the radioactive isotope cesium. Further, feelings related to words that are often omitted when discussing radiation need to be surveyed instead of just "radiation," "radioactivity," and "radioactive substances."

Whether a tweet is from a bot is determined based on whether the term *bot* is found before or after the user's ID. Therefore, all bots cannot be accurately identified. In addition, we believe that tweets from accounts that repeatedly post the same information that are not advertising accounts, that do not follow this format, or that are not RTs need to be surveyed as part of the RT group or split off into a separate group. However, considerable effort is required to look up and verify each tweet one by one. Communities are formed by Twitter's *follow* function, and sharing and propagating information occurs through RTs.

In the future, it may be important to elucidate the process by which people's attitudes become fixed through a survey of how information is propagated by community networks and RTs as well as the feelings that people become gripped by when an aversion to Fukushima increases.

The dataset analysed during the current study is available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The scripts used in this research. analyze_store_tweets_area_name.py: get area names in tweets. geocoder.py: get full address from area names by yahoo geocoder API. Thus, it can get prefecture name. calc_tweets_SOV.py: calculate the semantic orientation value of each tweets.

[ZIP File (Zip Archive), 10 KB - jmir_v22i9e18662_app1.zip]

References

1. Thielen H. The Fukushima Daiichi nuclear accident—An overview. Health Phys 2012;103(2):169-174. [doi: 10.1097/hp.0b013e31825b57ec]
2. International Expert Working Group, World Health Organization (WHO). Health Risk Assessment From the Nuclear Accident After the 2011 Great East Japan Earthquake and Tsunami: Based on a Preliminary Dose Estimation. Geneva, Switzerland: World Health Organization; 2013. URL: https://apps.who.int/iris/rest/bitstreams/110732/retrieve [accessed 2020-08-17]
3. United Nations Scientific Committee on the Effects of Atomic Radiation. Sources, Effects, and Risks of Ionizing Radiation: UNSCEAR 2013 Report. Volume I: Report to the General Assembly. Scientific Annex A: Levels and Effects of Radiation Exposure Due to the Nuclear Accident After the 2011 Great East-Japan Earthquake and Tsunami. New York, NY: United Nations; 2014. URL: https://www.unscear.org/docs/publications/2013/UNSCEAR_2013_Report_Vol.I.pdf [accessed 2020-08-17]

4. The Follow-Up IAEA International Mission on Remediation of Large Contaminated Areas Off-Site the Fukushima Daiichi Nuclear Power Plant: Tokyo and Fukushima Prefecture, Japan, 14 – 21 October 2013. Vienna, Austria: International Atomic Energy Agency; 2014 Jan 23. URL: https://www.iaea.org/sites/default/files/final_report230114.pdf [accessed 2020-08-17]
5. Kudo D, Nakayachi K. Reputational risk caused by the Great East Japan Earthquake: A study on consumer factors leading to restrained buying. Jpn J Soc Psychol 2014 Aug 18;30(1):35-44 [FREE Full text] [doi: 10.14966/jssp.30.1_35]
6. Yoshida K, Hayashida N, Fukushima Y, Ohtsuru A, Ohba T, Hasegawa A, et al. Changes in radiological imaging frequencies in children before and after the accident at the Fukushima Daiichi Nuclear Power Plant in Fukushima Prefecture, Japan. Jpn J Radiol 2015 Oct;33(10):619-626. [doi: 10.1007/s11604-015-0464-8] [Medline: 26219903]
7. Miyazaki O. Has pediatric CT practice in Japan changed since the Fukushima nuclear disaster? Pediatr Radiol 2015 Sep;45(10):1571-1574. [doi: 10.1007/s00247-015-3295-y] [Medline: 25792153]
8. Information and Communications in Japan. Tokyo, Japan: Ministry of Internal Affairs and Communications, Japan; 2011. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2011/2011-index.html [accessed 2020-08-13]
9. Lu X, Brelsford C. Network structure and community evolution on Twitter: Human behavior change in response to the 2011 Japanese earthquake and tsunami. Sci Rep 2014 Oct 27;4:6773 [FREE Full text] [doi: 10.1038/srep06773] [Medline: 25346468]
10. Toriumi F, Sakaki T, Shinoda K, Kazama K, Kurihara S, Noda I. Information sharing on Twitter during the 2011 catastrophic earthquake. In: Proceedings of the 22nd International World Wide Web Conference (WWW '13). 2013 Presented at: 22nd International World Wide Web Conference (WWW '13); May 13-17, 2013; Rio de Janeiro, Brazil. [doi: 10.1145/2487788.2488110]
11. Jung JY. Social media use and goals after the Great East Japan Earthquake. First Monday 2012 Aug 09;17(8):1 [FREE Full text] [doi: 10.5210/fm.v17i8.4071]
12. Kanda R, Tsuji S, Yonehara H. Text mining analysis of radiological information from newspapers as compared with social media on the Fukushima nuclear power plant accident. J Disaster Res 2014 Sep 01;9(sp):690-698. [doi: 10.20965/jdr.2014.p0690]
13. Ikegami Y, Kawai K, Namihira Y, Tsuruta S. Topic and opinion classification based information credibility analysis on Twitter. In: Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics. 2013 Presented at: 2013 IEEE International Conference on Systems, Man, and Cybernetics; October 13-16, 2013; Manchester, UK p. 4676-4681. [doi: 10.1109/smc.2013.796]
14. Takamura H, Inui T, Okumura M. Extracting semantic orientations of words using spin model. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL '05). 2005 Jun Presented at: 43rd Annual Meeting of the Association for Computational Linguistics (ACL '05); June 25-30, 2005; Ann Arbor, MI p. 133-140 URL: https://www.aclweb.org/anthology/P05-1.pdf [doi: 10.3115/1219840.1219857]
15. Wang J, Kim S. Searching for new directions for energy policy: Testing the cross-effect of risk perception and cyberspace factors on online/offline opposition to nuclear energy in South Korea. Sustainability 2019 Mar 05;11(5):1368 [FREE Full text] [doi: 10.3390/su11051368]
16. Nakayama C, Sato O, Sugita M, Nakayama T, Kuroda Y, Orui M, et al. Lingering health-related anxiety about radiation among Fukushima residents as correlated with media information following the accident at Fukushima Daiichi Nuclear Power Plant. PLoS One 2019;14(5):e0217285 [FREE Full text] [doi: 10.1371/journal.pone.0217285] [Medline: 31150483]
17. Takahashi M. Risk perception and risk reduction strategies of mothers after the Fukushima nuclear disaster: A questionnaire survey in the radioactively contaminated areas of Chiba and Ibaraki Prefectures [article in Japanese]. Disaster Recovery Res 2015 Sep 30;7:45-68 ISSN 18837530 [FREE Full text]
18. Kim S, Kim S. Impact of the Fukushima nuclear accident on belief in rumors: The role of risk perception and communication. Sustainability 2017 Nov 27;9(12):2188 [FREE Full text] [doi: 10.3390/su9122188]
19. Yagahara A, Hanai K, Hasegawa S, Ogasawara K. Relationships among tweets related to radiation: Visualization using co-occurring networks. JMIR Public Health Surveill 2018 Mar 15;4(1):e26 [FREE Full text] [doi: 10.2196/publichealth.7598] [Medline: 29549069]
20. Aoki T, Suzuki T, Yagahara A, Hasegawa S, Tsuji S, Ogasawara K. Analysis of the regionality of the number of tweets related to the 2011 Fukushima nuclear power station disaster: Content analysis. JMIR Public Health Surveill 2018 Dec 18;4(4):e70 [FREE Full text] [doi: 10.2196/publichealth.7496] [Medline: 30563815]
21. Kudo T, Yamamoto K, Matsumoto Y. Applying conditional random fields to Japanese morphological analysis. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2004 Presented at: 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP); July 25-26, 2004; Barcelona, Spain p. 230-237 URL: https://www.aclweb.org/anthology/W04-3230.pdf
22. Sato T. Neologism dictionary based on the language resources on the web for Mecab. GitHub. 2015. URL: https://github.com/neologd/mecab-ipadic-neologd [accessed 2020-08-13]
23. Yahoo! Geocoder API. Yahoo! Japan. URL: https://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/geocoder.html [accessed 2020-08-13]
24. Statistics Bureau of Japan. Population by sex and sex ratio for prefectures: Total population, Japanese population, October 1, 2011. e-Stat. Tokyo, Japan: Ministry of Internal Affairs and Communications; 2012 Apr 17. URL: https://www.e-stat.go.jp/

en/stat-search/files?page=1&layout=datalist&toukei=00200524&tstat=000000090001&cycle=7&year=20110&month=0&tclass1=000001011679&stat_infid=000012976596 [accessed 2020-08-13]
25. Mendoza M, Poblete B, Castillo C. Twitter under crisis: Can we trust what we RT? In: Proceedings of the First Workshop on Social Media Analytics (SOMA '10). 2010 Presented at: First Workshop on Social Media Analytics (SOMA '10); July 25, 2010; Washington, DC p. 71-79 URL: https://snap.stanford.edu/soma2010/papers/soma2010_11.pdf [doi: 10.1145/1964858.1964869]
26. Nuclear Safety Commission Japan. Organ dose of I-131 for infant at UTC=2011-03-23_15h [article in Japanese]. 2019. URL: https://radioactivity.nsr.go.jp/ja/contents/5000/4810/24/1305799_0324.pdf [accessed 2020-08-13]
27. Visualization of environmental radiation levels in Kanto area (Japan). micro sievert. URL: http://microsievert.net/html5.html [accessed 2020-08-13]
28. Wilson MJ, Fukurai H, Maruta T. Japan and Civil Jury Trials: The Convergence of Forces. Cheltenham, UK: Edward Elgar publishing; 2015.
29. Kasperson R, Renn O, Slovic P, Brown HS, Emel J, Goble R, et al. The social amplification of risk: A conceptual framework. Risk Anal 1988 Jun;8(2):177-187. [doi: 10.1111/j.1539-6924.1988.tb01168.x]
30. Pariser E. The Filter Bubble: What the Internet Is Hiding From You. New York, NY: Penguin Press; 2011.
31. Quattrociocchi W, Scala A, Sunstein CR. Echo chambers on Facebook. SSRN 2016:1-15 [FREE Full text] [doi: 10.2139/ssrn.2795110]
32. Gore RJ, Diallo S, Padilla J. You are what you tweet: Connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
33. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
34. Omer H, Alon N. The continuity principle: A unified approach to disaster and trauma. Am J Community Psychol 1994 Apr;22(2):273-287. [doi: 10.1007/BF02506866] [Medline: 7977181]
35. Jargin S. Some aspects of the Fukushima Daiichi nuclear accident. Environ Dis 2020;5(1):16. [doi: 10.4103/ed.ed_6_20]
36. Hinds J. Japanese conversation, discourse structure, and ellipsis. Discourse Process 2009 Nov 11;3(3):263-286. [doi: 10.1080/01638538009544490]

Abbreviations

CT: computed tomography
JST: Japan Standard Time
NTT: Nippon Telegraph and Telephone
QT: quote tweet
RT: retweet
TEPCO: Tokyo Electric Power Company

Edited by G Eysenbach; submitted 11.03.20; peer-reviewed by H Yasuda, R Gore; comments to author 13.04.20; revised version received 08.06.20; accepted 11.06.20; published 02.09.20.

Please cite as:
Hasegawa S, Suzuki T, Yagahara A, Kanda R, Aono T, Yajima K, Ogasawara K
Changing Emotions About Fukushima Related to the Fukushima Nuclear Power Station Accident—How Rumors Determined People's Attitudes: Social Media Sentiment Analysis
J Med Internet Res 2020;22(9):e18662
URL: https://www.jmir.org/2020/9/e18662
doi:10.2196/18662
PMID:32876574

Original Paper

Social Media Listening to Understand the Lived Experience of Presbyopia: Systematic Search and Content Analysis Study

James S Wolffsohn[1*], BSc, PgCert, PGDip, MBA, PhD; Claudia Leteneux-Pantais[2*], MSc; Sima Chiva-Razavi[2*], MSc; Sarah Bentley[3*], BSc, MSc; Chloe Johnson[3*], MSc; Amy Findley[3*], BSc, MSc; Chloe Tolley[3*], BSc; Rob Arbuckle[3*], MA, MSc; Jyothi Kommineni[4*], MPharm; Nishith Tyagi[4*], BA, MS, MBA

[1]Optometry and Vision Science, College of Health and Life Sciences, Aston University, Birmingham, United Kingdom

[2]Novartis Pharma AG, Basel, Switzerland

[3]Patient-Centered Outcomes, Adelphi Values Ltd, Bollington, United Kingdom

[4]Novartis Business Services, Product Lifecycle Services, Hyderabad, India

[*]all authors contributed equally

Corresponding Author:
Sarah Bentley, BSc, MSc
Patient-Centered Outcomes
Adelphi Values Ltd
Adelphi Mill
Grimshaw Lane
Bollington, SK10 5JB
United Kingdom
Phone: 44 1625576815
Email: sarah.bentley@adelphivalues.com

Abstract

Background: Presbyopia is defined as the age-related deterioration of near vision over time which is experienced in over 80% of people aged 40 years or older. Individuals with presbyopia have difficulty with tasks that rely on near vision. It is not currently possible to stop or reverse the aging process that causes presbyopia; generally, it is corrected with glasses, contact lenses, surgery, or the use of a magnifying glass.

Objective: This study aimed to explore how individuals used social media to describe their experience of presbyopia with regard to the symptoms experienced and the impacts of presbyopia on their quality of life.

Methods: Social media sources including Twitter, forums, blogs, and news outlets were searched using a predefined search string relating to symptoms and impacts of presbyopia. The data that were downloaded, based on the keywords, underwent manual review to identify relevant data points. Relevant posts were further manually analyzed through a process of data tagging, categorization, and clustering. Key themes relating to symptoms, impacts, treatment, and lived experiences were identified.

Results: A total of 4456 social media posts related to presbyopia were identified between May 2017 and August 2017. Using a random sampling methodology, we selected 2229 (50.0%) posts for manual review, with 1470 (65.9%) of these 2229 posts identified as relevant to the study objectives. Twitter was the most commonly used channel for discussions on presbyopia compared to forums and blogs. The majority of relevant posts originated in Spain (559/1470, 38.0%) and the United States (426/1470, 29.0%). Of the relevant posts, 270/1470 (18.4%) were categorized as posts written by individuals who have presbyopia, of which 37 of the 270 posts (13.7%) discussed symptoms. On social media, individuals with presbyopia most frequently reported experiencing difficulty reading small print (24/37, 64.9%), difficulty focusing on near objects (15/37, 40.5%), eye strain (12/37, 32.4%), headaches (9/37, 24.3%), and blurred vision (8/37, 21.6%). 81 of the 270 posts (30.0%) discussed impacts of presbyopia—emotional burden (57/81, 70.4%), functional or daily living impacts (46/81, 56.8%), such as difficulty reading (46/81, 56.8%) and using electronic devices (21/81, 25.9%), and impacts on work (3/81, 3.7%).

Conclusions: Findings from this social media listening study provided insight into how people with presbyopia discuss their condition online and highlight the impact of presbyopia on individuals' quality of life. The social media listening methodology can be used to generate insights into the lived experience of a condition, but it is recommended that this research be combined with prospective qualitative research for added rigor and for confirmation of the relevance of the findings.

(J Med Internet Res 2020;22(9):e18306) doi:10.2196/18306

KEYWORDS

presbyopia; near vision; social media; social media listening; infodemiology

Introduction

Presbyopia is the most common physiological change occurring in the adult eye. In presbyopia, the elasticity of the lens begins to deteriorate, causing universal near vision impairment with increasing age [1,2]. It is estimated that presbyopia is experienced in over 80% of people aged 40 years or above in western countries [1]. There is currently no way to stop or reverse the normal aging process that causes presbyopia; near vision impairments associated with presbyopia are typically corrected with glasses, contact lenses, surgery, or the use of a magnifying glass [3]. In 2015, it was estimated that presbyopia affected approximately 1.8 billion people (25% of the world population), with 826 million experiencing near vision impairments because they had no (or inadequate) vision correction [4]. As a result, the global unmet need for presbyopia correction methods in 2015 was estimated to be 45% [4].

Presbyopia impacts many domains of quality of life including difficulty with near vision tasks, such as reading printed text, using a smartphone, or threading a needle [5,6]. Difficulty performing these tasks can worsen over time and is typically more significant if the lighting is not optimal [2,6]. Such problems can, in turn, impact work productivity, social interactions, household activities, and emotional well-being [7,8]. Emotional impacts associated with presbyopia include having to rely more on others, feeling ashamed, and feeling embarrassed due to poor vision [8]. Individuals who do not wear glasses or contact lenses may experience headaches and eye strain due to their difficulty focusing on objects [9]. Some individuals with presbyopia describe holding objects (eg, a menu) progressively farther away from their eyes in order to be able to focus on them [1].

Minimal qualitative research [6] into the lived experience of presbyopia (not limited to a single form of correction) has been published, and what has been published focuses generally on refractive errors rather than being specific to presbyopia. Traditionally qualitative research is conducted via interviews or focus groups. However, social media sources can now be utilized to provide qualitative data for a large sample across multiple countries [10]. Approximately 68% of all US adults use Facebook, while over 20% of US adults use Instagram, Pinterest, LinkedIn, and Twitter [11]. Peer-to-peer exchange of health information is popular online; a recent study found that 51% of Americans had used social networking, and a further 66% had looked at blogs for health information [12].

Social media reviews are increasingly being used to investigate the patient experience of health conditions such as dry eye and chronic obstructive pulmonary disease [13-15]. Other studies have used social media listening to investigate the emotional impact of caregivers who look after patients with leukemia [16] and to identify posts discussing potential misuse or nonmedical use of antidepressants [17]. One study found social media listening generated more concepts relevant to the lived experience of specific conditions in comparison to those generated by concept elicitation interviews and the methodology of group concept mapping [18]. It has been theorized that some individuals may feel more comfortable discussing socially embarrassing symptoms online than discussing them in an interview or focus group setting [18,19]. It is acknowledged that the depth of data collected by social media reviews can be limited in comparison to those collected by methods such as concept elicitation interviews, particularly in cases where character counts are limited for each post (eg, Twitter with a 140 character limit per post, at the time of this research) [18]. To the authors' knowledge, there is currently no published research exploring how presbyopia is discussed on social media; this study aims to address this gap.

This study aimed to explore how individuals with presbyopia use social media to describe their experiences. Specifically, the study explored how the social media population described the visual and nonvisual symptoms that they experience as a result of presbyopia and the impacts that presbyopia has on their health-related quality of life (HRQoL). A secondary objective was to explore individuals' experiences of diagnosis and treatment or vision correction method options.

Methods

Study Design

This study was a noninterventional retrospective analysis of social media data available in the public domain.

Search Strategy

A predefined search string was used to identify social media posts and discussions that were relevant to the lived experience of presbyopia. The search string terms were initially identified through a literature search and review of online patient forums. Two approaches were taken to develop the final search string. The first approach was to search social media sources for indication-related keywords only (in English and translated to local languages). This first approach helped identify any further associated symptom or impact terms. The second approach involved searching social media sources using a combination of indication-related keywords and other disease journey–related keywords such as *symptoms*, *diagnosis*, and *vision correction*. Based on the results obtained from these two preliminary searches, appropriate terms were included in the search strategy.

The resulting search string (Multimedia Appendix 1) contained terminology related to the symptoms experienced, impacts of presbyopia, and vision correction options. Three key generic presbyopia search terms (*presbyopia*, *long sightedness*, and *elasticity*) were also translated into five additional languages (German, French, Spanish, Italian, and Japanese) and included in the search query. Boolean operators (AND, OR) were used to combine the keywords into a single search string.

Data Collection

Sales Force Social Studio database [20] was used to conduct the searches. The search terms (listed in Multimedia Appendix 1) were inputted into the database, and the Sales Force Social Studio software identified posts that matched the search terms across the following media channels: Twitter, forums, blogs, and news posts. Relevant forums and blog posts were identified on online community websites and discussion boards such as Medhelp and Optiboard. News posts were identified from general news websites across different countries.

Social media posts were searched in and identified from the United States, France, Germany, Italy, Spain, the United Kingdom, and Japan. All relevant posts were downloaded, and a random sampling methodology was employed using a simple randomization technique to reduce the number of posts by 50%. The number of posts were reduced by 50% to ensure it was a manageable amount for the research team to manually review. Posts that were originally non-English were translated into to English using Google translate. Posts were then manually reviewed by researchers against predefined criteria (Multimedia Appendix 2) to ensure they were relevant to the study objectives.

Data Analysis

Relevant posts were automatically tagged by channel type (eg, Twitter, Forums, Blogs, or News) and, where possible, manually categorized by type of stakeholder (eg, individuals with presbyopia, physicians, researchers, clinics, support groups, company handles, and media), sentiment (positive, neutral, or negative), key themes of discussion, and lived experience. The stakeholder category was established through the grammatical tense of the post. For example, a post that discussed presbyopia in the first person such as "I have difficulty reading because of my presbyopia" was categorized as an individual with presbyopia. Due to the amount of data evaluated, it was not possible to review every profile to confirm stakeholder categorization, and as expected, for many posts that were reviewed, the stakeholder category could not be deduced based on the post content. However, these findings give an indication of the approximate proportions of stakeholders discussing presbyopia online.

State-of-mind analysis was used to explore how individuals felt about presbyopia [21]. State-of-mind analysis is an analyst interpretation of the verbatim and focused on the key psychological and emotional state of audience. In order to maintain the consistency of the interpretation, the analysis went through three levels of review (review from analyst, quality control manager, and project manager).

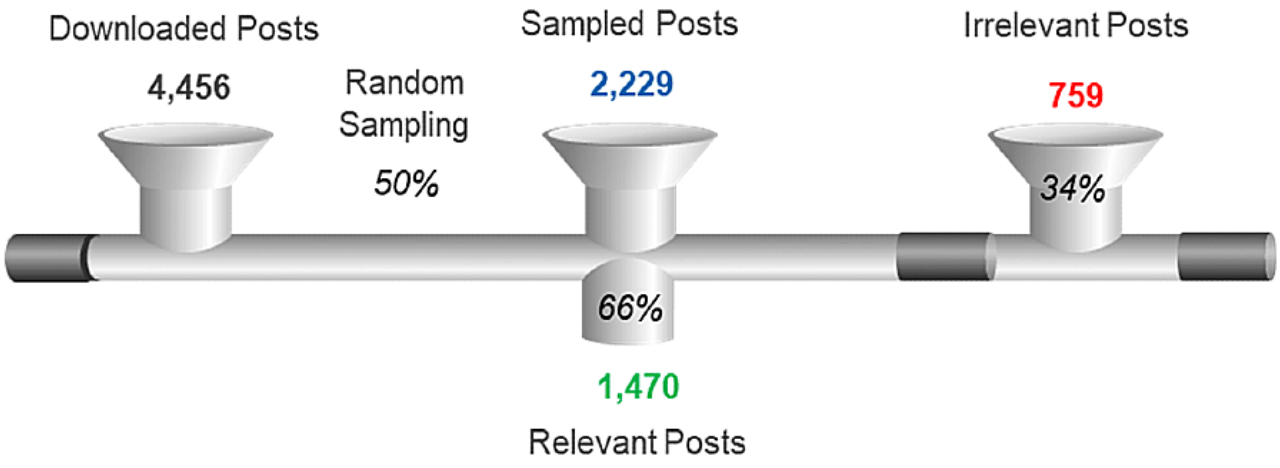
To illustrate and exemplify the results reported, the authors have included a number of quotations from the social media sources. The authors took several steps to anonymize these publicly reported quotations to ensure direct quotations from posts cannot be traced online. First, any direct quotations that were originally non-English were translated and as a result are not identifiable. Second, the username of the post's author was removed. Third, any individual patient or caregiver information was anonymized. Finally, any originally English quotations were edited to ensure these cannot be identified online. To edit quotations, the sentence structure and certain words were replaced by synonyms. This was done in a way that ensured the meaning of the quotation was retained. The edited quotations were reviewed by the research team, and consensus was reached.

Results

Overview of Data

Social media posts were collated from May 2017 to August 2017. Figure 1 shows the data relevancy process that was conducted to identify the most relevant social media posts. A total of 4456 social media posts were obtained from the initial search. The random sampling methodology selected 2229 (50.0%) posts for manual review, with 1470 (65.9%) of these 2229 posts identified as relevant to the study objectives.

Figure 1. Process of analysis of post relevancy.



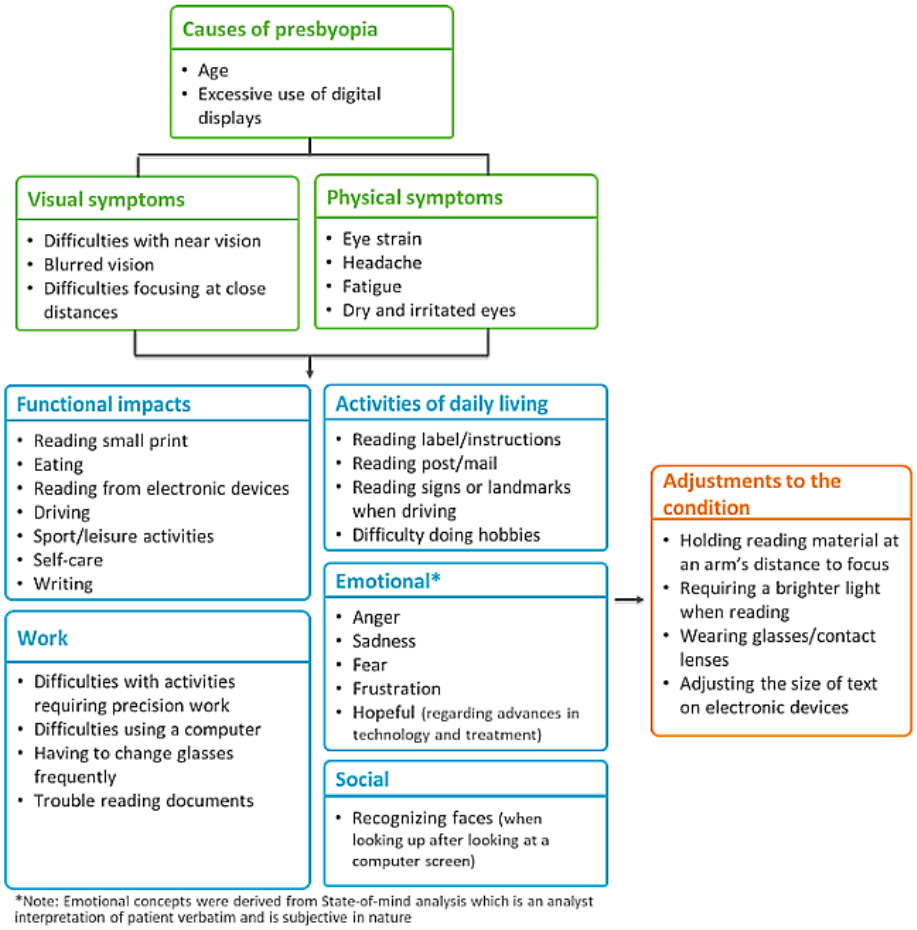
Twitter emerged as the most commonly used social media channel (Multimedia Appendix 3) and was the source of the majority of relevant posts (1182/1470, 80.4%), compared with the amount of posts derived from forums (164/1470, 11.2%), blogs (109/1470, 7.4%), and news sources (15/1470, 1.0%). Around 62.0% (733/1182) of discussions on Twitter were focused on disease awareness with symptoms of presbyopia being one of the most prominently tweeted topics. Forums

contained the highest level of conversation among the social media population. Discussions on forums were predominantly about quality of life (60/164, 36.6%) and treatment or correction methods (98/164, 59.8%). Discussions on blogs were primarily centered on lifestyle (23/109, 21.1%) and treatment or correction method options (66/109, 60.6%).

The country of origin of the relevant posts was analyzed (Multimedia Appendix 4). The majority of posts originated in Spain (559/1470, 38.0%) or the United States (426/1470, 29.0%). Fewer posts originated from France (147/1470, 10.0%), the United Kingdom (118/1470, 8.0%), Japan (103/1470, 7.1%), Italy (88/1470, 6.0%), and Germany (15/1470, 1.0%).

A total of 270 posts (270/1470, 18.4%) were categorized as posts from individuals who had presbyopia based on the language used (eg, "I am diagnosed"; "I have this condition"). Based on limited data (27 discussions), most of the individuals (19/27, 70.4%) were aged between 40 and 70 years, as would be expected given the typical age range of presbyopia. A peak in social media discussions was observed from May 29, 2017 to June 19, 2017, due to a number of tweets about presbyopia awareness in Spain, the United States, and Japan. In the week of July 31, 2017 to August 7, 2017, a progressive lens was launched and led to an increase in tweets from individuals with presbyopia and health care professionals. A conceptual model (see Figure 2) was formulated to provide an overview of the key causes, symptoms, and impacts that were reported by the presbyopia population on social media. The model also summarizes the key adjustments described to help individuals with presbyopia better cope with the impacts of presbyopia.

Figure 2. Conceptual model of social media listening findings.



Symptoms of Presbyopia

Of the 270 posts categorized as having been posted by individuals with presbyopia (based on the language used), 37 posts referred to symptoms of presbyopia (37/270, 13.7%). The majority of data reported by presbyopic individuals relating to symptoms of presbyopia were obtained via Twitter (21/37, 56.8%), although some discussions were identified in forums (10/37, 27.0%) and blogs (5/37, 13.5%). The social media population appeared to be aware of their near vision being impaired and discussed symptoms such as difficulty reading small print (24/37, 64.9%), diminished ability to focus on near objects (15/37, 40.5%), eye strain (12/37, 32.4%), headaches (9/37, 24.3%), and blurred vision (8/37, 21.6%).

> *I have no difficulty reading The Washington Post but I have to strain to read print that is smaller in size.*

Impacts of Presbyopia

A total of 81 posts referred to impacts that presbyopia has on an individual's life (81/270, 30.0%; see Table 1). The majority of data related to impacts of presbyopia were obtained via forums (42/81, 51.9%) and Twitter (30/81, 37.0%), and fewer were obtained from blogs (9/81, 11.1%). The social media

population discussed functional or daily life impacts of presbyopia including difficulties reading (46/81, 56.8%), difficulty using digital devices (21/81, 25.9%), and limitations in sport and leisure activities (8/81, 9.9%). Individuals with presbyopia discussed a number of difficulties related to reading including reading text in a small font size (16/46, 34.8%), reading printed text in books (13/46, 28.3%), and reading at a distance when presbyopia becomes more severe (9/46, 19.6%).

Table 1. Impacts of presbyopia reported on social media.

Impact	Posts[a] (n=81), n (%)
Emotional	57 (70.4)
Reading	46 (56.8)
Using digital devices	21 (25.9)
Sports and leisure activities	8 (9.9)
Work	3 (3.7)
Driving	2 (2.5
Recognizing people	2 (2.5)
Putting make-up on	1 (1.2)
Eating	1 (1.2)

[a]Posts could describe more than one.

> *I used to be able to read the comics from a distance. Actually, I can still do that. What I can't do is read them if they're right in front of me.*
>
> *The letters "dance," there are blurring even when looking from a distance, there is more visual fatigue and we become more dependent on a good light to be able to see the details up close. When the distance is not enough or is uncomfortable is when you start to use glasses of near vision.*

The social media population also reported difficulties using digital devices (21/81, 25.9%), specifically mobile phones (11/21, 52.4%), televisions (4/21, 19.0%), computers (4/21, 19.0%), and laptops (3/21, 14.3%).

> *The worst of all is that I can not see "the mobile" without my presbyopic glasses.*

Impacts of presbyopia relating to specific sports and leisure activities were discussed on social media (8/81, 9.8%) including difficulties diving (4/8, 50.0%), cycling (2/8, 25.0%), skiing (1/8, 12.5%), walking (1/8, 12.5%), and playing piano (1/8, 12.5%).

> *I find for walking that if I don't wear contacts I find it harder judging where the ground is on rough terrain, so the contacts help for that too.*

Other impacts of activities of daily living reported by individuals on social media included driving (2/81, 2.5%), recognizing people (2/81, 2.5%), putting makeup on (1/81, 1.2%), and eating (1/81, 1.2%).

> *People may refer to it as "reading vision," but it is the vision used for other near activities, such as eating, putting on make-up*

In terms of work impacts, 3 posts (3/81, 3.7%) discussed difficulties using a computer (1/3, 33.3%), recognizing people (1/3, 33.3%), and using a needle (1/3, 33.3%).

> *Presbyopia affects an individual's ability to enjoy and carry out a range of near vision activities – from reading, writing to precision tasks required in the workplace.*

A total of 57 posts (57/81, 70.4%) relating to the emotional impact of presbyopia were identified in this analysis, with individuals with presbyopia typically reporting feelings of sadness (35/57, 61.4%), happiness due to a positive treatment experience, new reading technology, or not needing to depend on glasses (9/57, 15.8%), anger (7/57, 12.3%), and fear (6/57, 10.5%).

> *I have a little difficulty to tell the difference between the 3 and 8, and the 6 and 8 !!!! It's very annoying!!!*

Adjustments to Presbyopia

A number of posts discussed adjustments that individuals made to help them cope with the effects of presbyopia (154/270, 57.0%). These adjustments were primarily using glasses (87/154, 56.5%) or contact lenses (59/154, 38.3%). However, a few also referred to adjusting the size of text on an electronic device (4/154, 2.6%), holding reading material farther away than an arm's length (2/154, 1.3%), and requiring a bright light (2/154, 1.3%).

> *I feel that my arms are now too short, and at the same time I find it hard to view from far, it could be presbyopia.*

Experience of Diagnosis of Presbyopia

A total of 51 posts (51/270, 18.9%) discussed diagnosis of presbyopia. Of these, 90.2% (46/51) reported that they were diagnosed by eye examination, and 9.8% (5/51) reported that they self-diagnosed their presbyopia. Out of those who reported they were diagnosed by eye examination, these included tests of visual acuity (14/46, 30.4%), retinal examination (13/46, 28.3%), slit lamp (12/46, 26.1%), a visual field test (11/46,

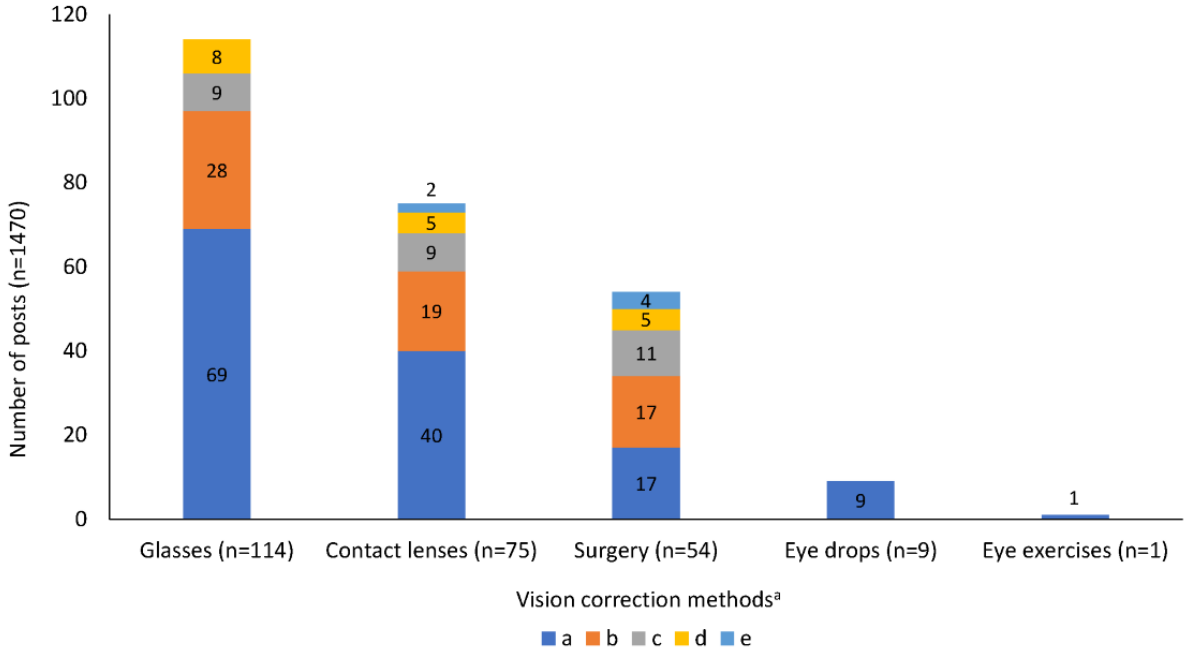
23.9%), evaluation of eye muscle integrity (11/46, 23.9%), refraction (2/46, 4.3%), and other tests (8/46, 17.4%).

> *It is the easiest way to determine presbyopia as a presbyopia based on whether you can read the letters of the newspaper from the position 30 cm away from the eyes or can not read.* [Self diagnosis]

Experience of Vision Correction Methods for Presbyopia

Figure 3 provides an overview of the vision correction methods reported in the social media posts. The most frequently reported vision correction options included glasses (114/1470 (7.8%)—of which 60.5% (69/114) were not specified further, 24.6% (28/114) were referred to as reading glasses, 7.9% (9/114) were bifocal glasses, and 7.0% (8/114) were progressive glasses—and contact lenses (75/1470, 5.1%)—of which 53.3% (40/75) were not specified further, 25.3% (19/75) were progressive, 12.0% (9/75) were bifocal, 6.7% (5/75) were monovision, and 2.7% (2/75) were described as monofocal. Other vision correction options discussed included surgery (54/1470, 36.7%)—which included 44.7% (17/38) posts that did not specify the type of surgery, 44.7% (17/38) posts that specified Lasik surgery, 28.9% (11/38) posts that specified corneal inlays, 13.2% (5/38) posts that specified intraocular lens surgery, 10.5% (4/38) posts that specified photorefractive keratectomy—and eye drops, as part of a clinical trial (9/1470, 0.6%), or eye exercises (1/1470, 0.1%).

Figure 3. Vision correction methods reported in social media posts.



[a]Legend key: **Glasses**: a (unspecified), b (reading glasses), c (bifocal glasses), d (progressive glasses). **Contact lenses:** a (unspecified), b (progressive), c (bifocal), d (monovision), e (monofocal). **Surgery:** a (unspecified), b (Lasik surgery), c (corneal inlays), d (intraocular lens surgery), e (photorefractive keratectomy). **Eye drops**: a (unspecified). **Eye exercises:** a (unspecified).

> *I bought a pair of bifocal glasses focusing to both distances for presbyopia.*

A number of impacts were found to be associated with current vision correction options. Impacts associated with wearing glasses included feeling "fed up" with having to use reading glasses, feeling unhappy with varifocal glasses, having to remove glasses or look under them while doing close work. Impacts associated with wearing contact lenses included having to have an additional pair of glasses to be able to see middle vision.

> *The varifocals have worked very well for most things apart from playing the piano... a very specific middle distance when reading music, so I have another pair of glasses just for that. If I need to do very near work, like tweezing eyebrows or a manicure, I put my glasses on my head or peer from under them.*

Discussion

Overview of Findings

The data collected in this study identified key concepts relevant to individuals with presbyopia and provided insight into how this condition is discussed online by multiple stakeholders across multiple countries. Key symptoms and vision impairment problems discussed by individuals online included difficulty reading small print, diminished ability to focus on near objects, and eye strain. Impacts of presbyopia discussed in the social media posts included difficulty reading, difficulty using electronic devices, and difficulty taking part in sport and leisure activities. Individuals with presbyopia also discussed the emotional impact of presbyopia with most expressing sadness.

Value of Findings

The findings from this study provide qualitative insights into the lived experience of presbyopia, where there is currently limited published qualitative research. Nevertheless, the symptoms discussed by this social media population were

consistent with those reported in other qualitative studies, specifically difficulties with near vision and blurred vision [5,6,8]. The findings obtained in this study also reflect those described in other qualitative studies that have explored the impact of refractive error (including individuals with presbyopia) on individuals' daily lives [1,5-9,22]. Difficulties carrying out tasks that require near vision have also been reported in a number of other qualitative studies [5,6,8,23,24]; these include reading small print, using digital devices such as a mobile phones, self-care (for example, putting make-up on), hobbies that require detailed near vision (for example, sewing or weaving), driving, watching television, difficulties with sports and exercise, and difficulties preparing food.

Emotional impacts associated with presbyopia reported in other qualitative studies [6,8] included requiring more help from others, feeling ashamed or embarrassed, feeling scared, depressed, and isolated. Thus, while there is overlap between the published literature and the findings reported here (particularly in terms of depression and sadness), there were also differences, with shame and embarrassment not emerging from the social media review. Impacts on work have also been associated with presbyopia in the literature [6]. Thus, in general, the findings in this social media listening study corroborate other findings in the literature, arguably providing greater confidence that the findings can be generalized across populations, given the large sample size of the social media listening study, which included posts from individuals from a range of countries around the world. This study highlights the value of social media listening for identifying and confirming relevant quality of life concepts for any given disease.

The findings from this social media listening study have contributed to the development of a conceptual model of presbyopia and helped inform development of an interview guide for prospective, in-depth qualitative research and to inform the modification of a patient-reported outcome measure to assess near vision functioning in presbyopia. Knowledge of how individuals discuss their presbyopia and the key concepts associated with their condition could aid communication between health care professionals and their patients and provide a basis for a patient-reported outcome measure that could be used to evaluate quality of life. The findings add to the published evidence regarding the lived experience of presbyopia and are of value to inform the design of clinical trials and other research studies to ensure adequate measurement of the identified visual functioning and quality of life concepts.

Study Limitations

The social media posts were analyzed in two ways—first by automated analysis and second by manual analysis. The benefit of automated analysis was that it allowed large amounts of data to be analyzed quickly and efficiently. This methodology was used to reduce the large amount of posts to a more manageable size by identifying and dismissing irrelevant posts. Using automated analysis, however, meant that some intricacies of human expression may not have been captured, and relevant posts may have been dismissed.

As the data were retrospectively collected from social media posts in the public domain, the only information about participants that could, generally, be obtained was the country of origin. Demographic or clinical information about the social media population in most instances could not be obtained; therefore, there was no way to confirm that the individuals truly had a diagnosis of presbyopia [25]. As such, it is not certain that all posts were from individuals with presbyopia (or other relevant stakeholders), and the authors acknowledge that some data may be incorrectly categorized.

When using social media as research media, certain biases need to be considered. People who post on social media represent a biased sample which may not be representative of the whole population of interest. Older people are often underrepresented on social media [26]; this needs to be considered when researching a condition such as presbyopia which becomes more prevalent with increasing age. Internet usage has increased in the older population over the last 20 years, but research shows that it is still used 20% less frequently by older populations than by younger populations and its use varies vastly based on socioeconomic factors, with lower economic status older populations being less active on the internet [27]. This study only found a small proportion of posts about quality of life from the total number of posts identified during the search. As presbyopia increases in severity with age, it is possible that individuals whose quality of life is most affected by their presbyopia are of an age where social media use is less common [26]. An alternate explanation could be that individuals with presbyopia do not discuss the impact presbyopia has on quality of life on social media. It is also acknowledged in the results that social media campaigns related to presbyopia awareness in Spain, the United States, and Japan likely led to increased tweets related to presbyopia in those countries relative to others. Thus, the data regarding the relative number of posts in each country should be interpreted with caution.

Social media listening research brings about its own ethical challenges, given individuals cannot formally consent to the use of their data. It can be argued that, since the social media posts are in the public domain, there is implicit consent that they can be read and used for research. Nevertheless, there is some guidance available regarding steps that can be taken to protect the privacy of individuals posting on social media platforms [28]. Recommendations include only collecting the data necessary to answer the research question, presenting data carefully to avoid participant identification, and understanding the risk of and not using direct text quotations from research participants. However, there is a lack of agreement on the correct way to conduct and present this form of research [29,30]. In particular, the introduction of new General Data Protection Regulation [30] guidelines in the European Union require privacy of any data by design and by default, but those guidelines do not currently offer specific solutions to ensure privacy is maintained in social media research. To ensure anonymity in this study, appropriate steps were taken to deidentify direct quotations. This included removing any identifiable information and editing quotations where necessary to ensure original quotations cannot be discovered online.

Conclusion

Presbyopia has a substantial impact on individuals' daily life. Individuals with presbyopia can experience particular difficulties with reading and electronic device usage that impact their daily lives and quality of life. Despite the limitations and considerations described, the social media listening methodology provides a quick and person-focused starting point to identify topics and relevant themes of interest that corroborate and provide greater ecological validity to other qualitative findings in the literature. These findings have been used to inform the design of more in-depth, prospective qualitative research and to support research outcomes. For data collected as part of a research outcomes study, the findings should be supplemented with further qualitative research (eg, interviews with individuals with presbyopia) for added rigor [31].

Authors' Contributions

JK and NT conducted data collection and data analysis. RA, CT, SB, CJ, and AF developed the first draft of the manuscript. All authors contributed to the interpretation of the findings and writing and reviewing of this manuscript.

Conflicts of Interest

JW has been engaged with Novartis on other projects and has received honoraria for speaker events, ad boards and other related activities. RA, CT, SB, and AF are employed by Adelphi Values Patient-Centered Outcomes, which received funding from Novartis Pharma AG to write this manuscript. CJ was employed by Adelphi Values Patient-Centered Outcomes at the time of the study. JK and NT are employed by Novartis Business Services – Product Lifecycle Services, which was funded by Novartis Pharma AG to conduct this research. CL-P and SC-R are employed by Novartis Pharma AG.

Multimedia Appendix 1
Social media search strategy.
[PDF File (Adobe PDF File), 115 KB - jmir_v22i9e18306_app1.pdf]

Multimedia Appendix 2
Predefined inclusion and exclusion criteria.
[PDF File (Adobe PDF File), 114 KB - jmir_v22i9e18306_app2.pdf]

Multimedia Appendix 3
Data source of relevant posts.
[PDF File (Adobe PDF File), 121 KB - jmir_v22i9e18306_app3.pdf]

Multimedia Appendix 4
Country of origin of relevant posts.
[PDF File (Adobe PDF File), 125 KB - jmir_v22i9e18306_app4.pdf]

References

1. Holden BA, Fricke TR, Ho SM, Wong R, Schlenther G, Cronjé S, et al. Global vision impairment due to uncorrected presbyopia. Arch Ophthalmol 2008 Dec 08;126(12):1731-1739. [doi: 10.1001/archopht.126.12.1731] [Medline: 19064856]
2. Charman WN. The eye in focus: accommodation and presbyopia. Clin Exp Optometry 2008 May;91(3):207-225. [doi: 10.1111/j.1444-0938.2008.00256.x]
3. Charman WN. Developments in the correction of presbyopia I: spectacle and contact lenses. Ophthalmic Physiol Opt 2014 Jan 10;34(1):8-29. [doi: 10.1111/opo.12091] [Medline: 24205890]
4. Fricke TR, Tahhan N, Resnikoff S, Papas E, Burnett A, Ho SM, et al. Global Prevalence of Presbyopia and Vision Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling. Ophthalmology 2018 Oct;125(10):1492-1499 [FREE Full text] [doi: 10.1016/j.ophtha.2018.04.013] [Medline: 29753495]
5. Goertz AD, Stewart WC, Burns WR, Stewart JA, Nelson LA. Review of the impact of presbyopia on quality of life in the developing and developed world. Acta Ophthalmol 2014 Sep;92(6):497-500 [FREE Full text] [doi: 10.1111/aos.12308] [Medline: 24910300]
6. Kandel H, Khadka J, Goggin M, Pesudovs K. Impact of refractive error on quality of life: a qualitative study. Clin. Experiment. Ophthalmol 2017 May 15;45(7):677-688. [doi: 10.1111/ceo.12954]
7. Frick KD, Joy SM, Wilson DA, Naidoo KS, Holden BA. The Global Burden of Potential Productivity Loss from Uncorrected Presbyopia. Ophthalmology 2015 Aug;122(8):1706-1710. [doi: 10.1016/j.ophtha.2015.04.014]
8. Lu Q, Congdon N, He X, Murthy GVS, Yang A, He W. Quality of life and near vision impairment due to functional presbyopia among rural Chinese adults. Invest Ophthalmol Vis Sci 2011 Jun 13;52(7):4118-4123. [doi: 10.1167/iovs.10-6353] [Medline: 21508106]

9. What is Presbyopia? American Academy of Opthalmology. 2017. URL: https://www.aao.org/eye-health/diseases/what-is-presbyopia [accessed 2017-10-26]
10. Snelson CL. Qualitative and Mixed Methods Social Media Research. International Journal of Qualitative Methods 2016 Mar 04;15(1):160940691562457. [doi: 10.1177/1609406915624574]
11. Greenwood S, Perrin A, Duggan M. Social media update 2016. Pew Research Center 2016;11:83. [doi: 10.1002/9781119085621.wbefs533]
12. Song H, Omori K, Kim J, Tenzek KE, Hawkins JM, Lin W, et al. Trusting Social Media as a Source of Health Information: Online Surveys Comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25. [doi: 10.2196/jmir.4193]
13. Cook N, Mullins A, Gautam R, Medi S, Prince C, Tyagi N, et al. Evaluating Patient Experiences in Dry Eye Disease Through Social Media Listening Research. Ophthalmol Ther 2019 Jun 3;8(3):407-420. [doi: 10.1007/s40123-019-0188-4]
14. Gutzwiller F, Gruenberger J, Kaur V, Pathak P, Mudumby A, Cook N. Generating patient insights in chronic obstructive pulmonary disease (COPD) with social media listening study. Eur Respiratory Soc 2018;52:PA2005. [doi: 10.1183/13993003.congress-2018.pa2005]
15. Cook NS, Kostikas K, Gruenberger J, Shah B, Pathak P, Kaur VP, et al. Patients' perspectives on COPD: findings from a social media listening study. ERJ Open Res 2019 Feb 10;5(1):00128-2018. [doi: 10.1183/23120541.00128-2018]
16. Gabriel S, Phillips K, Slater M, Zhou J, Pathak A, Lee L. The emotional journey among caregivers of patients with leukemia: The caregivers' perspective using social media listening. J Clin Oncol 2017 May 20;35(15_suppl):e18568. [doi: 10.1200/jco.2017.35.15_suppl.e18568]
17. Anderson L, Bell HG, Gilbert M, Davidson JE, Winter C, Barratt MJ, et al. Using Social Listening Data to Monitor Misuse and Nonmedical Use of Bupropion: A Content Analysis. JMIR Public Health Surveill 2017 Feb 01;3(1):e6. [doi: 10.2196/publichealth.6174]
18. Humphrey L, Willgoss T, Trigg A, Meysner S, Kane M, Dickinson S, et al. A comparison of three methods to generate a conceptual understanding of a disease based on the patients' perspective. J Patient Rep Outcomes 2017 Dec 19;1(1):9 [FREE Full text] [doi: 10.1186/s41687-017-0013-6] [Medline: 29757313]
19. Sjödahl Hammarlund C, Nilsson MH, Hagell P. Measuring outcomes in Parkinson's disease: a multi-perspective concept mapping study. Qual Life Res 2011 Aug 26;21(3):453-463. [doi: 10.1007/s11136-011-9995-3]
20. Sales Force Social Studio. Sales Force Studio. 2019. URL: https://www.salesforce.com/products/marketing-cloud/social-media-marketing/#
21. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. Computational Intelligence 2012 Sep 04;29(3):436-465. [doi: 10.1111/j.1467-8640.2012.00460.x]
22. Kandel H, Khadka J, Shrestha MK, Sharma S, Neupane Kandel S, Dhungana P, et al. Uncorrected and corrected refractive error experiences of Nepalese adults: a qualitative study. Ophthalmic Epidemiology 2017 Oct 06;25(2):147-161. [doi: 10.1080/09286586.2017.1376338]
23. McDonnell PJ. Associations of Presbyopia With Vision-Targeted Health-Related Quality of Life. Arch Ophthalmol 2003 Nov 01;121(11):1577-1581. [doi: 10.1001/archopht.121.11.1577]
24. Williams S, Brian G, Toit RD. Measuring Vision-specific Quality of Life among Adults in Fiji. Ophthalmic Epidemiology 2012 Nov 21;19(6):388-395. [doi: 10.3109/09286586.2012.716896]
25. Cesare N, Grant C, Nsoesie E. Detection of user demographics on social media: A review of methods and recommendations for best practices. arXiv preprint arXiv 2017;1702:01807 [FREE Full text]
26. Schwartz HA, Ungar LH. Data-Driven Content Analysis of Social Media. The Annals of the American Academy of Political and Social Science 2015 Apr 09;659(1):78-94. [doi: 10.1177/0002716215569197]
27. Hunsaker A, Hargittai E. A review of Internet use among older adults. New Media & Society 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
28. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of Social Media Research: Common Concerns and Practical Considerations. Cyberpsychology, Behavior, and Social Networking 2013 Sep;16(9):708-713. [doi: 10.1089/cyber.2012.0334]
29. Golder S, Ahmed S, Norman G, Booth A. Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review. J Med Internet Res 2017 Jun 06;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]
30. Kotsios A, Magnani M, Vega D, Rossi L, Shklovski I. An Analysis of the Consequences of the General Data Protection Regulation on Social Network Research. Trans. Soc. Comput 2019 Dec 17;2(3):1-22. [doi: 10.1145/3365524]
31. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. US Food and Drug Administration. 2009. URL: https://www.fda.gov/media/77832/download [accessed 2020-06-05]

Edited by G Eysenbach; submitted 21.02.20; peer-reviewed by E Papadatou, S Ma; comments to author 23.03.20; revised version received 05.06.20; accepted 07.07.20; published 21.09.20.

Please cite as:
Wolffsohn JS, Leteneux-Pantais C, Chiva-Razavi S, Bentley S, Johnson C, Findley A, Tolley C, Arbuckle R, Kommineni J, Tyagi N
Social Media Listening to Understand the Lived Experience of Presbyopia: Systematic Search and Content Analysis Study
J Med Internet Res 2020;22(9):e18306
URL: http://www.jmir.org/2020/9/e18306/
doi:10.2196/18306
PMID:32955443

Original Paper

Impact of the Internet on Medical Decisions of Chinese Adults: Longitudinal Data Analysis

Qianqian Ma[1,2], MSc; Dongxu Sun[1,2], MSc; Fangfang Cui[1,2], MSc; Yunkai Zhai[1,2,3*], PhD; Jie Zhao[1,2*], PhD; Xianying He[1,2], MSc; Jinming Shi[1,2], MSc; Jinghong Gao[1,2], MD; Mingyuan Li[1,2], PhD; Wenjie Zhang[1,2], PhD

[1]National Engineering Laboratory for Internet Medical Systems and Applications, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

[2]National Telemedicine Center of China, Zhengzhou, Henan, China

[3]School of Management Engineering, Zhengzhou University, Zhengzhou, Henan, China

[*]these authors contributed equally

Corresponding Author:
Yunkai Zhai, PhD
National Engineering Laboratory for Internet Medical Systems and Applications
The First Affiliated Hospital of Zhengzhou University
1 Jianshe Road, Erqi District
Zhengzhou, Henan, 450052
China
Phone: 86 371 67966215
Email: zhaiyunkai@zzu.edu.cn

Abstract

Background: The internet has caused the explosive growth of medical information and has greatly improved the availability of medical knowledge. This makes the internet one of the main ways for residents to obtain medical information and knowledge before seeking medical treatment. However, little has been researched on how the internet affects medical decisions.

Objective: The purpose of this study was to explore the associations between internet behaviors and medical decisions among Chinese adults aged 18 or over, including whether to go to the hospital and which level of medical institution to choose.

Methods: With the adult residents (≥18 years old) in 12 regions including urban and rural areas taken as the research objects, the differences in medical choices of adults with various characteristics were analyzed, and generalized linear mixed models were adopted to analyze the longitudinal data of the China Health Nutrition Survey from 2006 to 2015.

Results: Adult groups with different ages, genders, education levels, regions, places of residence, severities of illness and injury, years of suffering from hypertension, and history of chronic diseases showed diverse medical decisions, and the differences were statistically significant (P<.05). After controlling for these potential confounding factors and taking self-care as the reference, the probability of Chinese adults who participated in online browsing activities selecting hospital care was 0.82 (95% CI 0.69-0.98; P=.03) times that of residents who did not participate in online browsing activities. In terms of medical institution choices, adults who participated in online browsing activities were 1.86 (95% CI 1.35-2.58; P<.001) times more likely to opt for municipal medical treatment than primary care. However, the effect of online browsing on the selection probability of county-level hospitals was not significant compared with primary hospitals (P=.59). Robust analysis verified that accessing the internet had a similar effect on Chinese adults' medical decisions.

Conclusions: Chinese adults who use the internet are a little less likely to go to the hospital than self-care. The internet has broken down the barriers to obtain knowledge of common diseases and thus has a slight substitution effect of self-care on hospital care. Internet use may increase the probability of adults going to municipal hospitals. The rising tendency of visiting high-level medical institutions may be consequently exacerbated due to knowledge monopoly of severe and complicated diseases that is difficult to eliminate, and the increase in inconsistent and incomplete medical information online will blur the residents' cognitive boundary of common diseases and severe diseases. Exploring the substantive impact of the internet on medical decision making is of great significance for further rational planning and utilization of the internet, in order to guide patients to appropriate medical institution.

(J Med Internet Res 2020;22(9):e18481) doi:10.2196/18481

KEYWORDS

internet; medical decision; health care provider choice; adult; longitudinal data analysis; hierarchical medical policy

Introduction

Medical resources have been unequally distributed in different levels of hospitals in China for a long time, and this has resulted in the chaos of residents' health-seeking behaviors [1,2]. According to statistics, the number of hospital beds in medical institutions in urban areas was 8.70 per 1000 people, compared with 4.56 per 1000 people in rural areas and 1.43 per 1000 people in township hospitals in 2018 [3]. Furthermore, the number of licensed doctors (assistants) per 1000 people in urban and rural areas was 10.91 and 4.63, respectively [3], which indicates the imbalance of medical resources in China. In other words, health resources have been allocated in big cities and superior hospitals. As a result, patients who suffer from acute or severe diseases, or those with stable or mild diseases, tend to select larger and professional hospitals rather than the smaller ones. The overcrowded high-level medical institutions and other medical institutions with insufficient patients reflect the inefficient utilization and wasting of health resources, and the chaos existing in medical practice [4]. In recent years, the hierarchical medical policy (HMP) has been proposed to steer patients from higher- to lower-cost providers [5]. The main idea behind the implementation of HMP is that the initial diagnosis is recommended to be carried out at the grassroots level, chronic and common diseases treated in the primary hospitals, and acute and intractable diseases referred to higher-level hospitals for treatment [6]. In 2018, the Chinese government proposed *Internet & health care*, a crucial development strategy which regards the internet as an important means to promote the implementation of HMP and optimize allocation of health care resources [7].

The period from 2006 to 2011 represents the initial development stage of China's internet health care. Telemedicine gradually emerged in 2015, while online medicine and remote consultation were not widely used. However, patients or healthy people can obtain disease knowledge online through search engines such as *Baidu*, and professional medical websites such as *Good Doctor Online* and *Weiyi*. The statistical report on internet development in China has suggested that internet penetration has continued to grow and the popularity of the internet has gradually spread to the elderly from the young [8]. Given the convenience, high usability, and wider accessibility, internet has become a medium for the dissemination of health information, which has led to the inflated growth of online medical information and achieved the widespread sharing of medical knowledge. The internet provides us with a variety of health information, including drug information, basic definition and symptoms, treatment methods, and mental health information [9]. People browse health information through the internet to make further medical decisions. For example, patients who originally plan to go to a large hospital change their minds and choose home-based self-care or a nearby primary medical institution for treatment instead after browsing online, or they realize the seriousness of the disease from the internet and immediately seek treatment at a high-level medical institution.

Studies have shown that an increasing number of people tend to use the internet to obtain health care information [10-12], including older adults [13,14]. One study revealed that 57% of adults with acute coronary syndromes in 6 hospitals in Massachusetts and Georgia sought health information online [11]. Another survey found that 88% of participants with opioid treatment searched for information on medical topics online [10]. Furthermore, another study showed that 89% of 335 Chinese pregnant women who attended the antenatal clinic in a general hospital in Guangzhou used the internet to retrieve health information from the beginning of the pregnancy [15]. For Chinese patients with invasive breast cancer, the rate of internet information searching was reported to be 49% [16]. Previous studies have demonstrated that using the internet to obtain health information is very common in the internet era. Exploring the substantive impact of the internet on medical decision making is of great significance for further rational planning and use of the internet, to guide patients to the appropriate medical institution based on their illness.

However, there is limited evidence about how internet affects adult hospital choices. When it comes to the factors influencing residents' choice of medical treatment, most previous studies have focused more on hospital-related factors, including hospital equipment, distance, time, cost, reputation, doctor level, etc [17-22]. A survey showed that the primary reasons for choosing private hospitals were the presence of a specialist, availability of good equipment and technology, and trust in treatment, whereas proximity, receiving enough information, and being well-treated were the reasons why participants chose family health centers in Samsun Province in Turkey [20]. A semistructured interview study involving 13 pregnant women in Denmark noted that the experience of pregnant women themselves or their peers and travel distance played a role in the women's choice of delivery hospital [21], while Schuldt et al [22] believed that factors such as the distance to hospital, level of information about the treatment, number of respective treatments performed in the hospital per year, and complication rate had a significant impact on hospital choice. However, these studies have rarely involved internet use.

Several studies that involve internet use indicate that the relationship between the internet and medical decision making still needs to be clarified. Lee et al [23] pointed out that online health information had the potential to powerfully influence the health attitude and behaviors of a large proportion of the population, and affected the management of chronic diseases. However, an earlier study [24] suggested that the internet could enhance residents' health-related knowledge and attitudes to a certain extent, but rarely changed their health-related behaviors. Similarly, in the study by Zwijnenberg et al [25], patients showed interest in online comparative health care information, but the impact of internet on patients' decision making remained limited. Consequently, it is still unclear whether searching online information through the internet will affect patient's decision to go to the hospital and the choice of health care providers.

Therefore, the purpose of this study was to explore residents' decision-making behavior under the background of the internet era, and to analyze whether the internet could guide and channel patients to the suitable medical institution, so as to achieve hierarchical treatment. Based on measurements of longitudinal data from 2006 to 2015, the generalized linear mixed model was employed to explore the associations between internet use and medical decisions in general Chinese adult population, combined with other relevant factors influencing patients' preference for hospital types.

Methods

Data Source

Data were extracted from the China Health and Nutrition Survey [26], an international cooperation project jointly conducted by the Carolina Population Center of University of North Carolina at Chapel Hill and National Institute of Nutrition and Food Safety of the Chinese Center for Disease Control and Prevention. The survey is a continuously open cohort with a multistage, stratified cluster random sampling method, covering 12 regions including Heilongjiang, Liaoning, Hunan, Shandong, Guizhou, Jiangsu, Guangxi, Hubei, Henan, Beijing, Shanghai, and Chongqing. These regions differ in geographical location, economic development, public resources, health conditions, and other demographic measures, making the survey informative, high-quality, nationally representative data. The entire data collection and collation process has been subject to good quality control. In addition, the desensitized and anonymous data have been publicly released online, without patient privacy.

Questions about internet behaviors have been set in the original Chinese questionnaire after 2006, such as the internet location, online browsing, online chat, online game playing, and the duration of internet behaviors. Thus, this paper selected 2006-2015 longitudinal data. Because the medical treatment of minors is often decided by the guardian rather than by the minors themselves, residents younger than 18 years were excluded, and the research object included only the adult group. After data cleaning, the records with missing key variables, such as health care–seeking behavior and internet use, were excluded. The final analysis included 10,164 records, of which there were 2032, 2280, 3145, and 2707 records in 2006, 2009, 2011, and 2015, respectively. Among them, 4877 records were obtained from the same individuals by repeated observations. A total of 7408 adult participants were involved in the analysis and 2121 participants had records that were repeated at least twice. Each participant was followed up for 1, 2, 3, or 4 times (ie, not everyone had 4 records), which suggests unbalanced longitudinal data.

Internet Use and Health Care–Seeking Behaviors

The internet behavior was obtained through the questions "Do you participate in surfing the internet? (Yes/No)" and "Can you access the internet? (Yes/No)" on the questionnaire. Browsing online was chosen as a proxy variable for internet usage, because only individuals who participated in online browsing activities had the opportunity to access the internet medical information. At the same time, in order to test the robustness of the association between internet use and health care–seeking behaviors, internet access was used as another explanatory variable in the robustness analysis.

Health-seeking behaviors were obtained from the questions "What did you do when you felt ill?" and "Which medical institution did you seek first?" The patients made medical decisions in the following 2 steps: (1) Whether to go to the hospital and (2) Which hospital to go to. Therefore, the analysis of the impact of the internet on medical behavior was divided into 2 parts according to the decision-making process. First, did it affect the patient's choice of whether to seek medical treatment, self-care, or hospital care? This was a 2-category event. Second, based on the level of medical institutions, we classified health care provider choices into primary-level hospital, county-level hospital, and municipal-level hospital, which was a 3-category event.

Potential Confounders

Some other factors might influence residents' decision to seek medical treatment. For example, people of different ages and genders show distinct preferences for hospitals. Geographical differences indicate diverse levels of modernization as well as economic and medical development, and thus residents' choice of the hospital may be affected by the supply of local medical resources. Given the potential confounders, it is not enough to consider only the single-factor influence of internet behavior on medical decision making. Thus, the adjusted model included sociodemographic characteristics (marriage, age, gender, education), health supply (medical insurance, district, urban or rural), health needs (body mass index, severity of illness or injury, history of chronic illness, hypertension), and other factors as covariates to in-depth verify the influence of internet behavior. Details of the variables are presented in Table 1.

Table 1. Description of variables.

Variables and description	Variable assignment
Explained variable	
Medical choice	0=Self-care, 1=Hospital care
Tier of hospital care	1=Primary hospital, 2=County hospital, 3=Municipal hospital
Explanatory variables	
Online browsing	0=No, 1=Yes
Internet access	0=No, 1=Yes
Confounders	
Marital status	0=Married, 1=Other (single, widowed, divorced, or separated)
Age	0=18-44 years old, 1=45-59 years old, 3=60-74 years old, 4=≥75 years old
Gender	0=Female, 1=Male
Education level[a] (years)	Years of being educated
Medical insurance	0=No, 1=Yes
District	0=Center, 1=East, 2=West
Residence site	0=Rural, 1=Urban
Time	Survey year (1=2006, 2=2009, 3=2011, 4=2015)
Disease/injury severity	1=Not severe, 2=Somewhat severe, 3=Quite severe
Chronic diseases[a]	The number of chronic diseases diagnosed by doctors, including hypertension, diabetes, myocardial infarction, stroke, asthma, tumor
BMI[a] (kg/m^2)	Body mass index, calculated by weight (kg)/height (m^2)
Hypertension (years)[a]	Years of suffering from hypertension

[a]Continuous variable.

Statistical Analysis and Methodology

Data collation and cleaning were performed using RStudio 1.1.456 software (RStudio, Inc.). The random forest algorithm was applied to fill in the missing values of potential confounders (<10%) after removing duplicate records and missing samples of key variables. In descriptive statistical analysis, statistical charts and tables were adopted to analyze the changes in health care provider choices among Chinese adults, and the differences in health-seeking behaviors among adults with different characteristics. The quantitative data were described by mean and standard deviation, whereas qualitative data were analyzed using rate or composition ratio. Univariate analysis was performed by Wilcoxon rank-sum test, multisample Kruskal–Wallis rank-sum test, chi-square test, and Cochran–Mantel–Haenszel test. In multivariate analysis, because the data were longitudinal and the health care–seeking behaviors were characterized with 2 categories in the first step and 3 categories in the second step as the dependent variable, the mixed-effects binary or multinomial logit model, (ie, a generalized linear mixed model with binomial or multinomial distribution and logit link function) was perhaps the most appropriate statistical perspective for analyzing such data when accounting for the potential lack of independence in longitudinal data [27,28].

Methodologically, combining the strengths of both the generalized linear model and linear mixed model, the generalized linear mixed model extends the generalized linear model further to account for variation and correlation of longitudinal data. A random effect b_{ik} (i=1, 2, ..., m) was introduced and the logit link function was selected in the model. With *k*=0 serving as the reference, the model was expressed using the following equation [28]:

$$\log(P_{ijk}/P_{ij0})=X_{ij}'\beta+b_{ik}+\varepsilon_{ijk}\ k=1, ..., K \quad \textbf{(1)}$$

where P_{ijk} denotes the probability that adult i makes a medical decision of k in survey year j, $P_{ijk}=Pr(Y_{ij}=k|X_{ij})$; ε_{ijk} is the within-subject random error and was normally distributed as $N(0,\sigma_{ijk}^2)$; b_{ik} is the between-subjects random effect on the *k*th logit component, and was assumed to be distributed as $N(0,\sigma^2_{bk})$; and X_{ij} is the covariate vector. The mixed-effects binomial logit model (K=1) and mixed-effects multinomial logit model (K=2) were established by GLIMMIX Proc Step in SAS software, version 9.4 (SAS Institute Inc.) [29]. All tests were two-sided at the significance level α=.05 and *P*<.05 indicated statistical significance.

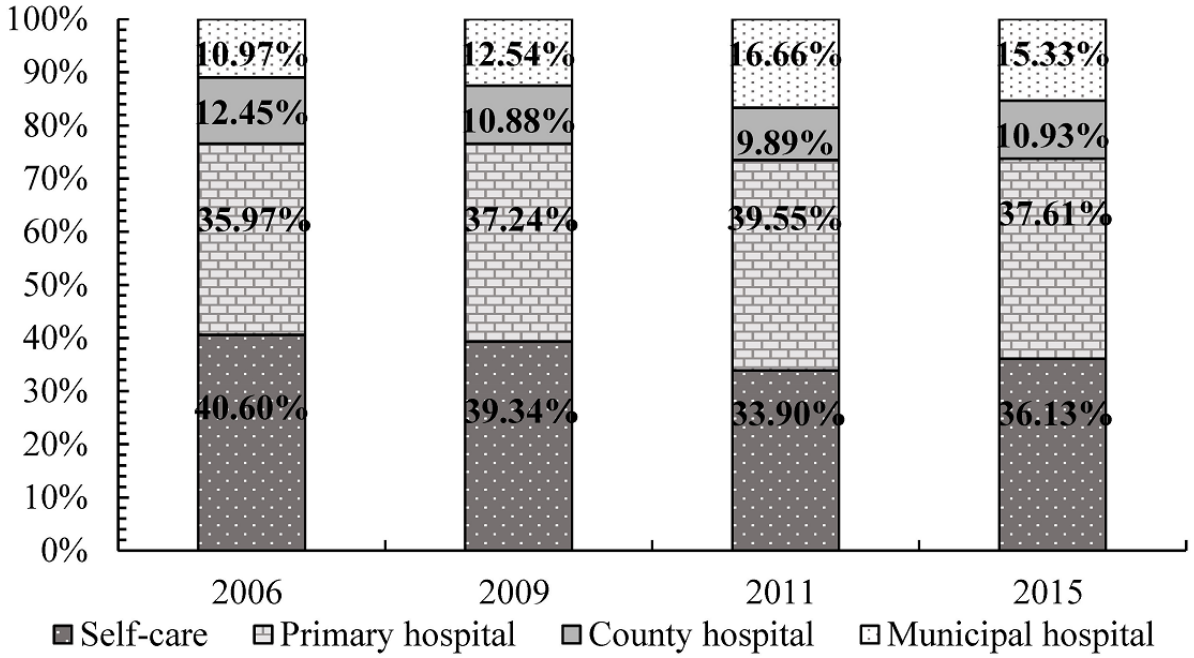
Results

Health Care Provider Choices for Chinese Adult Residents

On the whole, primary care and self-care were the main medical treatment choices for Chinese adults after they were sick or

injured, accounting for 37.80% (3842/10,164) and 37.05% (3766/10,164) of the total records, respectively, followed by municipal and county hospitals. From 2006 to 2015, the proportion of consultations at primary medical institutions increased by 1.64% (from 35.97% to 37.61%), which indicated the moderate effects of HMP. The proportion of residents choosing municipal hospitals grew by 4.36% (from 10.97% to 15.33%), whereas the figures for choosing self-care and county hospitals both decreased (Figure 1).

Figure 1. Health care provider choices of Chinese adult residents from 2006 to 2015.



Factors Influencing Adults' Medical Decisions

Taking the data in 2015 as an example, people with different characteristics had various preferences for health care (Tables 2 and 3).

First, for self-care versus hospital care choice, there was a significant correlation between adults' age and their medical choices (ie, as they grew older, more patients chose to go to the hospital for treatment instead of self-treatment; χ^2_3=63.0, *P*<.001). The factor of disease or injury severity was also found to be statistically significantly associated with medical choices. Patients with more severe illness or injury were more likely to choose hospital care (χ^2_2=94.3, *P*<.001). In addition, education levels (*P*=.005), residence sites (*P*<.001), years of suffering from hypertension (*P*<.001), and history of chronic diseases (*P*<.001) differed significantly between those who chose self-care and those who chose hospital care (α=.05). However, gender, marital status, medical insurance, and BMI were not significantly associated with the choice of self-care or hospital care in univariate analysis (Table 2).

Second, for the tier of hospital care in Table 3, adult groups with different genders, education levels, regions, places of residence, severities of illness and injury, years of suffering from hypertension, and history of chronic diseases showed diverse choices of medical institutions, and the differences were statistically significant (*P*<.05). For instance, those with higher education mainly selected municipal hospitals (χ^2_2=76.1, *P*<.001). The proportions of urban residents' choices of hospital were ranked as primary hospitals, municipal hospitals, county-level hospitals, whereas rural residents' choices of hospitals were ranked as primary hospitals, county hospitals, municipal hospitals (χ^2_2=159.7, *P*<.001).

Table 2. Medical choices for people with different characteristics in 2015 (N=2707).

Variables	Medical choice		Hypothetical test	
	Self-care (N=978)	Hospital care (N=1729)	χ^2 (*df*)/W	*P* value
Age (years), n (%)			63.0 (3)	<.001
18-44	232 (23.72)	226 (13.07)		
45-59	320 (32.72)	528 (30.54)		
60-74	323 (33.03)	725 (41.93)		
≥75	103 (10.53)	250 (14.46)		
Gender, n (%)			2.4 (1)	.12
Male	449 (45.91)	739 (42.74)		
Female	529 (54.09)	990 (57.26)		
Marriage status, n (%)			2.0 (1)	.16
Married	836 (85.48)	1441 (83.34)		
Others	142 (14.52)	288 (16.66)		
Education level (years), mean (SD)	8.140 (4.48)	7.57 (4.53)	899,572[a]	.005
Region, n (%)			4.3 (2)	.11
East	388 (39.67)	737 (42.63)		
Center	323 (33.03)	506 (29.27)		
West	267 (27.30)	486 (28.11)		
Residence site, n (%)			13.8 (1)	<.001
Urban	518 (52.97)	786 (45.46)		
Rural	460 (47.03)	943 (54.54)		
Medical insurance, n (%)			0.01 (1)	.92
No	36 (3.68)	61 (3.53)		
Yes	942 (96.32)	1668 (96.47)		
Disease or injury severity, n (%)			94.3 (2)	<.001
Not severe	487 (49.80)	548 (31.69)		
Somewhat severe	444 (45.40)	1005 (58.13)		
Quite severe	47 (4.81)	176 (10.18)		
Hypertension (years), mean (SD)	2.01 (5.38)	4.13 (8.16)	717,065[a]	<.001
BMI (kg/m^2), mean (SD)	24.15 (3.84)	24.31 (3.78)	821,957[a]	.23
Chronic diseases, mean (SD)	0.34 (0.62)	0.62 (0.80)	685,490[a]	<.001

[a]Wilcoxon rank-sum test.

Table 3. Hospital choices for people with different characteristics in 2015 (N=1729).

Variables	Tier of hospital care			Hypothetical test	
	Primary hospital (N=1018)	County hospital (N=296)	Municipal hospital (N=415)	χ^2 (*df*)	*P* value
Age (years), n (%)				10.2 (6)	.12
18-44	137 (13.46)	37 (12.50)	52 (12.53)		
45-59	293 (28.78)	106 (35.81)	129 (31.08)		
60-74	451 (44.30)	111 (37.50)	163 (39.28)		
≥75	137 (13.46)	42 (14.19)	71 (17.11)		
Gender, n (%)				6.4 (2)	.04
Male	415 (40.77)	145 (48.99)	179 (43.13)		
Female	603 (59.23)	151 (51.01)	236 (56.87)		
Marriage status, n (%)				4.2 (2)	.12
Married	836 (82.12)	258 (87.16)	347 (83.61)		
Others	182 (17.88)	38 (12.84)	68 (16.39)		
Education level (years), mean (SD)	6.87 (4.47)	7.84 (4.45)	9.12 (4.35)	76.1 (2)[a]	<.001
Region, n (%)				11.9 (4)	.02
East	401 (39.39)	136 (45.95)	200 (48.19)		
Center	307 (30.16)	85 (28.72)	114 (27.47)		
West	310 (30.45)	75 (25.34)	101 (24.34)		
Residence site, n (%)				159.7 (2)	<.001
Urban	413 (40.57)	79 (26.69)	294 (70.84)		
Rural	605 (59.43)	217 (73.31)	121 (29.16)		
Medical insurance, n (%)				3.7 (2)	.16
No	41 (4.03)	5 (1.69)	15 (3.61)		
Yes	977 (95.97)	291 (98.31)	400 (96.39)		
Disease or injury severity, n (%)				30.7 (4)	<.001
Not severe	371 (36.44)	78 (26.35)	99 (23.86)		
Somewhat severe	563 (55.30)	179 (60.47)	263 (63.37)		
Quite severe	84 (8.25)	39 (13.18)	53 (12.77)		
Hypertension (years), mean (SD)	3.67 (7.74)	3.71 (7.05)	5.58 (9.63)	13.9 (2)[a]	<.001
BMI (kg/m^2), mean (SD)	24.26 (3.72)	24.77 (4.20)	24.12 (3.61)	4.0 (2)[a]	.13
Chronic diseases, mean (SD)	0.54 (0.73)	0.65 (0.85)	0.79 (0.89)	26.3 (2)[a]	<.001

[a]Multisample Kruskal–Wallis rank-sum test.

Relationship Between Internet Use and Medical Decisions

Adults who did not browse the internet presented an obvious preference for primary hospitals, supplemented by self-diagnosis and treatment. By contrast, people who browsed the internet had different medical treatment–seeking behaviors, and they preferred self-care, followed by medical care from primary hospitals and municipal hospitals (Table 4). With the time (year) as a stratified variable, it was found that the use of the internet was significantly related to the choice of health care provider among adults after controlling the time variable by the Cochran–Mantel–Haenszel test (χ^2_3=170.4, *P*<.001).

Table 4. Association between internet use and medical decisions in different survey years.

Year: Internet use	Self-care, n (%)	Primary hospital, n (%)	County hospital, n (%)	Municipal hospital, n (%)
2006: Online browsing				
Yes (N=124)	65 (52.42)	20 (16.13)	16 (12.90)	23 (18.55)
No (N=1908)	760 (39.83)	711 (37.26)	237 (12.42)	200 (10.48)
2009: Online browsing				
Yes (N=225)	129 (57.33)	41 (18.22)	14 (6.22)	41 (18.22)
No (N=2055)	768 (37.37)	808 (39.32)	234 (11.39)	245 (11.92)
2011: Online browsing				
Yes (N=502)	207 (41.24)	128 (25.50)	42 (8.37)	125 (24.90)
No (N=2643)	859 (32.50)	1116 (42.22)	269 (10.18)	399 (15.10)
2015: Online browsing				
Yes (N=441)	205 (46.49)	117 (26.53)	39 (8.84)	80 (18.14)
No (N=2266)	773 (34.11)	901 (39.76)	257 (11.34)	335 (14.78)

Impact of the Internet on Choosing Self-Care Versus Hospital Care

Taking self-care as the reference group, the mixed-effects binomial logit model was employed to analyze whether online browsing would influence patient's decision to visit hospital. Based on the univariate analysis of Model 1, Models 2 and 3 further introduced different confounders that potentially affect patients' medical decision to validate whether the relationship between online browsing and patient decisions was still significant. Models 1-3 all clarified that Chinese adults who participated in online browsing activities were less likely to go to the hospital than those who did not participate in online browsing activities. As revealed in Model 3, the odds ratio was 0.82 ($e^{-0.20}$; 95% CI 0.69-0.98; P=.03) in the group that participated in online browsing activities compared with those that did not participate in online browsing activities (Table 5).

Table 5. Results of a generalized linear mixed-effects binomial logit model analyzing the influence of internet use on choosing self-care versus hospital care.

Effects	Model 1 (unadjusted model)		Model 2		Model 3	
	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Fixed effects						
Intercept	0.70 (0.64 to 0.76)	<.001	0.72 (0.53 to 0.92)	<.001	0.55 (0.15 to 0.95)	.007
Online browsing (ref=No)						
Yes	–0.52 (–0.66 to 0.38)	<.001	–0.24 (–0.41 to –0.08)	.004	–0.20 (–0.37 to –0.02)	.03
Age (ref=18-44)						
45-59			0.16 (0.02 to 0.30)	.03	–0.01 (–0.16 to 0.14)	.91
60-74			0.21 (0.06 to 0.37)	.008	–0.09 (–0.26 to 0.07)	.27
≥75			0.34 (0.13 to 0.55)	.001	–0.05 (–0.28 to 0.17)	.64
Gender (ref=Female)						
Male			–0.05 (–0.16 to 0.06)	.40	–0.07 (–0.19 to 0.04)	.23
Time (ref=2006)						
2009			0.07 (–0.07 to 0.20)	.32	0.07 (–0.08 to 0.22)	.36
2011			0.38 (0.25 to 0.51)	<.001	0.33 (0.18 to 0.48)	<.001
2015			0.29 (0.16 to 0.43)	<.001	0.25 (0.10 to 0.40)	.001
Region (ref=Center)						
East			0.04 (–0.08 to 0.17)	.49	–0.02 (–0.15 to 0.11)	.76
West			0.04 (–0.10 to 0.18)	.55	0.02 (–0.12 to 0.17)	.76
Residence site (ref=Rural)						
Urban			–0.60 (–0.71 to –0.49)	<.001	–0.71 (–0.82 to –0.60)	<.001
Marriage status (ref=Married)						
Others			–0.15 (–0.29 to –0.01)	.04	–0.15 (–0.30 to 0.003)	.046
Education level			–0.02 (–0.03 to 0.00)	.01	–0.01 (–0.03 to 0.001)	.08
Disease/injury severity (ref= Not severe)						
Somewhat severe					0.85 (0.74 to 0.95)	<.001
Quite severe					1.49 (1.29 to 1.68)	<.001
Chronic diseases					0.32 (0.22 to 0.42)	<.001
Hypertension					0.01 (–0.001 to 0.02)	.09
Medical insurance (ref= No)						
Yes					0.03 (–0.13 to 0.19)	.73
BMI					–0.02 (–0.03 to 0.004)	.15
Random effect						
Intercept, variance	2.40	<.001	2.47	<.001	2.82	<.001

Impact of the Internet on the Choices of Tier of Hospital Care

Taking the primary medical institution as the reference group, 3 mixed-effects multinomial logit models (Models 4-6) were established by using different factors that might affect hospital choices as control variables. All parameter estimates of the models were shown in Multimedia Appendix 1, and the key results we were most interested in are presented in Table 6. The result showed that Chinese adults who participated in online

browsing activities were more likely to choose municipal hospitals than primary medical institutions, whether in the unadjusted analysis (Model 4) or in the models adjusted for confounding factors (Models 5 and 6). The multifactor Model 6 hinted that after controlling for as many confounding factors as possible, residents participating in online browsing activities were 1.86 ($e^{0.62}$; 95% CI 1.35-2.58; P<.001) times more likely to opt for municipal medical treatment than those who did not participate in online browsing activities (Figure 2). However, the effect of online browsing on the selection probability of county-level hospitals was not significant compared with primary hospitals (P=.59).

Table 6. Results of generalized linear mixed-effects multinomial logit model analyzing the influence of online browsing on medical provider choice (ref=primary hospital).

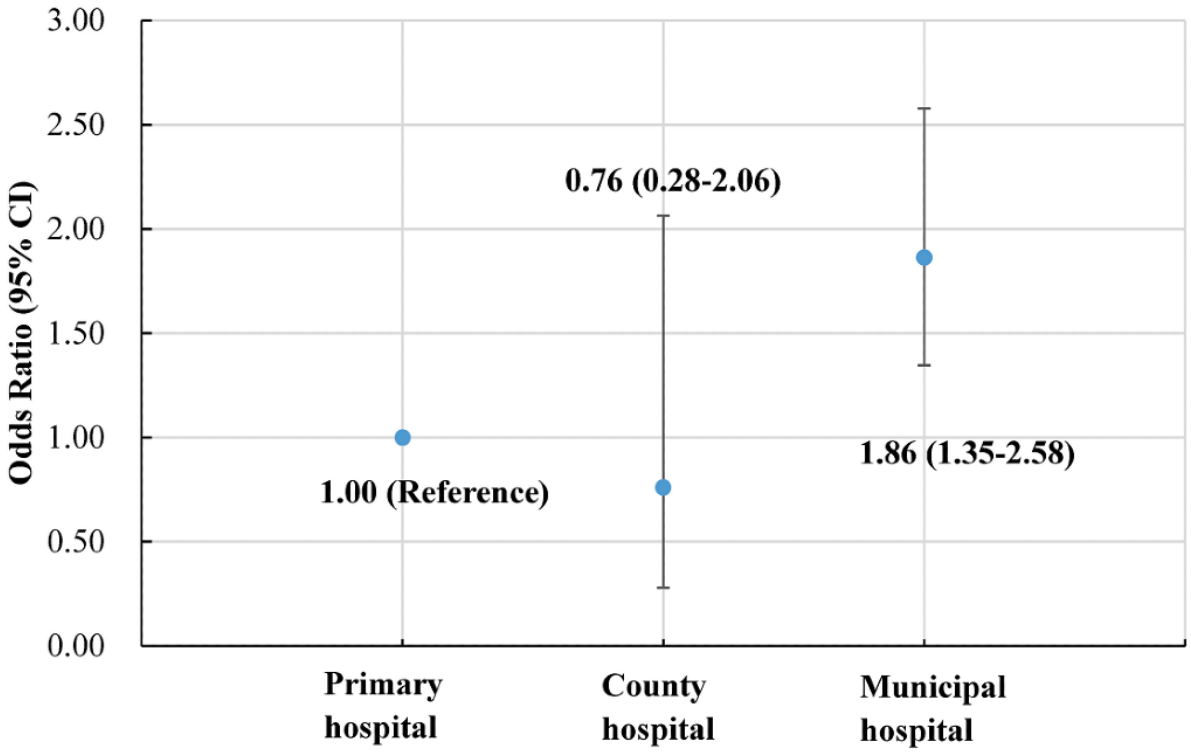
Model and dependent variable	Online browsing		
	Coefficient	(95% CI)	*P* value
4[a]			
County hospital	–0.05	(–0.97 to 0.86)	.90
Municipal hospital	1.15	(0.51 to 1.78)	<.001
5[b]			
County hospital	–0.31	(–1.28 to 0.66)	.53
Municipal hospital	0.51	(0.20 to 0.81)	.001
6[c]			
County hospital	–0.27	(–1.27 to 0.73)	.59
Municipal hospital	0.62	(0.30 to 0.95)	<.001

[a]Only explanatory variable was included in the model.

[b]The confounders included in the model were the same as those in Model 2.

[c]The confounders included in the model were the same as those in Model 3.

Figure 2. Odds ratio estimates based on Model 6.



Robust Analysis

Tables 7 and 8 present the results of robust analysis using "internet access" as another explanatory variable instead of the existing online browsing. As revealed in Model 9, the odds ratio was 0.85 ($e^{-0.16}$; 95% CI 0.74-0.99, P=.03) in the group that could access the internet compared with that which could not

access the internet. Model 12 showed that compared with primary hospital, the probability of residents who could access the internet selecting municipal hospital was 1.57 ($e^{0.45}$; 95% CI 1.20-2.07, *P*=.001) times that of residents who did not access the internet. Besides, there was no preference gap for primary and county hospitals (*P*=.98). Robust analysis verified similar results that the internet had a certain effect on adults' medical choices.

Table 7. Results of analyzing the influence of "accessing the internet" on medical choice behaviors (self-care versus hospital care, ref=self-care).

Model	Internet access		
	Coefficient	(95% CI)	*P* value
7[a]	–0.45	(–0.57 to –0.34)	<.001
8[b]	–0.18	(–0.32 to –0.04)	.01
9[c]	–0.16	(–0.31 to –0.01)	.03

[a]Only explanatory variable was included in the model.

[b]The confounders included in the model were the same as those in Model 2.

[c]The confounders included in the model were the same as those in Model 3.

Table 8. Results of analyzing the influence of "accessing the internet" on the choice of hospital (ref=primary hospital).

Model and dependent variable	Internet access		
	Coefficient	(95% CI)	*P* value
10[a]			
County hospital	0.16	–0.52 to 0.84	.64
Municipal hospital	0.81	0.27 to 1.35	.003
11[b]			
County hospital	–0.03	–0.78 to 0.71	.93
Municipal hospital	0.35	0.09 to 0.60	.008
12[c]			
County hospital	0.01	–0.76 to 0.78	.98
Municipal hospital	0.45	0.18 to 0.73	.001

[a]Only explanatory variable was included in the model.

[b]The confounders included in the model were the same as those in Model 2.

[c]The confounders included in the model were the same as those in Model 3.

Discussion

Principal Findings

Based on longitudinal data from 2006 to 2015, this paper analyzed the impact of internet on the medical decisions among Chinese adults through generalized linear mixed models. The results showed that the internet had a certain effect on adults' medical decisions. First, regarding the impact of whether to go to the hospital, adults with internet behaviors (eg, browsing information online, accessing the internet) were less likely to go to the hospital. Patients tended to self-care, which presented a partial substitutive effect of self-diagnosis and treatment on hospital care. Second, in terms of hospital selection, compared with primary hospitals, the use of the internet might not change the probability of choosing county hospitals, but it might increase the probability of going to municipal hospitals for advanced treatment. The study has theoretical and practical implications on how to regulate internet health care and guide patients to seek medical institutions, and has a reference to the promotion and application of internet medical treatment.

Chinese adults with internet behaviors are more likely to self-diagnose and treat at home than visiting hospitals, which is consistent with some research descriptions [30,31]. Yang et al [30] pointed out that in the "internet +" era, online medical platforms provided an effective way to alleviate the high demand for hospitals. As the popularity of the internet has increased dramatically among people, browsing and selecting health information have become a basic approach before determining whether to visit hospitals further [31]. A study of 164 perinatal women in Korea showed that some women, who sought informal medical help online, would be more likely to change their medical decisions only according to internet information, without consulting doctors (*P*<.001) [32]. Concerning the reasons for choosing self-care instead of primary care, some studies have given explanations [10,33,34]. One study noted that almost half of health information searchers (48%) reported that health information online could help them take better care

of themselves, and two-thirds of adults (67%) showed increasing awareness of health issues through internet [33]. Turan et al [34] and others suggested that online access to reliable disease information could abate anxiety, boost the feelings of self-efficacy, and reduce the use of medical services. The popularity of the internet can effectively overcome traditional obstacles and achieve easy access to health information for prevention and treatment [10]. All in all, the internet can break down the barriers to the knowledge of common diseases, reduce the asymmetry of information between patients and doctors to some extent, and improve patients' awareness and access to basic health knowledge, thereby reducing the possibility of using medical services.

By contrast, this study found that the internet might exacerbate the tendency of going to higher-level medical institutions for medical treatment. The information browsed on the internet is not able to resolve the monopoly of knowledge about intractable and severe diseases. In addition, residents' misunderstanding of medical expertise can cause health anxiety, for instance, misinterpretation of physical symptoms as signs of serious diseases, accompanied by persistent fear of serious illness [35]. Some studies have reported that internet health information searchers were more likely to have health concerns than nonseekers, and adult seekers tended to rate their health status as poor [36,37]. Furthermore, a random effect meta-analysis demonstrated that online health information seeking was positively correlated with health anxiety (r=0.34, 95% CI 0.20-0.48, P<.001) [37]. At the same time, given the privacy principles, the medical information that can be retrieved is often partial, subjective, and even biased, which aggravates the limitations and incompleteness of residents' awareness of the disease. In a semistructured interview on the use of Chinese language internet information on cancer, most of the 20 respondents reported that they encountered internet health information with questionable quality [38]. An observational study showed that some sites provide harmful information, and the proportion of these sites was much higher than sites providing reliable information on cancer treatment (N=247) [39]. The studies above hint at the reasons why the use of internet might increase the probability of residents going to high-level medical institutions.

Unlike previous studies that have paid more attention to the impact of hospital-related factors on patients' medical decision making, our study focused on internet use. Especially in the internet era, as mentioned previously, the internet has played a vital role in residents' decision making on their choice of hospitals [40]. Li et al [41] demonstrated that there was a strong association between online health communities information and patient decisions of switching from online to offline medical services. One study suggested an association between online health information–seeking behaviors and some health behaviors, such as physical activity, fruit and vegetable consumption, alcohol use, and hypertension medication adherence [23]. However, there are few studies that deal with such health behaviors as whether to go to a hospital and what level of hospital to choose, under the influence of the internet. This research has innovatively analyzed the influence of internet behavior on medical choices by following the 2 steps of the decision-making process. In addition, some factors such as age or education level might be associated with medical decisions. According to a survey in Samsun Province in Turkey, patients aged 18 years or younger and 65 years or older preferred family health centers, whereas those aged 19-64 presented a higher preference for private hospitals [20]. In addition, it was pointed out that the level of education affected patients' choices [20]. Our study not only explored the impact of the internet use on the residents' choice of health care provider by univariate analysis, but also deeply took other confounding factors into account, including age, gender, region, urban or rural, education, disease severity, chronic medical history, and BMI, that might affect health care choices from the perspective of residents.

Limitations

This study has employed the generalized linear mixed models to delve into the associations between internet use and medical decisions with longitudinal data, which fills in the gaps of current related research and provides a reference for policy makers. To our knowledge, this is the first time that the mixed-effects multinomial logit regression, an appropriate method for processing longitudinally correlated multiclass data [42,43], is adopted for modeling medical institution choices in China. However, there are some limitations in this study. First, variables such as occupation, income, transportation mode, self-perceived life happiness index, and internet browsing time were not included in the model as confounding factors due to high percentage of missing data. In addition, when interpreting the results, only the internet behaviors in the main forms of "online browsing" and "having access to the internet" were considered, rather than interactive internet medical behaviors, such as online consultation with doctors. As a result, further study focusing more on medical information can be conducted with an in-depth assessment of network usage, including network usage time, languages of online health information (in English or in Chinese) [36], content of information (Western medicine or traditional Chinese medicine), level of trust in online information, etc, which can deeply portray the impact of the internet on residents' health care–seeking behaviors.

Conclusions

With the advent of the internet, the availability of health care information has improved. The internet has become a pivotal source of medical information for Chinese residents [13,44]. This study has found that compared with self-care, internet use slightly reduces the probability of patients going to the hospital to some extent. In addition, compared with primary hospitals, the internet seems not to change the probability of choosing county hospitals, although it may increase the probability of adults going to municipal hospitals for high-level health care. The internet has broken down the barriers to the knowledge of common diseases, shortened the gaps in health information accessibility, and has produced a slight substitution effect of self-diagnosis and treatment on hospital care. However, the knowledge monopoly of difficult and complicated diseases cannot be eliminated, and at the same time, the increase in inconsistent, incomplete, and commercialized medical information has also brought noise to decision making, and will blur the residents' cognitive boundary of common diseases and

severe diseases. Consequently, the rising tendency of visiting high-level medical institutions may be exacerbated, which is unable to guide patients to hierarchical diagnosis and treatment. It is necessary to further regulate the normativeness of medical-related websites, ensure the correctness and scientificity of medical knowledge online, and reduce the noise of medical information correspondingly in order to achieve the standardized dissemination of medical knowledge. For example, it is recommended to promote the implementation of telemedicine and internet hospitals, and make it an important means to support health self-management and rehabilitation with extensive application of internet technology, and guide patients to make medical decisions, which will ultimately contribute to the formation of hierarchical diagnosis and treatment order.

Acknowledgments

This study was supported by grants from the National Key Research and Development Program of China (Grant No. 2017YFC0909900), the National Natural Science Foundation of China (Grant No. 71673254), and Innovation Research Team of Higher Education in Henan Province (Grant No. 20IRTSTHN028). This research uses data from China Health and Nutrition Survey (CHNS). We thank the National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700), and the Fogarty International Center, NIH for financial support for the CHNS data collection and analysis files from 1989 to 2006 and both parties plus the China–Japan Friendship Hospital, Ministry of Health for their support for CHNS 2009 and future surveys.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Estimation results of generalized linear mixed-effects multinomial logit models analyzing the influence of online browsing on medical provider choice.
[PDF File (Adobe PDF File), 385 KB - jmir_v22i9e18481_app1.pdf]

References

1. Hu Y, Zhang Z. Skilled doctors in tertiary hospitals are already overworked in China. Lancet Glob Health 2015 Dec;3(12):e737 [FREE Full text] [doi: 10.1016/S2214-109X(15)00192-8] [Medline: 26566744]
2. Anand S, Fan VY, Zhang J, Zhang L, Ke Y, Dong Z, et al. China's human resources for health: quantity, quality, and distribution. Lancet 2008 Nov 15;372(9651):1774-1781. [doi: 10.1016/S0140-6736(08)61363-X] [Medline: 18930528]
3. National Health Commission. Chinese Health Statistics Yearbook. Beijing: Beijing Union Medical University Press; 2019.
4. Yu W, Li M, Ye F, Xue C, Zhang L. Patient preference and choice of healthcare providers in Shanghai, China: a cross-sectional study. BMJ Open 2017 Nov 01;7(10):e016418 [FREE Full text] [doi: 10.1136/bmjopen-2017-016418] [Medline: 29092898]
5. Zhai Y, Gao J, Chen B, Shi J, Wang L, He X, et al. Design and Application of a Telemedicine System Jointly Driven by Videoconferencing and Data Exchange: Practical Experience from Henan Province, China. Telemed J E Health 2020 Jan 01;26(1):89-100. [doi: 10.1089/tmj.2018.0240] [Medline: 31050599]
6. Zheng Y, He Y, Xu Z, Pedrycz W. Assessment for hierarchical medical policy proposals using hesitant fuzzy linguistic analytic network process. Knowledge-Based Systems 2018 Dec;161:254-267. [doi: 10.1016/j.knosys.2018.07.005]
7. General Office of the State Council. Opinions on Promoting the Development of Internet + Medical Health. 2018. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm?trs=1 [accessed 2020-08-20]
8. China Internet Network Information Center. China Internet Network Development Statistic Report. 2014. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/P020140721507223212132.pdf [accessed 2020-08-20]
9. Villagonzalo K, Arnold C, Farhall J, Rossell SL, Foley F, Thomas N. Predictors of overall and mental health-related internet use in adults with psychosis. Psychiatry Res 2019 Aug;278:12-18. [doi: 10.1016/j.psychres.2019.05.034] [Medline: 31132571]
10. Masson CL, Chen IQ, Levine JA, Shopshire MS, Sorensen JL. Health-related internet use among opioid treatment patients. Addict Behav Rep 2019 Jun;9:100157 [FREE Full text] [doi: 10.1016/j.abrep.2018.100157] [Medline: 31193741]
11. Waring ME, McManus DD, Amante DJ, Darling CE, Kiefe CI. Online health information seeking by adults hospitalized for acute coronary syndromes: Who looks for information, and who discusses it with healthcare providers? Patient Educ Couns 2018 Nov;101(11):1973-1981 [FREE Full text] [doi: 10.1016/j.pec.2018.06.016] [Medline: 30305253]
12. Wiegers TA, Hendriks M, Malanda U, de Boer D. Users' Experiences With Web-Based Health Care Information: Qualitative Study About Diabetes and Dementia Information Presented on a Governmental Website. J Med Internet Res 2019 Jul 08;21(7):e11340 [FREE Full text] [doi: 10.2196/11340] [Medline: 31287066]

13. Wu D, Li Y. Online health information seeking behaviors among Chinese elderly. Library & Information Science Research 2016 Jul;38(3):272-279. [doi: 10.1016/j.lisr.2016.08.011]
14. Chang SJ, Im E. A path analysis of Internet health information seeking behaviors among older adults. Geriatr Nurs 2014;35(2):137-141. [doi: 10.1016/j.gerinurse.2013.11.005] [Medline: 24332965]
15. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003] [Medline: 22958935]
16. Li Y, Ye S, Zhou Y, Mao F, Guo H, Lin Y, et al. Web-Based Medical Information Searching by Chinese Patients With Breast Cancer and its Influence on Survival: Observational Study. J Med Internet Res 2020 Apr 17;22(4):e16768 [FREE Full text] [doi: 10.2196/16768] [Medline: 32301740]
17. Avdic D, Moscelli G, Pilny A, Sriubaite I. Subjective and objective quality and choice of hospital: Evidence from maternal care services in Germany. J Health Econ 2019 Dec;68(4):102229-102252. [doi: 10.1016/j.jhealeco.2019.102229] [Medline: 31521024]
18. Sahama T, Stranieri A, Butler-Henderson K, Golden I. Medical System Choice: Information That Affects the Selection of Healthcare Provider in Australia? Stud Health Technol Inform 2018;247:596-600. [Medline: 29678030]
19. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The Effect of Online Effort and Reputation of Physicians on Patients' Choice: 3-Wave Data Analysis of China's Good Doctor Website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
20. Dündar C. Health-seeking behavior and medical facility choice in Samsun, Turkey. Health Policy 2017 Sep;121(9):1015-1019. [doi: 10.1016/j.healthpol.2017.07.002] [Medline: 28734683]
21. Tayyari Dehbarez N, Lou S, Uldbjerg N, Møller A, Gyrd-Hansen D, Søgaard R. Pregnant women's choice of birthing hospital: A qualitative study on individuals' preferences. Women Birth 2018 Dec;31(6):e389-e394. [doi: 10.1016/j.wombi.2017.11.006] [Medline: 29198502]
22. Schuldt J, Doktor A, Lichters M, Vogt B, Robra B. Insurees' preferences in hospital choice-A population-based study. Health Policy 2017 Oct;121(10):1040-1046 [FREE Full text] [doi: 10.1016/j.healthpol.2017.08.004] [Medline: 28823608]
23. Lee YJ, Boden-Albala B, Jia H, Wilcox A, Bakken S. The Association Between Online Health Information-Seeking Behaviors and Health Behaviors Among Hispanics in New York City: A Community-Based Cross-Sectional Study. J Med Internet Res 2015;17(11):e261 [FREE Full text] [doi: 10.2196/jmir.4368] [Medline: 26611438]
24. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
25. Zwijnenberg NC, Hendriks M, Bloemendal E, Damman OC, de JJD, Delnoij DM, et al. Patients' Need for Tailored Comparative Health Care Information: A Qualitative Study on Choosing a Hospital. J Med Internet Res 2016 Nov 28;18(11):e297 [FREE Full text] [doi: 10.2196/jmir.4436] [Medline: 27895006]
26. China Health and Nutrition Survey. Data Sets From China Health and Nutrition Survey. 2020. URL: https://www.cpc.unc.edu/projects/china/data/datasets [accessed 2020-02-20]
27. Chen Z, Fan W. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. International Journal of Transportation Science and Technology 2019 Mar;8(1):43-52. [doi: 10.1016/j.ijtst.2018.10.001]
28. Liu X. Mixed-effects multinomial logit model for nominal outcomes. In: Methods and Applications of Longitudinal Data Analysis. Oxford: Academic Press; 2016:343-378.
29. Ene M, Leighton EA, Blue GL, Bell BA. Multilevel models for categorical data using SAS® PROC GLIMMIX: The basics. 2014. URL: http://support.sas.com/resources/papers/proceedings15/3430-2015.pdf [accessed 2020-08-20]
30. Yang Y, Zhang X, Lee PK. Improving the effectiveness of online healthcare platforms: An empirical study with multi-period patient-doctor consultation data. International Journal of Production Economics 2019 Jan;207:70-80. [doi: 10.1016/j.ijpe.2018.11.009]
31. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
32. Chung K, Cho HY, Kim YR, Jhung K, Koo HS, Park JY. Medical Help-Seeking Strategies for Perinatal Women With Obstetric and Mental Health Problems and Changes in Medical Decision Making Based on Online Health Information: Path Analysis. J Med Internet Res 2020 Mar 04;22(3):e14095 [FREE Full text] [doi: 10.2196/14095] [Medline: 32130139]
33. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
34. Turan N, Kaya N, Aydın G. Health Problems and Help Seeking Behavior at the Internet. Procedia - Social and Behavioral Sciences 2015 Jul;195:1679-1682. [doi: 10.1016/j.sbspro.2015.06.248]
35. Jones SL, Hadjistavropoulos HD, Gullickson K. Understanding health anxiety following breast cancer diagnosis. Psychol Health Med 2014;19(5):525-535. [doi: 10.1080/13548506.2013.845300] [Medline: 24102547]
36. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]

37. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Dec 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
38. Zhang X, Warren J, Gu Y. The Use of Chinese-language Internet Information about Cancer by Chinese Health Consumers. Sydney: University of Wollongong Faculty of Business; 2016 Dec Presented at: Proceedings of the 27th Australasian Conference on Information Systems (ACIS2016); 2016 December 5-7; Wollongong.
39. Ogasawara R, Katsumata N, Toyooka T, Akaishi Y, Yokoyama T, Kadokura G. Reliability of Cancer Treatment Information on the Internet: Observational Study. JMIR Cancer 2018 Dec 17;4(2):e10031 [FREE Full text] [doi: 10.2196/10031] [Medline: 30559090]
40. Bao Y, Hoque R, Wang S. Investigating the determinants of Chinese adult children's intention to use online health information for their aged parents. Int J Med Inform 2017 Jun;102:12-20. [doi: 10.1016/j.ijmedinf.2017.01.003] [Medline: 28495340]
41. Li Y, Song Y, Zhao W, Guo X, Ju X, Vogel D. Exploring the Role of Online Health Community Information in Patients' Decisions to Switch from Online to Offline Medical Services. Int J Med Inform 2019 Oct;130:103951. [doi: 10.1016/j.ijmedinf.2019.08.011] [Medline: 31473534]
42. Zeng Z, Zhu W, Ke R, Ash J, Wang Y, Xu J, et al. A generalized nonlinear model-based mixed multinomial logit approach for crash data analysis. Accid Anal Prev 2017 Feb;99(Pt A):51-65. [doi: 10.1016/j.aap.2016.11.008] [Medline: 27870986]
43. Ye F, Lord D. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. Analytic Methods in Accident Research 2014 Jan;1:72-85. [doi: 10.1016/j.amar.2013.03.001]
44. Viswanath K, McCloud R, Minsky S, Puleo E, Kontos E, Bigman-Galimore C, et al. Internet use, browsing, and the urban poor: implications for cancer control. J Natl Cancer Inst Monogr 2013 Dec;2013(47):199-205 [FREE Full text] [doi: 10.1093/jncimonographs/lgt029] [Medline: 24395992]

Abbreviations

CHNS: China Health and Nutrition Survey
HMP: hierarchical medical policy

Edited by G Eysenbach; submitted 03.03.20; peer-reviewed by A Stranieri, Y Si; comments to author 13.04.20; revised version received 04.08.20; accepted 10.08.20; published 03.09.20.

Please cite as:
Ma Q, Sun D, Cui F, Zhai Y, Zhao J, He X, Shi J, Gao J, Li M, Zhang W
Impact of the Internet on Medical Decisions of Chinese Adults: Longitudinal Data Analysis
J Med Internet Res 2020;22(9):e18481
URL: https://www.jmir.org/2020/9/e18481
doi:10.2196/18481
PMID:32880581

Original Paper

How People with Parkinson's Disease and Health Care Professionals Wish to Partner in Care Using eHealth: Co-Design Study

Carolina Wannheden[1], MSc, PhD; Åsa Revenäs[2,3], MSc, PhD

[1]Medical Management Centre, Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden

[2]Centre for Clinical Research, County of Västmanland, Uppsala University, Västerås, Sweden

[3]School of Health, Care and Social Welfare, Division of Physiotherapy, Mälardalen University, Västerås, Sweden

Corresponding Author:
Carolina Wannheden, MSc, PhD
Medical Management Centre
Department of Learning, Informatics Management and Ethics (LIME)
Karolinska Institutet
Tomtebodavägen 18a
Stockholm, 171 77
Sweden
Phone: 46 0852483934
Email: carolina.wannheden@ki.se

Abstract

Background: Worldwide, the number of people with Parkinson's disease (PD) is predicted to double between the years 2005 and 2030. Chronic care management requires active collaboration and knowledge exchange between patients and health care professionals (HCPs) for best possible health outcomes, which we describe as co-care. eHealth services have the potential to support the realization of co-care between people with PD (PwP) and HCPs.

Objective: This study aimed to explore how co-care could be operationalized in PD care, supported by eHealth. More specifically, this study explores PwP's and HCPs' expectations and desired eHealth functionalities to achieve co-care.

Methods: Principles of participatory design were used to enable the identification of co-care needs and design ideas, in a series of 4 half-day co-design workshops. The sample included 7 (4 women) PwP and 9 (4 women) HCPs, including 4 neurologists, 3 nurses, and 2 physiotherapists. The co-design process resulted in a functional prototype that was evaluated by the co-design participants in the last workshop. Data were collected through note cards produced by the participants during the first 3 workshops and focus group discussions during the 3rd and 4th workshops. The data were analyzed using qualitative thematic analysis. After the workshop series, the prototype was demonstrated at a Mini Fair for ongoing PD research and evaluated using a self-developed questionnaire with 37 respondents: 31 PwP (14 women) and 6 informal caregivers (3 women). Descriptive statistics are reported.

Results: The qualitative analysis of data resulted in 2 main themes. The first theme, core eHealth functionalities and their expected values, describes 6 desired eHealth functionalities for supporting PD co-care between PwP and HCPs: (1) self-tracking, (2) previsit forms, (3) graphical visualization, (4) clinical decision support, (5) self-care recommendations, and (6) asynchronous communication. The second theme, individual and organizational constraints, describes constraints that need to be addressed to succeed with an eHealth service for co-care. Individual constraints include eHealth literacy and acceptance; organizational constraints include teamwork and administrative workload. The majority of the questionnaire respondents (31/37, 84%) perceived that they would benefit from an eHealth service similar to the demonstrated prototype. All prototype functionalities were rated as very important or important by the majority of respondents (ranging from 86% to 97% per functionality).

Conclusions: This study adds to our knowledge on how PD co-care could be operationalized. Co-care implies a shift from episodic routine-driven care to more flexible care management that is driven by the mutual needs of patients and HCPs and supported by active information exchange between them, as well as automated information processing to generate patient-specific advice. More research is needed to further explore the concept of co-care in chronic care management and what it means for self-care and health care.

International Registered Report Identifier (IRRID): RR2-10.2196/11278

(J Med Internet Res 2020;22(9):e19195) doi:10.2196/19195

KEYWORDS

chronic care; Parkinson's disease; co-creation; co-design; participatory design; eHealth; mHealth; clinical decision support

Introduction

Chronic conditions affect more than 80% of people aged over 65 years in the European Union and represent a major challenge for health and social care systems [1]. Parkinson's disease (PD) is the second most common neurodegenerative disorder following Alzheimer's disease. It causes motor and nonmotor symptoms and results in significant burden for individual patients and their families, as well as health care and society [2]. Worldwide, the number of people with PD (PwP) is predicted to double between the years 2005 and 2030 [3]. Given this predicted increase and the limited availability of health care resources, self-management in everyday life is crucial for PwP as well as for people with other chronic conditions.

Co-Care

Chronic disease management requires a different practice of health care compared to the management of acute conditions [4]. This practice emphasizes both patients' and health care professionals' (HCPs') knowledge and active engagement for best possible health outcomes [5,6]. The term co-care, as defined by von Thiele Schwarz [7], emphasizes the use of appropriate tools, such as health information technologies, to enable the creation, shaping, sharing, and application of knowledge between different actors who are involved in an individual's care.

eHealth and e-Patients

eHealth refers to "health services and information delivered or enhanced through the internet and related technologies" [8]. The internet is an important resource for individuals with chronic conditions to acquire disease-specific knowledge [9] and also among PwP [10]. In 2004, Ferguson and Frydman [11] described patients and informal caregivers who sought online health guidance, for example through health communities, as the first generation of e-patients. Fifteen years later, the second generation of e-patients was described as patients who engage actively in their self-care and health care by producing and sharing their own health data as well as contributing to digital health innovations [12], which indicates a transition towards co-care.

eHealth in Parkinson's Disease Care

There is considerable evidence indicating that eHealth can be effective or at least promising in somatic care [13]. Identified values include real-time monitoring, better tailored personalized services, and patient empowerment [14]. A great number of PD applications are available to support individuals with diagnosis-specific information, assessments, and treatment [15]. In particular, mobile health technologies can support remote monitoring of PD motor symptoms by use of wireless motor sensors [16-18]. Such technologies have been widely recognized as promising [19,20], but the clinical utility of the self-tracked data and their value to improve health care requires further research [17,21,22].

It has been suggested that eHealth tools that support direct patient-provider communication may be more effective at improving patient self-management and self-efficacy [23]. PD technologies still have a tendency to prioritize the physician's perspective, while the needs of PwP and informal caregivers may not be fully supported [24]. To the best of our knowledge, how to enable PwP and HCPs to partner effectively in co-care, supported by eHealth, has not been described.

Aim

The aim of this study was to explore how co-care could be operationalized in PD care, supported by eHealth. More specifically, this study explores PwP's and HCPs' expectations and desired eHealth functionalities to achieve co-care.

Methods

Study Design and Participants

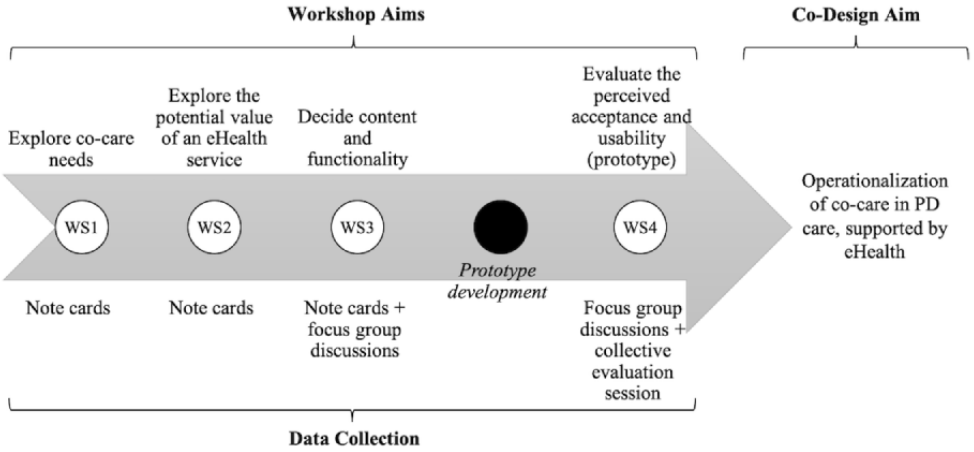
Principles of participatory design were used to enable the identification of PwP's and HCPs' views and expectations on co-care [25]. Participatory design shares similarities with action research and offers a method for combining health service and technology development in close collaboration with the intended users of the future service [25].

Co-Design Workshops

The collaborative work was performed in a series of 4 half-day co-design workshops during May and June of 2016 [26,27]. Participants in the co-design workshops included 7 PwP (4 women) and 9 HCPs (4 women). Among the HCPs, 3 were registered nurses, 4 were neurologists, and 2 were registered physiotherapists. The overall aim with the workshops was to explore and identify co-care needs, important functionalities in an eHealth service to enable co-care, and its potential impact for PD care. In workshops 1-3, participants engaged in co-design to explore needs and generate ideas (see Figure 1). In the third workshop, the participants prioritized functionalities to include in a functional prototype of an eHealth service. The prototype was developed by a software developer in the time period between the third and fourth workshops (3 weeks), in collaboration with the workshop facilitators and researchers. Content to include in the prototype was collected from the participants. The PwP contributed by sharing their most recent medication list and prescriptions through an anonymous questionnaire, and HCPs contributed by sharing self-care instructions and assessment instruments that are used in routine care. The PwP user interface was designed for mobile devices (smartphone or tablet), and the HCP user interface was designed for a computer screen (see Multimedia Appendix 1). In the fourth workshop, the participants discussed their perceived usability and acceptance of the prototype and potential impacts. More details about the recruitment process, participants, and

structure and content of the co-design workshops are described in [28]. The regional ethical committee approved the study (2015/2216-31/5).

Figure 1. Overview of the co-design workshops performed to explore how people with Parkinson's disease (PD) and health care professionals would like to partner in chronic care management, with particular emphasis on how eHealth could support them. WS: workshop.



Mini Fair for Parkinson's Disease Research

After the workshop series, in October 2016, we demonstrated the co-care prototype at a Mini Fair for ongoing Parkinson's disease research at Karolinska Institutet. After our demonstration, PwP and informal caregivers in the audience were invited to evaluate the prototype and its different functionalities in a questionnaire. The questionnaire was answered by 37 respondents: 31 PwP (14/31, 45% female) and 6 informal caregivers (3/6, 50% female). One of the responding PwP had also participated in the co-design workshops.

Data Collection and Analysis

Co-Design

User needs and design ideas were collected in workshops 1-3 through note cards produced by the participants in co-design sessions, based on the nominal group technique [29], and focus group discussions with HCPs and PwP separately (workshop 3) [30]. In workshop 4, the prototype was evaluated in focus group discussions, first separately with HCPs and PwP and then collectively with all participants. All workshops were audio recorded.

We followed the principles of a qualitative thematic analysis to analyze the data in 6 phases, using both an inductive and deductive approach [31]. In the first phase, the handwritten note cards created by individual participants (n=139) and the ones created collectively (n=83) were transcribed and labeled with the co-design session number and participant role where applicable. Selected parts from the co-design sessions were transcribed to complement the note cards when more descriptive details were needed for the analysis. The focus group discussions were transcribed verbatim. In the second phase, we identified meaning units in the data and generated initial codes to reflect their content. In the third phase, the meaning units with their initial codes were printed out on individual paper slips and sorted deductively into themes guided by the research question: (1) experienced needs, (2) desired functionalities of a co-care service, and (3) expected value. Subthemes were created inductively by grouping data according to similarities and differences. We first categorized data from the co-design sessions and thereafter added data from the focus group discussions. In the fourth phase, we reviewed and discussed the thematization of all data. When agreement was reached between the authors, the thematic map (ie, themes, subthemes, codes, and meaning units) was transferred to mind-mapping software (FreeMind version 1.0.1). In the fifth phase, themes and subthemes were refined and renamed in several iterations until they reflected a condensed analysis of the participants' expectations and desired eHealth functionalities to achieve co-care. In the final phase, we selected illustrative quotes that were translated from Swedish into English. Presented quotations are complemented with information about the source or respondent group (ie, note card, PwP or HCP) and workshop number (eg, WS1), which is provided in brackets.

Prototype Evaluation

The evaluation questionnaire contained 2 questions about general impressions and 7 questions about the perceived importance of different functionalities. The questions were answered using Likert-type response options with 5 levels. In 2 final questions, respondents were asked to list the 3 most important functionalities and were given the opportunity to provide their own suggestions. We analyzed the questions as Likert-type items, reporting variability as frequencies and percentage and central tendency as mode [32] (Multimedia Appendix 2).

Results

Co-Design

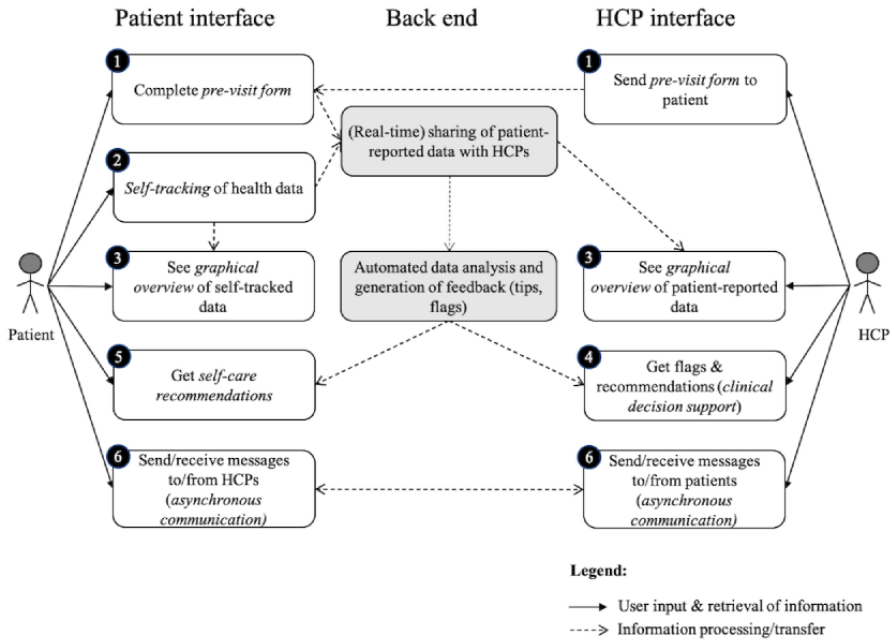
The qualitative analysis resulted in 2 main themes: (1) core eHealth functionalities and their expected values and (2) individual and organizational constraints.

Theme 1: Core eHealth Functionalities and Their Expected Values

We identified 6 core eHealth functionalities for supporting PD co-care between PwP and HCPs. These include a previsit form (desired functionality 1), patient self-tracking (desired functionality 2), graphical overview (desired functionality 3), clinical decision support (desired functionality 4), self-care recommendation (desired functionality 5), and asynchronous communication (desired functionality 6) and are illustrated in a use case diagram (Figure 2) and described in the following text.

Regarding desired functionality 1, an electronic previsit form was suggested as a PwP-facing and HCP-facing functionality that enables both PwP and HCPs to prepare for planned patient visits and use time during visits more efficiently. The participants discussed that collecting information that is necessary for assessing the PwP's health status during a visit can be a time-consuming activity. The dialogue in Textbox 1 illustrates how one of the physicians expressed their ambition and challenge of completing the Unified Parkinson's Disease Rating Scale (UPDRS) during visits.

Figure 2. Use case diagram describing the desired functionalities of an eHealth service to support co-care in Parkinson's disease care management, targeting the patient (left) and health care professionals (HCPs; right). The numbers in circles refer to the numbered eHealth functionalities in the text.



Textbox 1. Example of how one of the physicians expressed their ambition and challenge of completing the Unified Parkinson's Disease Rating Scale (UPDRS) during visits.

Physician 1: It is very challenging to find the time to complete it [UPDRS]. But when I do, I feel that I do – also for the long run – very high-quality work that is also easier to follow up.

[...]

Facilitator: Is this something the patient can do beforehand?

Physician 3: No, it's a status. Patients can help with other parts of it [UPDRS]. If there would be a validated translation in Swedish it could work.

Physiotherapist: Mhm, but there is none. [WS3]

The participants were of the opinion that electronic forms could be a way of collecting information from PwP prior to planned visits, which would be beneficial for both the PwP and HCPs. For example, main concerns and expectations for the upcoming visit, as well as a structured summary of experienced symptoms, medication intake, and diet could be reported beforehand. They pointed out that it would make work easier for HCPs and could support history taking and documentation:

> *Now patients come with their notes, which are not documented in any way [...] this may be a good instrument for documenting issues that may arise time and again.* [HCP, WS4]

One of the HCPs also pointed out that in this way, they would not need to remember to ask each of their patients about every possible issue on every occasion they meet, which would save time and enable more efficient consultations.

Regarding desired functionality 2, patient self-tracking was suggested as a PwP-facing functionality that allows PwP to track their own health and wellbeing on a daily basis. Self-tracking was discussed as a method that allows PwP to make their own measurements and register health-related parameters continuously whenever they experience a need for doing so. As one of the HCPs expressed, rather than limiting the collection of health data to the (few) instances when PwP

have scheduled visits with HCPs, self-tracking could be "a way to collect [data] and monitor one's health condition over many years" [HCP, WS4]. The PwP believed that they would not mind spending time on self-tracking and filling in self-assessment forms if this could benefit their health. On the contrary, they saw value in the ability to monitor their health parameters over time. As one of the PwP described, this would make it "easier to understand what is Parkinson's disease and what is something else" [note card, WS2]. Also, they believed that they would feel more confident in their communication with HCPs because the self-tracked data would make it easier to adequately describe how their condition has varied over time.

The participants believed that HCPs would also benefit from PwP's self-tracked data as the gathered information could make it easier to make adequate treatment decisions and provide insights about treatment adherence and effectiveness. If negative trends in health or wellbeing can be detected through self-tracking, they also anticipated an opportunity for more timely care interventions and prevention of undesired effects. Given these potential benefits, PwP expressed that the tracking and sharing of data would allow them to feel safer and calmer about their care.

Regarding desired functionality 3, the participants emphasized the need of a PwP-facing and HCP-facing graphical overview of health data collected through self-tracking and previsit forms. It should contain PwP's reported symptoms and wellbeing, prescribed and consumed medication, as well as their tracked self-care activities, such as number of steps per day or other physical activities. As one of the PwP pointed out:

> *If you were to get out the best of this type of system, I think it would be to see trends.* [PwP, WS4]

By visualizing trends, the PwP expected to gain more understanding about their situation and thereby gain insights about the effectiveness of their self-care efforts.

The participants, mainly HCPs, further emphasized the need to integrate patient-reported data with other data sources, such as data documented in the electronic health record and the national Parkinson's quality registry. As one of the HCPs pointed out:

> *But it is also convenient if everything is gathered in one and the same portal so that you don't need to find your way in different systems.* [HCP, WS4]

By gathering information from different sources in one overview, HCPs and PwP envisioned a strengthened collaboration also among HCPs. For example, a test result from the physiotherapist might be valuable for the physician when seeing the patient. Both HCPs and PwP also emphasized the possibility for improving collaboration with primary care:

> *And that possibly also primary care could get access to some of this, maybe not to use it, but to look into the system.* [PwP, WS4]

Regarding desired functionality 4, the need for HCP-facing clinical decision support functionality was emphasized by the HCPs. In particular, they desired automated guidance for planning when to schedule patient visits based on PwP's individual needs:

> *The system signals when it's time for a visit or a telephone contact. [...] If there are several issues, we may need to schedule an earlier consultation.* [HCP, WS4]

Automated flags and alerts were suggested as a type of decision support functionality that could support HCPs in planning consultations according to identified needs, rather than a routine care protocol that does not consider the health status of individual patients. For example, the HCPs desired to be notified through alerts or flags that indicate the occurrence of extraordinary events or negative trends based on patient-reported data. As one of them reflected:

> *When something does not follow the pattern, that you react earlier, and then it may be easier to manage; not wait until the next follow-up visit which may be months ahead, and then it turns into a big problem instead.* [HCP, WS4]

The participants emphasized that automated alerts and recommendations based on patient-reported data are necessary for HCPs to make individual adjustments in care plans as it would not be possible for them to actively monitor all the patients' self-tracking data. As one of the HCPs pointed out:

> *It [the anticipated eHealth system] needs to be so good that it simplifies - and not just by providing the doctor with information. It needs to be processed.* [HCP, WS1]

Regarding desired functionality 5, provision of self-care recommendations was one of the essential PwP-facing functionalities that the participants emphasized. As one of the PwP described it:

> *Self-care is something we do every day. All of us who have Parkinson's disease [...] It's when we get uncertain about our self-care that we need this - to look something up. It can be about symptoms, general wellbeing… some uncertainty we cannot manage on our own. That's when we need this.* [PwP, WS1]

The PwP expressed a desire for information about the causes and determinants of PD, symptoms, treatment, and ongoing research. Moreover, to be able to manage their own health with more confidence, PwP expressed that they need general as well as individually tailored recommendations for self-care. In particular, they desired recommendations regarding administration of medication, diet, physical activity, and exercises they could perform on a daily basis (eg, "What and how should I exercise?" or "What can I do to feel better?" [note cards, WS3]).

With adequate support, the participants expected that PwP could take responsibility for more of the activities important for their health and thus be more autonomous in their self-care. As one of the participants expressed:

> *It is possible to support patients in different ways – if there is something to, do they get more independent, I think.* [HCP, WS4]

One of the HCPs suggested that a sophisticated eHealth service would ideally generate simple self-care advice based on patient-reported data:

> *I imagine that a form of self-care could be that you feed in a bunch of data and then something pops out of the eHealth service. But I also understand that it is difficult to do. [...] But simple things: "Have you tried to take the medication without food?" or, if feeling nauseous: "Have you tried to take it with food?"* [HCP, WS1]

The possibility of a self-care scoring system and rewards was discussed as a means to motivate PwP to engage in their self-care. However, the idea of rewards and who should receive these was controversial. As one of the PwP emphasized:

> *Maybe health care should receive the movie theatre gift card instead [of patients], so that they log into the system – because I think that's where the resistance will be.* [PwP, WS4]

However, as one of the HCPs pointed out:

> *I think that can be intimidating for our staff. We are so pressured. If there is too much of that [rewards and bonus points], I think there will probably be many who choose not to use it [the system].* [HCP, WS4]

Regarding desired functionality 6, text-based messaging for asynchronous communication was suggested as PwP-facing and HCP-facing functionality. The PwP emphasized the need for continuous and maybe more frequent and regular contact with their HCPs. The participants discussed that a messaging service would make it easier for PwP to get in contact with their HCPs, for example to inform about newly experienced symptoms or to ask for renewals of prescriptions or health certificates — issues that may be resolved without having to meet face-to-face. However, the participants maintained that the idea is not to replace physical visits. As one of the PwP pointed out:

> *I just think that the planned visit with the health care professional is very important as a check-up — so that this is not systematized too much without noticing when the system goes to hell. The concept is good, but we need — the ultimate checkpoint is to see the patient in front of the health care professional.* [PwP, WS1]

The main expected benefit of asynchronous communication was better access to care and more timely support because PwP would be able to contact and reach health care at any time when a need occurs. As one PwP expressed:

> *This is a way to break into health care.* [PwP, WS4]

As one of the HCPs described:

> *I also think that, many patients we meet want to get in contact, it is difficult to call and no one responds, so it can be very convenient and calming to feel that "I have reported an issue; I have sent it and received a confirmation that it is sent to [HCP]."* [HCP, WS4]

Theme 2: Individual and Organizational Constraints

Two subthemes of constraints were identified that need to be addressed for succeeding with an eHealth service for co-care. These reflected constraints on an individual level (eg, relating to eHealth literacy and technology acceptance) and constraints on an organizational level (eg, how teamwork and collaboration between PwP and HCPs are organized).

Regarding the subtheme of eHealth literacy and acceptance, the necessity of eHealth literacy and acceptance was discussed as a major constraint of eHealth services. The participants pointed out that communicating through an eHealth service may not improve access to care for all PwP as some individuals may not be willing or able to use an eHealth service. As one of the HCPs pointed out:

> *Everyone will not be able to manage this and then we come to what [physician] said — that we need flexibility. There are those who will never have the energy to acquire the knowledge that is needed to manage this [anticipated eHealth system].* [HCP, WS1]

Specifically, motor symptoms of PD may cause difficulty filling in forms. Concerns were also raised in relation to nonmotor symptoms:

> *Unfortunately, one of the nonmotor symptoms is that you don't have the energy to fill in these forms. Parkinson-related fatigue is a concern.* [PwP, WS4]

An HCP also raised that some health problems may require informal caregivers' collaboration:

> *In case of hallucinosis or impulse control problems, patients won't report these issues themselves. In such cases, family members would need to report.* [HCP, WS4]

Regarding the subtheme of teamwork and administrative workload, the participants anticipated that an eHealth service for co-care may require HCPs to engage more in teamwork.

> *But they [health care] need to somehow organize themselves. This [anticipated eHealth service] could maybe enforce a more holistic perspective.* [PwP, WS4]

They emphasized the importance of organizing the team around the patient and clarifying roles and responsibilities:

> *It needs to be limited, so that the patient-reported issues do not end up being owned by many but addressed by no one.* [HCP, WS4)]

For example, they stressed that it is important to know who should take responsibility for corresponding with the patient.

The participants further raised that an eHealth service for patient-provider collaboration might cause additional administrative workload for HCPs and thereby limit the time available to interact face-to-face with patients. As one HCP expressed it:

> *My fear is that one gets tied up with the system and forgets about what is important, the patients.* [HCP, WS4]

As one of the PwP expressed:

> *If a nurse or physician communicates something to a patient, does this have to be documented in the patient's record or not? This is something health care will come down on immediately. And if it needs to be documented, this implies a double documentation burden for the HCP.* [PwP, WS4]

HCPs also feared that there is a risk that individual patients might overuse the opportunity to report health issues, which would increase HCPs' workload. Hence, this could scare HCPs from using this type of eHealth service. As one of the HCPs pointed out:

> *Because it is painful [...] not to be able to meet existing expectations. And there will be conflicts in the sense that we know that we would be able to do things better, but we can't.* [HCP, WS1]

Prototype Evaluation

The majority of the 37 questionnaire respondents (31/37, 84%) perceived that they would benefit from an eHealth service similar to the demonstrated prototype, while some (5/37, 14%) were neutral and one (1/37, 3%) saw no benefit. Using rewards was perceived as a benefit by 24 (24/37, 65%), while 12 (12/37, 32%) were neutral and one (1/37, 3%) saw no benefit. All prototype functionalities were rated as very important or important by the majority of respondents (ranging from 86% to 97% per functionality). Previsit forms were rated lowest with a mode of 4 (Important). For all other functionalities (self-tracking, graphical overview, self-care recommendations, asynchronous communication), the mode was 5 (Very important). The 3 functionalities that were rated most important were the ability to *send* messages to HCPs (asynchronous communication), graphical overview, and self-tracking. The ability to *receive* messages from HCPs got the fewest ratings as one of the 3 most important functionalities, followed by the previsit forms. The following additional functionalities were suggested: accumulated statistics, synchronization with new models of care, drug information and interaction alerts, and reminders to take medication. Details are presented in Multimedia Appendix 2.

Discussion

Principal Findings

The aim of this study was to explore how co-care could be operationalized in PD care, supported by eHealth. We identified 6 core eHealth functionalities that were desired by both PwP and HCPs, and all of them were rated important or very important in a group of PwP and informal caregivers that did not participate in the co-design. The co-design participants believed that these functionalities could contribute to higher quality of care in terms of safety, timeliness, effectiveness, efficiency, and patient-centeredness. Concerns that were raised included constraints on individual and organizational levels, which would need to be addressed to succeed with the implementation of a future co-care service.

Comparison With Previous Work

In the comparison of our results with prior work, we categorized the 6 desired functionalities into 3 groups, based on their purpose: (1) collection and sharing of health data (previsit forms, self-tracking), (2) feedback and recommendations (graphical overview, self-care recommendations, clinical decision support), and (3) asynchronous communication.

Functionalities for Collection and Sharing of Health Data

Self-tracking was described as functionality that is initiated and driven by PwP. In line with the participants' expectations, previous research has reported that self-tracking could contribute to a deeper understanding about PD manifestations among PwP and enhance both self-care and communication with health care, while also pointing out the importance for PwP to find a balance between the burdens and benefits of self-tracking [33]. The PwP in our study were not worried about the burden of frequent tracking, but the high attrition rate of eHealth services in general needs to be considered [34]. In contrast to self-tracking, previsit forms were suggested as a type of data collection that could be initiated by HCPs to save time during consultations. Previous research indicates that self-reported symptom assessments based on the UPDRS may be a viable option as patients' own assessments are not less reliable than clinicians' assessments [35]. Other patient-oriented assessment methods still need refinement (eg, the assessment of visual hallucinations, which are common among PwP [36], or assessments of cognitive impairments [37]). Our study participants emphasized the importance of including open questions in previsit forms that allow PwP to describe their main concerns in their own words. However, previous research has shown that discrepancy in patients' and HCP's perceptions of which information is important for health care staff to know and respond to may lead to disappointments [38]. A central aspect in the design of such functionalities is the "alignment of concerns" between patients and HCPs to ensure that the shared information is considered meaningful, actionable, and feasible from the perspectives of both patients and HCPs [39].

Functionalities for Providing Feedback and Recommendations

Feedback can support PwP in their self-care by providing insights on aggregated symptoms and medication data [40]. While there is an upsurge of self-monitoring applications available to support PD care, comprehensive systems that can support both the assessment of health data and provide treatment recommendations are limited [15,22]. A recent study identified that available eHealth solutions for PD do not always present graphical visualizations of self-tracked data to patients, which defeats the purpose of supporting individuals' self-care [24]. Research about clinical decision support in PD focuses largely on early detection and diagnosis of PD [41,42], but the participants in our study stressed the value of clinical decision support and self-care recommendations related to the management of already diagnosed PD. Active feedback,

including alerts and personalized recommendations, was considered essential to support individual needs-based health care and self-care. In relation to rewards, previous research corroborates our study participants' mixed feelings about game-based approaches to enhance motivation and suggests that alternative motivation strategies should be considered [43]. For HCP interfaces, the importance of workflow integration to support data-driven consultations has been emphasized [44]. A European Union–funded project has reported on the design of a clinical decision support system for PD that takes a holistic approach, which is in line with the functionalities suggested in this study [45-47]. The researchers emphasize the importance of enabling shared decision making [46]. Machine learning techniques and medical knowledge are used to generate alerts and suggest appropriate actions to patients, informal caregivers, and health care professionals [47].

Functionalities for Asynchronous Communication

To add flexibility in patient-provider collaboration and improve access to care, the participants in this study emphasized the need for functionality to support 2-way asynchronous communication between PwP and HCPs. Limitations in access to PD care in Europe are characterized by a lack of consultations with PD specialists [48], a need for more multidisciplinary care [49], and PD consultations that are based on clinicians' routines rather than driven by patient needs [50]. If eHealth is used effectively, it has been suggested that the traditional model of care with annual follow-up visits may in future be decomposed into several shorter needs-based consultations [51], which our study participants also expected. However, it has been reported that eHealth services for PD still have a tendency to prioritize the doctor's perspective, while patients' and informal caregivers' needs may not be fully supported [24]. For example, PwP are rarely able to initiate a consultation with their HCPs or signal that their treatment needs adjustment [24]. The PwP in our study emphasized the importance of being able to contact HCPs with free-text messages allowing them to ask questions or signal experienced needs or concerns. However, based on a Cochrane review, there is yet little evidence to justify the use of text messages to support self-management [52]. Therefore, despite the anticipated value of functionality for asynchronous communication that can be initiated by PwP, the potential risks and benefits need further research.

Addressing Individual and Organizational Constraints of Using eHealth for Co-Care

As has been shown in previous research, poor eHealth literacy and acceptance could lead to disparities [53]. In particular, factors such as age and disabilities have been negatively associated with the digital divide [54]. The median age of PD onset is 60 years [2]. In comparison, the median age of the PwP in our study was 73 years. To address symptom-related disabilities of PD, design guidelines for touch screen gestures have been suggested [55], as well as calibration of touch screen sensitivity [56]. Previous research also emphasizes the important role of informal caregivers in supporting self-care and symptom assessments [57,58]. Acknowledging caregivers' role in self-care, it has been suggested that eHealth services for PD should be designed to support collaboration between PwP and their informal caregivers [59]. While our study participants also regretted the absence of informal caregivers in the co-design process, they nevertheless expected that the main challenge ahead would be to engage HCPs. We acknowledge that the implementation and adoption of PD technologies in health care require integration with clinical workflow [19].

Limitations

A limitation of this study is that the prototyped eHealth service has not been implemented and evaluated in clinical practice. Thus, our study does not allow us to draw conclusions about the actual value of the desired eHealth functionalities for co-care, including cost-effectiveness and clinical outcomes. Nevertheless, we believe that the multistakeholder co-design workshops provided an effective forum for the PwP and HCPs to discuss and align their concerns, which has been described as a prerequisite for successful design and implementation [39]. This may have been confirmed by the high ratings of the importance of all co-designed functionalities in our evaluation questionnaire. However, these results should be interpreted with caution as the questionnaire was not based on a validated instrument. Our intention was to use the results to guide the next step in the design process, rather than as a summative evaluation of the prototype.

The transferability of our results is inevitably limited by contextual factors of the overall study design, the participant constellation, and the Swedish health care system and standards of care. As has already been discussed elsewhere [28], the constellation of participants had several shortcomings: We failed in our attempt to involve informal caregivers; most of the participating PwP were highly educated and experienced (ie, expert patients); and there were existing professional or patient-provider relationships between some of the participants. This may have influenced the results to not fully reflect the needs of all stakeholder groups. Because the participants were recruited from different health care organizations, a contextual analysis of (local) care workflows was not considered meaningful at this stage. Instead, we discussed care practices based on the national guidelines for PD care in Sweden [60]. Despite these limitations, we believe that our results capture experienced needs and desired eHealth functionalities for co-care that may be of relevance also in other PD settings, as well as other areas of chronic care management.

Conclusions

This study adds to our knowledge on how co-care in PD care could be operationalized. It provides a description of 6 core eHealth functionalities that were desired by PwP and HCPs to support co-care. Co-care implies a shift from episodic routine-driven care to more flexible care management that is driven by the mutual needs of patients and HCPs and supported by active information exchange between them, as well as automated information processing to generate patient-specific advice. While various eHealth applications have been developed and tested for different purposes, as of yet, we lack evidence for services that enable PD co-care by supporting the mutual needs and requirements of PwP, informal caregivers, and HCPs. More research is needed to further explore the concept and operationalization of co-care in chronic care management and

what it means for self-care, health care, and ultimately, individuals' health and wellbeing.

Acknowledgments

We give special thanks to all co-design participants for their valuable time and contributions. The authors also thank the following people for their support in different phases of the research: Per Svenningsson, Sara Riggare, and Eleonor Högström in the recruitment; Elena Eftimovska, Oscar Frykholm, Sara Tolf, Sara Riggare, Helena Hvitfeldt, and Emma Granström in the workshop planning and/or data collection; and Ulrica von Thiele Schwarz for valuable input to improve the manuscript. This research was funded by The Swedish Research Council for Health, Work Life and Welfare, grant number 2014-4238.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the co-care prototype that was developed.
[PPTX File , 5092 KB - jmir_v22i9e19195_app1.pptx]

Multimedia Appendix 2
Descriptive statistics of the evaluation questionnaire.
[DOCX File , 28 KB - jmir_v22i9e19195_app2.docx]

References

1. European Commission Health and Consumers Directorate-General. Conference Conclusions. The 2014 EU Summit on Chronic Diseases Brussels. Luxembourg: European Commission; 2014 Apr 3. URL: https://ec.europa.eu/health/sites/health/files/major_chronic_diseases/docs/ev_20140403_mi_en.pdf [accessed 2020-08-30]
2. Lees AJ, Hardy J, Revesz T. Parkinson's disease. The Lancet 2009 Jun 13;373(9680):2055-2066. [doi: 10.1016/S0140-6736(09)60492-X] [Medline: 19524782]
3. Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 2007 Jan 30;68(5):384-386. [doi: 10.1212/01.wnl.0000247740.47667.03] [Medline: 17082464]
4. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
5. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 09;288(14):1775-1779. [doi: 10.1001/jama.288.14.1775] [Medline: 12365965]
6. Batalden M, Batalden P, Margolis P, Seid M, Armstrong G, Opipari-Arrigan L, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul;25(7):509-517 [FREE Full text] [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
7. von Thiele Schwarz U. Co-care: Producing better health outcome through interactions between patients, care providers and information and communication technology. Health Serv Manage Res 2016 Apr 04;29(1-2):10-15 [FREE Full text] [doi: 10.1177/0951484816637746]
8. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
9. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How Can eHealth Technology Address Challenges Related to Multimorbidity? Perspectives from Patients with Multiple Chronic Conditions. J Gen Intern Med 2015 Aug 18;30(8):1063-1070 [FREE Full text] [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
10. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
11. Ferguson T, Frydman G. The first generation of e-patients. BMJ 2004 May 15;328(7449):1148-1149 [FREE Full text] [doi: 10.1136/bmj.328.7449.1148] [Medline: 15142894]
12. Scott Duncan T, Riggare S, Koch S, Sharp L, Hägglund M. From Information Seekers to Innovators: Qualitative Analysis Describing Experiences of the Second Generation of E-Patients. J Med Internet Res 2019 Aug 15;21(8):e13022 [FREE Full text] [doi: 10.2196/13022] [Medline: 31418421]
13. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
14. Berler A, Ainali C, Spyrou S, Chondropolous K, Lampros D, Dalasti C. Report on Studies Concerning Added Value of eHealth/mHealth Services. Joint Action to support the eHealth Network. 2017 Nov 28. URL: https://jasehn.eu/wordpress/

wp-content/uploads/2018/05/D7.3_REPORT_Studies_concerning_added_value_eHealth-mHealth_services.pdf [accessed 2020-08-30]
15. Linares-Del Rey M, Vela-Desojo L, Cano-de la Cuerda R. Mobile phone applications in Parkinson's disease: A systematic review. Neurologia 2019;34(1):38-54 [FREE Full text] [doi: 10.1016/j.nrl.2017.03.006] [Medline: 28549757]
16. Pan D, Dhall R, Lieberman A, Petitti DB. A mobile cloud-based Parkinson's disease assessment system for home-based monitoring. JMIR Mhealth Uhealth 2015 Mar 26;3(1):e29 [FREE Full text] [doi: 10.2196/mhealth.3956] [Medline: 25830687]
17. Ossig C, Antonini A, Buhmann C, Classen J, Csoti I, Falkenburger B, et al. Wearable sensor-based objective assessment of motor symptoms in Parkinson's disease. J Neural Transm (Vienna) 2016 Jan;123(1):57-64. [doi: 10.1007/s00702-015-1439-8] [Medline: 26253901]
18. Maetzler W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. Mov Disord 2013 Oct;28(12):1628-1637. [doi: 10.1002/mds.25628] [Medline: 24030855]
19. Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: Challenges and opportunities. Mov Disord 2016 Sep;31(9):1272-1282 [FREE Full text] [doi: 10.1002/mds.26642] [Medline: 27125836]
20. Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019 May;34(5):657-663 [FREE Full text] [doi: 10.1002/mds.27671] [Medline: 30901495]
21. Johansson D, Malmgren K, Alt Murphy M. Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: a mixed-methods systematic review. J Neurol 2018 Aug;265(8):1740-1752 [FREE Full text] [doi: 10.1007/s00415-018-8786-y] [Medline: 29427026]
22. Rovini E, Maremmani C, Cavallo F. Automated Systems Based on Wearable Sensors for the Management of Parkinson's Disease at Home: A Systematic Review. Telemed J E Health 2019 Mar;25(3):167-183. [doi: 10.1089/tmj.2018.0035] [Medline: 29969384]
23. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The Use and Effects of Electronic Health Tools for Patient Self-Monitoring and Reporting of Outcomes Following Medication Use: Systematic Review. J Med Internet Res 2018 Dec 18;20(12):e294 [FREE Full text] [doi: 10.2196/jmir.9284] [Medline: 30563822]
24. Nunes F, Andersen T, Fitzpatrick G. The agency of patients and carers in medical care and self-care technologies for interacting with doctors. Health Informatics J 2019 Jun;25(2):330-349 [FREE Full text] [doi: 10.1177/1460458217712054] [Medline: 28653552]
25. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct 01;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
26. Sanders E, Stappers P. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18 [FREE Full text] [doi: 10.1080/15710880701875068]
27. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
28. Revenäs Å, Hvitfeldt Forsberg H, Granström E, Wannheden C. Co-Designing an eHealth Service for the Co-Care of Parkinson Disease: Explorative Study of Values and Challenges. JMIR Res Protoc 2018 Oct 30;7(10):e11278 [FREE Full text] [doi: 10.2196/11278] [Medline: 30377143]
29. Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I. The nominal group technique: a research tool for general practice? Fam Pract 1993 Mar;10(1):76-81. [doi: 10.1093/fampra/10.1.76] [Medline: 8477899]
30. Fern EF. Advanced focus group research. Newbury Park, CA: Sage Publications, Inc; Jun 25, 2001.
31. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
32. Boone D, Boone H. Analyzing likert data. Journal of Extension 2012 Apr;50(2):1-5 [FREE Full text]
33. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
34. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11-e19 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
35. Little M, Wicks P, Vaughan T, Pentland A. Quantifying short-term dynamics of Parkinson's disease using self-reported symptom data from an Internet social network. J Med Internet Res 2013 Jan 24;15(1):e20 [FREE Full text] [doi: 10.2196/jmir.2112] [Medline: 23343503]
36. Jackson ML, Bex PJ, Ellison JM, Wicks P, Wallis J. Feasibility of a web-based survey of hallucinations and assessment of visual function in patients with Parkinson's disease. Interact J Med Res 2014 Jan 06;3(1):e1 [FREE Full text] [doi: 10.2196/ijmr.2744] [Medline: 24394559]
37. Piau A, Wild K, Mattek N, Kaye J. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care:

Systematic Review. J Med Internet Res 2019 Aug 30;21(8):e12785 [FREE Full text] [doi: 10.2196/12785] [Medline: 31471958]
38. Nielsen KD. Involving Patients with E-health – The Dialogic Dynamics of Information Filtration Work. Science & Technology Studies 2015 Jan 1;28(2):29 [FREE Full text]
39. Andersen T, Bansler J, Kensing F, Moll J, Mønsted T, Nielsen K, et al. Aligning Concerns in Telecare: Three Concepts to Guide the Design of Patient-Centred E-Health. Comput Supported Coop Work 2018 May 19;27(3-6):1181-1214 [FREE Full text] [doi: 10.1007/s10606-018-9309-1]
40. Memedi M, Tshering G, Fogelberg M, Jusufi I, Kolkowska E, Klein G. An Interface for IoT: Feeding Back Health-Related Data to Parkinson's Disease Patients. JSAN 2018 Mar 12;7(1):14 [FREE Full text] [doi: 10.3390/jsan7010014]
41. Pereira CR, Pereira DR, Weber SAT, Hook C, de Albuquerque VHC, Papa JP. A survey on computer-assisted Parkinson's Disease diagnosis. Artif Intell Med 2019 Apr;95:48-63. [doi: 10.1016/j.artmed.2018.08.007] [Medline: 30201325]
42. Belić M, Bobić V, Badža M, Šolaja N, Đurić-Jovičić M, Kostić VS. Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease-A review. Clin Neurol Neurosurg 2019 Sep;184:105442. [doi: 10.1016/j.clineuro.2019.105442] [Medline: 31351213]
43. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-Care Technologies in HCI. ACM Trans. Comput.-Hum. Interact 2015 Dec 14;22(6):1-45 [FREE Full text] [doi: 10.1145/2803173]
44. Kim Y, Heo E, Lee H, Ji S, Choi J, Kim JW, et al. Prescribing 10,000 Steps Like Aspirin: Designing a Novel Interface for Data-Driven Medical Consultations. 2017 May Presented at: 2017 CHI Conference on Human Factors in Computing Systems; Denver, CO; May 6–11, 2017 p. 5787-5799 URL: https://doi.org/10.1145/3025453.3025570 [doi: 10.1145/3025453.3025570]
45. Tsiouris KM, Gatsios D, Rigas G, Miljkovic D, Koroušić Seljak B, Bohanec M, et al. PD_Manager: an mHealth platform for Parkinson's disease patient management. Healthc Technol Lett 2017 Jun;4(3):102-108 [FREE Full text] [doi: 10.1049/htl.2017.0007] [Medline: 28706727]
46. Timotijevic L, Hodgkins CE, Banks A, Rusconi P, Egan B, Peacock M, et al. Designing a mHealth clinical decision support system for Parkinson's disease: a theoretically grounded user needs approach. BMC Med Inform Decis Mak 2020 Feb 19;20(1):34 [FREE Full text] [doi: 10.1186/s12911-020-1027-1] [Medline: 32075633]
47. Bohanec M, Miljković D, Valmarska A, Mileva Boshkoska B, Gasparoli E, Gentile G, et al. A decision support system for Parkinson disease management: expert models for suggesting medication change. Journal of Decision Systems 2018 May 09;27(sup1):164-172 [FREE Full text] [doi: 10.1080/12460125.2018.1469320]
48. Bloem BR, Stocchi F. Move for change part I: a European survey evaluating the impact of the EPDA Charter for People with Parkinson's disease. Eur J Neurol 2012 Mar;19(3):402-410 [FREE Full text] [doi: 10.1111/j.1468-1331.2011.03532.x] [Medline: 21967281]
49. Stocchi F, Bloem BR. Move for Change Part II: a European survey evaluating the impact of the EPDA Charter for people with Parkinson's disease. Eur J Neurol 2013 Mar;20(3):461-472 [FREE Full text] [doi: 10.1111/j.1468-1331.2012.03876.x] [Medline: 23034057]
50. Bloem BR, Stocchi F. Move for Change Part III: a European survey evaluating the impact of the EPDA Charter for People with Parkinson's Disease. Eur J Neurol 2015 Jan;22(1):133-41, e8. [doi: 10.1111/ene.12544] [Medline: 25196038]
51. Dorsey ER, Glidden AM, Holloway MR, Birbeck GL, Schwamm LH. Teleneurology and mobile technologies: the future of neurological care. Nat Rev Neurol 2018 May;14(5):285-297. [doi: 10.1038/nrneurol.2018.31] [Medline: 29623949]
52. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
53. Smith B, Magnani JW. New technologies, new disparities: The intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
54. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
55. Nunes F, Silva P, Cevada J, Correia Barros A, Teixeira L. User interface design guidelines for smartphone applications for people with Parkinson's disease. Univ Access Inf Soc 2015 Oct 27;15(4):659-679 [FREE Full text] [doi: 10.1007/s10209-015-0440-1]
56. de BA, Cevada J, Bayés A, Alcaine S, Mestre B. User-centred design of a mobile self-management solution for Parkinson's disease. 2013 Dec Presented at: 12th International Conference on Mobile and Ubiquitous Multimedia; December 2-5, 2013; Luley, Sweden p. 1-10. [doi: 10.1145/2541831.2541839]
57. Rastgardani T, Armstrong MJ, Marras C, Gagliardi AR. Improving patient-centred care for persons with Parkinson's: Qualitative interviews with care partners about their engagement in discussions of "off" periods. Health Expect 2019 Jun;22(3):555-564 [FREE Full text] [doi: 10.1111/hex.12884] [Medline: 30977262]
58. Hiseman J, Fackrell R. Caregiver Burden and the Nonmotor Symptoms of Parkinson's Disease. Int Rev Neurobiol 2017;133:479-497. [doi: 10.1016/bs.irn.2017.05.035] [Medline: 28802929]

59. Nunes F, Fitzpatrick G. Self-Care Technologies and Collaboration. International Journal of Human-Computer Interaction 2015 Jul 30;31(12):869-881 [FREE Full text] [doi: 10.1080/10447318.2015.1067498]
60. National Board of Health and Welfare. National guidelines for care of multiple sclerosis and Parkinson's disease. 2016 Dec 1. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-riktlinjer/2016-12-1.pdf [accessed 2020-08-30]

Abbreviations

HCP: health care professional
PD: Parkinson's disease
PwP: person/people with Parkinson’s disease
UPDRS: Unified Parkinson's Disease Rating Scale
WS: workshop

Edited by G Eysenbach; submitted 08.04.20; peer-reviewed by T Andersen, S Prior; comments to author 26.05.20; revised version received 04.07.20; accepted 26.07.20; published 21.09.20.

Please cite as:
Wannheden C, Revenäs Å
How People with Parkinson's Disease and Health Care Professionals Wish to Partner in Care Using eHealth: Co-Design Study
J Med Internet Res 2020;22(9):e19195
URL: http://www.jmir.org/2020/9/e19195/
doi:10.2196/19195
PMID:32955448

Original Paper

Understanding the Intention to Use Telehealth Services in Underserved Hispanic Border Communities: Cross-Sectional Study

Suad Ghaddar[1], PhD; Kristina P Vatcheva[2], PhD; Samantha G Alvarado[3], MSHS; Laryssa Mykyta[4], PhD

[1]Department of Health and Biomedical Sciences, University of Texas Rio Grande Valley, Edinburg, TX, United States

[2]School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, Brownsville, TX, United States

[3]School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, United States

[4]Social, Economic and Housing Statistics Division, Health and Disability Statistics Branch, US Census Bureau, Washington, DC, United States

Corresponding Author:
Suad Ghaddar, PhD
Department of Health and Biomedical Sciences
University of Texas Rio Grande Valley
EHABW 2.206
1201 W. University Dr.
Edinburg, TX,
United States
Phone: 1 956 665 5269
Email: suad.ghaddar@utrgv.edu

Abstract

Background: Despite the United States having one of the leading health care systems in the world, underserved minority communities face significant access challenges. These communities can benefit from telehealth innovations that promise to improve health care access and, consequently, health outcomes. However, little is known about the attitudes toward telehealth in these communities, an essential first step toward effective adoption and use.

Objective: The purpose of this study is to assess the factors that shape behavioral intention to use telehealth services in underserved Hispanic communities along the Texas-Mexico border and examine the role of electronic health (eHealth) literacy in telehealth use intention.

Methods: We used cross-sectional design to collect data at a community health event along the Texas-Mexico border. The area is characterized by high poverty rates, low educational attainment, and health care access challenges. Trained bilingual students conducted 322 in-person interviews over a 1-week period. The survey instrument assessed sociodemographic information and telehealth-related variables. Attitudes toward telehealth were measured by asking participants to indicate their level of agreement with 9 statements reflecting different aspects of telehealth use. For eHealth literacy, we used the eHealth Literacy Scale (eHEALS), an 8-item scale designed to measure consumer confidence in finding, evaluating, and acting upon eHealth information. To assess the intention to use telehealth, we asked participants about the likelihood that they would use telehealth services if offered by a health care provider. We analyzed data using univariate, multivariate, and mediation statistical models.

Results: Participants were primarily Hispanic (310/319, 97.2%) and female (261/322, 81.1%), with an average age of 43 years. Almost three-quarters (219/298) reported annual household incomes below $20,000. Health-wise, 42.2% (136/322) self-rated their health as fair or poor, and 79.7% (255/320) were uninsured. The overwhelming majority (289/319, 90.6%) had never heard of telehealth. Once we defined the term, participants exhibited positive attitudes toward telehealth, and 78.9% (254/322) reported being somewhat likely or very likely to use telehealth services if offered by a health care provider. Based on multivariate proportional odds regression analysis, a 1-point increase in telehealth attitudes reduced the odds of lower versus higher response in the intention to use telehealth services by 23% (OR 0.77, 95% CI 0.73-0.81). Mediation analysis revealed that telehealth attitudes fully mediated the association between eHealth literacy and intention to use telehealth services. For a 1-point increase in eHEALS, the odds of lower telehealth use decreased by a factor of 0.95 (5%; OR 0.95, 95% CI 0.93-0.98; $P<.001$) via the increase in the score of telehealth attitudes.

Conclusions: Telehealth promises to address many of the access challenges facing ethnic and racial minorities, rural communities, and low-income populations. Findings underscore the importance of raising awareness of telehealth and promoting eHealth literacy as a key step in fostering positive attitudes toward telehealth and furthering interest in its use.

(J Med Internet Res 2020;22(9):e21012) doi:10.2196/21012

KEYWORDS

telehealth; eHealth literacy; health information technologies

Introduction

The United States has one of the leading health care systems in the world, offering highly specialized and technologically advanced medical care. At the same time, the US health care system faces many challenges, especially when serving vulnerable communities (eg, low socioeconomic status groups, minority populations, and uninsured people). Primary among these challenges is access to care, including lack of health insurance coverage [1], shortages of primary and specialty care providers [2], transportation difficulties [3], and language barriers [4,5], among others. With recent advances in technology, telehealth promises to address many of these access challenges.

Telehealth is commonly defined as "the use of electronic information and telecommunications technologies to support and promote long-distance clinical health care, patient and professional health-related education, public health, and health administration" [6]. This definition encompasses a broad scope of remote health care services (eg, telemedicine, telemonitoring, mobile health [mHealth] apps, patient portals). For the purposes of this study, telehealth as presented to participants and supported by the statements assessing their attitudes is more representative of telemedicine rather than other health information technologies (HIT). We opted to use the term telehealth in the survey because of its broader scope and higher likelihood of public recognition.

Several models, such as the technology acceptance model [7] and the unified theory of acceptance and use of technology [8], have been developed to depict consumer interest and willingness to use technology. With the expansion of technology into the health care sector, these models, with various modifications, have been applied to the adoption of HIT such as mHealth apps, patient portals, telemonitoring, and telemedicine [9]. Only recently have efforts been directed at developing technology use models specific to the health context [10,11]. In addition to the common key concepts across previous models (attitudes, behavioral intention, and behavior), HIT-specific models expand the focus from the technology's features to incorporate end-user characteristics (eg, health status, internet self-efficacy) and the realm of social influence. The role of electronic health (eHealth) literacy, commonly defined as "consumers' combined knowledge, comfort, and perceived skills at finding, evaluating, and applying electronic health information to health problems" [12], has rarely been integrated in these models [11]. Yet evidence has been mounting in support of its role in the adoption of various HIT apps such as patient portals [13,14] and mHealth [15,16]. In line with conceptual models depicting end-user characteristics influencing behavioral intention to use HIT through the mediating effects of perceived ease of use, perceived usefulness, and attitudes [10], we hypothesize that eHealth literacy will exhibit a similar indirect effect on the behavioral intention to use telehealth.

Most telehealth research in the United States has initially focused on telehealth adoption from the perspective of health care providers [17-21] and health systems [20,22] or on policies [23] and reimbursement models that facilitate its adoption [24]. The consumer/patient perspective has just recently been more extensively considered. However, most patient research on telehealth has been conducted at the international level [21,25,26] or has focused on white, non-Hispanic populations within the United States [27]. The perspective and characteristics of individuals from vulnerable US minority communities has not received much attention, although there are a few notable exceptions focused primarily on African Americans [28,29]. Given that Hispanics are the largest minority group in the United States [30], it is important to understand the factors that influence their acceptance of telehealth, especially given that acceptance is a predictor of adoption [31]. This entails examining several dimensions, primary among which are end-user characteristics and attitudes toward health technology. Thus, the purpose of this study is to assess the factors that shape behavioral intentions to use telehealth services in vulnerable, marginalized Hispanic communities along the Texas-Mexico border and examine the role of eHealth literacy in telehealth intention use.

Methods

Study Setting

We collected data from participants at Operation Lone Star (OLS), a joint military and civilian public health emergency preparedness exercise that takes place annually along the Texas-Mexico border. OLS is a cooperative effort between the Texas Department of State Health Services, Cameron County Public Health Department, Hidalgo County Health and Human Services, City of Laredo Health Department, the Texas State Guard, and various community volunteer organizations. The event brings free health care services to area residents. These include child immunizations, sports physicals, hearing screenings, vision screenings for prescription glasses, diabetes and blood pressure screenings, and dental services, among others.

In 2018, OLS events took place during the week of July 23-27 at 6 locations across 4 border counties (Cameron, Hidalgo, Starr, and Webb). For this study, we collected data at one of the two Hidalgo County sites, which provided services to 2294 children and adult county residents over the course of OLS week. Hidalgo

County is the largest county along the Texas-Mexico border; it is home to almost 850,000 people, the overwhelming majority of whom are of Hispanic or Latino origin (92%) [30]. The county is characterized by high poverty rates (almost a third of the population lives below the federal poverty level) and low educational attainment (36% of individuals aged 25 years and over do not have a high school degree) [30]. Lack of health care coverage is a major access challenge with 43% of individuals aged 18 to 64 years being uninsured in 2018 [32].

Recruitment and Data Collection

Data were collected in person by students participating in a special course-based undergraduate research experience, two graduate research assistants, and the first author. All data collection team members completed training in the ethical conduct of research, survey administration, and interviewing techniques, as well as additional requirements for participation at OLS. Most team members were bilingual (English and Spanish).

We employed a convenience sampling design to recruit participants. The data collection team approached event attendees waiting to receive health services at various stations, provided them with information about the study and invited them to participate. The 15- to 20-minute interviews were conducted in either English or Spanish, based on the participant's preferred language. After completing the anonymous interview, participants were provided with a drawstring bag, a bottle of water, and a chance to enter in a raffle for one of sixty $50 gift cards from a local grocery store.

All study procedures were approved by the institutional review board at the University of Texas Rio Grande Valley (UTRGV).

Survey Instrument

The survey instrument included questions assessing sociodemographic information, health status, eHealth literacy, and telehealth-related variables measuring attitudes and behavioral intentions to use telehealth. For sociodemographic characteristics and health status variables, we used questions from existing national surveys (eg, US Census Bureau, Centers for Disease Control and Prevention [CDC]) for which existing Spanish translation was available. For the remaining variables, where no Spanish translation was available, a bilingual (English and Spanish) graduate student translated the survey. The survey was then piloted by bilingual and native Spanish-speaking interviewers who assessed participant understanding of both the English and Spanish versions. Minor modifications were made to reflect the area's culture and local linguistic Spanish use.

We provided participants with the following definition before asking about the two main telehealth-related variables (outcome measure: behavioral intention to use telehealth; predictor measure: attitudes toward telehealth): Telehealth uses technology to access and manage health care outside of doctors' offices or clinics. Some examples are receiving care from your health care provider by video, remote monitoring of blood pressure or heart rate, or checking your laboratory results online.

Outcome Measure: Behavioral Intention to Use Telehealth Services

We assessed the behavioral intention to use telehealth by the question, "How likely are you to use telehealth services if they were offered by your provider?" Response options included: very likely, somewhat likely, not very likely, and would not use telehealth services.

Predictor Measures

Attitudes Toward Telehealth

We assessed attitudes toward telehealth by asking respondents to rate their level of agreement (5-point Likert scale: 1=strongly disagree to 5=strongly agree) with 9 statements reflecting different aspects of telehealth use such as perceived ease of use, perceived usefulness, and perceived cost effectiveness. The statements were adopted with minor modifications from a study on patient telemedicine readiness in a Louisiana oncology practice [28]; the instrument was developed based on the technology acceptance model [7,33] and the fit between individuals, task, and technology framework [34]. The summated scale (range 9 to 45 with higher scores reflecting more positive attitudes) demonstrated good internal consistency (Cronbach alpha of .794); the internal consistency was lower for the Spanish surveys (Cronbach alpha of .757) as compared with the English surveys (Cronbach alpha of .839) but still above the acceptable .70 threshold value [35].

Telehealth Readiness

We assessed telehealth readiness with 3 questions related to the methods participants used to (1) make an appointment with their health care provider, (2) communicate with their health care provider, and (3) keep track of their personal health information. A participant was considered telehealth-ready if they had communicated with their health care provider to make an appointment or discussed their test results via a website and/or email or kept track of their personal health information using an online system.

eHealth Literacy

We used the 8-item eHealth Literacy Scale (eHEALS) to assess eHealth literacy [12]. For each item, respondents indicated their level of agreement on a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Higher scores on the summation of responses (range 8 to 40) reflect higher levels of eHealth literacy. The reliability and validity of eHEALS has been previously established in both English and Spanish [12,36]. Cronbach alpha, assessing the internal consistency of eHEALS, was .916 for our sample (.904 and .924 for the English- and Spanish-speaking subsamples, respectively).

Sociodemographic Variables

We assessed several sociodemographic variables, including age, sex, employment status, marital status, educational attainment, country of birth, ethnicity, annual household income, and health care coverage.

Internet Knowledge and Skills

We used 2 questions to examine participant internet knowledge and skills. The first question (dichotomized yes/no) asked

participants whether they have "ever gone online to access the internet or to send and receive emails." The second question asked respondents to self-rate their internet skills. Responses were dichotomized as yes=fairly skilled, very skilled, or expert and no=not at all skilled or not very skilled.

Health Status

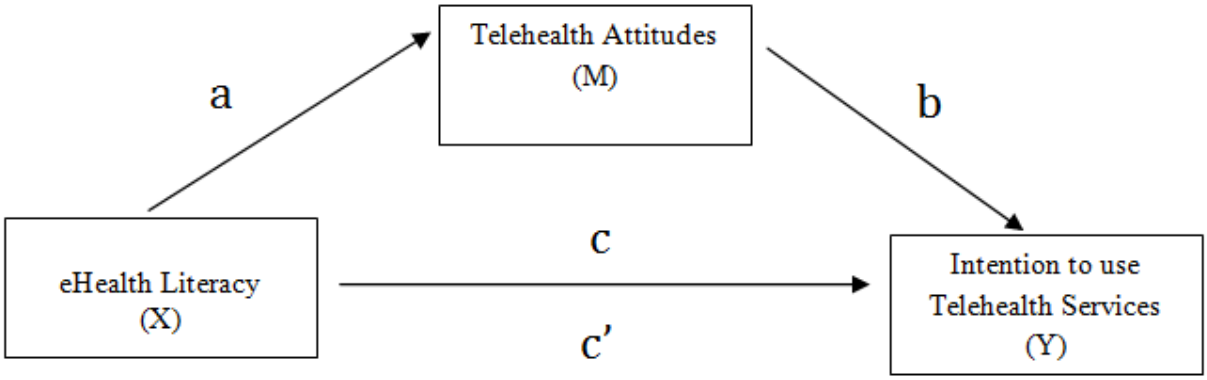
To assess the health of participants, we used 2 measures. The first measure examined general health status through the CDC's validated self-rated health status measure [37,38]: "Would you say that in general your health is?" The 5 response categories were excellent, very good, good, fair, and poor. We dichotomized the responses: 1=fair or poor health versus 0=otherwise. The second measure, chronic condition, assessed the presence of chronic health conditions by asking participants whether they have ever been told by a health professional that they had diabetes, asthma, heart disease, cancer, or arthritis. Those reporting a diagnosis of one or more conditions were coded as 1=having at least one chronic condition versus 0=otherwise.

Statistical Analysis

We used descriptive analyses (frequencies and percentages for categorical variables and means and standard deviations for continuous variables) to summarize and examine study data by the levels of the outcome variable (intention to use telehealth services). To take into account the ordinal nature of the outcome variable (that there is a clear ordering of the levels but with unknown absolute distances between them) and the fact that the number of the ordered levels is fewer than 5, we conducted bivariate and multivariate proportional odds regression analyses [39,40]. To identify independent factors associated with the intention to use telehealth services, we fitted proportional odds regression models with variables selected based on the bivariate analyses and controlling for potential confounders. We assessed for potential multicollinearity and 2-way interactions between the variables included in the models. Model-based adjusted odds ratios (ORs) for lower versus higher response levels for the intention to use telehealth services and their respective 95% confidence intervals were estimated. The assumption of the proportional odds model that the effects of any explanatory variables are proportional across any response levels were tested using the score test and likelihood ratio test. To select a good model, we used Akaike information criterion.

To test for the potential mediation effect of telehealth attitudes on the effect of eHealth literacy on the intention to use telehealth services (Figure 1), we conducted mediation analysis. A mediator is a variable M (eg, telehealth attitudes) that falls into the casual pathway between an independent variable X (eg, eHealth literacy) and an outcome variable Y (eg, intention to use telehealth services) and at least partially explains the effects of X on Y. To examine a potential mediation effect, we decomposed the total effect of eHealth literacy on the intention to use telehealth services (c path) into two causal paths, direct effect (c′ path between eHealth literacy and intention to use telehealth services not passing through telehealth attitudes) and indirect effect (path between eHealth literacy and the intention to use telehealth services passing through telehealth attitudes) [41].

Figure 1. Path diagram of the hypothesized mediation model.



The focus of the mediation analysis was to evaluate and estimate the indirect effect of eHealth literacy on the intention to use telehealth services using the product of coefficients approach, c=ab and employing the method proposed by VanderWeele et al [42] by fitting the regression models (1) and (2) below in the settings of an ordinal outcome and a continuous mediator:

where equation (1) is a linear regression equation for the continuous mediator M on the explanatory variable X and a covariate Z with intercept i_0 and slopes a and b, respectively, and a random error $\varepsilon \sim N(0,\sigma^2)$; and equation (2) is a proportional odds regression model for the log of probability of a smaller response $Y \leq j$ compared with the probability of a larger response $Y > j$ on independent variable X with regression coefficient c′, mediator variable M with regression coefficient b, and covariate Z with regression coefficient d^*, l_j are the intercepts, and j=1,2,3,4 is the number of the ordered categories in the outcome variable Y.

Under the assumptions that the reference level in the outcome variable is common, the model is correctly specified, and the mediator follows a normal distribution with a constant

conditional variance, the natural indirect effect (NIE) and natural direct effect (NDE) for an exposure X on the outcome Y comparing any two X=x and X=x* for a proportional odds model are given by:

NIE ≈ exp {ab(x–x*)} (**3**)

NDE ≈ exp {c′(x–x*)} (**4**)

The standard errors of the log of the aforementioned effects were estimated using the Delta method [42,43].

Proportion of mediated effect in the total effect was computed as:

Multivariate regression and mediation analyses were conducted using complete case analysis under the missing at random (MAR) assumptions. The 294 participants used to build the model did not differ from the original sample in terms of sociodemographic characteristics: age, gender, education level, marital status, employment status, language, income, and health insurance status.

All statistical analyses were generated using SAS software version 9.4 (SAS Institute Inc). All statistical tests were 2-sided and performed at .05 significance level.

Results

Sample Characteristics

Table 1 reports sample characteristics by the levels of the outcome variable (intention to use telehealth services). Most of the participants were female (261/322, 81.1%), Hispanic (310/319, 97.2%), and of low socioeconomic status (219/298, 73.5% reported annual household incomes below $20,000). Only 59.3% (191/322) were high school graduates and 41.6% (131/315) were employed. Around a third (117/320, 36.6%) were born in the United States; a similar percentage (126/322, 39.1%) chose to complete the survey in English. Over half of participants (181/322, 56.2%) were married. The average age was 43 (SD 14.1) years. Not surprisingly, given the nature of the event where free health care services are the main attraction, only a fifth (65/320, 20.3%) were insured. A large proportion (136/322, 42.2%) self-rated their health as fair or poor, considerably higher than the corresponding numbers of 30% and 21% at the county and state levels, respectively [44]. Over a third (124/321, 38.6%) indicated having at least one chronic health condition.

Regarding familiarity with technology, only 59.9% (193/322) indicated having gone online to access the internet or send/receive emails. Almost 40% (118/306, 38.6%) self-rated their internet skills as not skilled at all or not very skilled. As a result, a small proportion of participants exhibited telehealth readiness: only 23.9% (77/322) reported having used technology to communicate with their health care provider or keep track of their personal health information. While only 9.4% (30/319) had heard of telehealth, once provided with a telehealth definition, 4 in 5 respondents were either very likely (137/322, 42.5%) or somewhat likely (117/322, 36.3%) to use telehealth services, if offered by their health care provider. The remainder were either not very likely (25/322, 7.8%) or indicated that they would not use these services (43/322, 13.4%). The average eHEALS score was 28 (SD 7.1, range 8 to 40) and the average telehealth attitudes score was 30 (SD 5.9, range 13 to 45).

Bivariate analyses (Table 1), using proportional odds regressions to predict lower versus higher intentions to use telehealth services, revealed that only country of birth (P=.01), language in which survey was administered (P=.01), eHEALS score (P=.03), and telehealth attitudes (P<.001) were significantly associated with the intention to use telehealth services.

Overall, respondents reported positive attitudes toward the use of telehealth. The majority indicated agreement with the idea that telehealth can save time and money and provide access to specialized care (Figure 2). An area of concern for more than half of respondents (169/321, 52.6%) was related to their ability to understand the physician through a telehealth video or call.

The significant variables based on bivariate analysis, with the exception of language, were then included in a multivariate proportional odds regression model. Given the high level of collinearity between language and country of birth (r=.732), we opted to include only the country of birth for its more objective level of measurement relative to the language variable. The language variable was based on the language in which participants chose to complete the survey (English or Spanish); while indicative of language preference, it does not necessarily measure the level of English language proficiency or correspond to language measures in other studies or national datasets. Country of birth, on the other hand, is a standard measure and allows for comparability across studies [45]. To ensure that the language variable did not impact the results differently, we repeated the analysis with language instead of country of birth and the results were similar.

Table 1. Sample characteristics by the levels of the intention to use telehealth services if offered by a health care provider.

Variables	Total sample (n=322)	Very likely (n=137)	Somewhat likely (n=117)	Not very likely (n=25)	Would not use (n=43)	*P* value[a]
Categorical, n (%)						
Female	261 (81)	111 (81)	91 (78)	21 (84)	38 (88)	.54
Hispanic	310 (97)	133 (98)	112 (97)	24 (100)	41 (95)	.58
Income[b] <$20K	219 (74)	100 (79)	71 (65)	18 (72)	30 (81)	.33
High school graduate	191 (59)	78 (57)	71 (61)	19 (76)	23 (54)	.56
Employed	131 (42)	53 (40)	56 (48)	12 (50)	10 (24)	.70
US-born	117 (37)	39 (29)	49 (42)	11 (46)	18 (43)	.01
English survey	126 (39)	40 (29)	51 (44)	14 (56)	21 (48)	.01
Married	181 (56)	77 (56)	66 (56)	11 (44)	27 (63)	.92
Insured	65 (20)	29 (21)	22 (19)	4 (16)	10 (23)	.84
Health status						
Fair/poor	136 (42)	55 (40)	48 (41)	9 (36)	24 (56)	.26
Chronic condition	124 (39)	57 (42)	42 (36)	9 (36)	16 (37)	.37
Internet skills						
Gone online[c]	193 (60)	76 (56)	77 (66)	19 (76)	21 (49)	.45
Not/not very skilled	118 (39)	56 (43)	35 (31)	9 (38)	18 (46)	.53
Telehealth						
Telehealth ready	77 (24)	33 (24)	30 (26)	6 (24)	8 (19)	.69
Heard of telehealth	30 (9)	11 (8)	14 (12)	4 (16)	1 (2)	.92
Continuous, mean (SD)						
Age in years	43 (14)	44 (14)	41 (14)	41 (12)	46 (17)	.62
eHEALS[d]	28 (7)	28 (7)	28 (7)	29(6)	25 (9)	.03
Telehealth attitudes	30 (6)	34 (5)	30 (4)	28 (5)	23 (6)	<.001

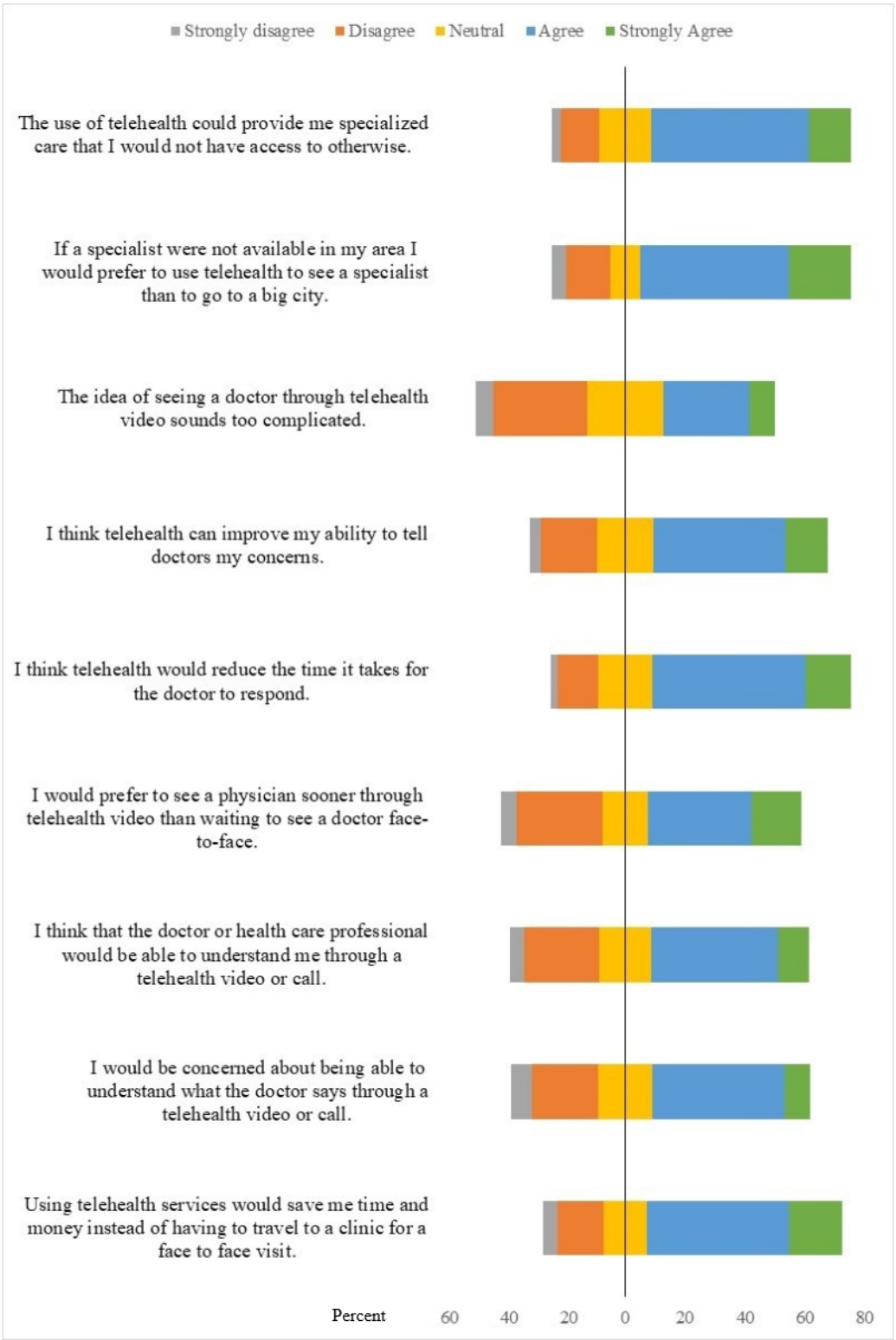
[a]Wald chi-square test.

[b]Annual household income.

[c]Gone online to access the internet or to send and receive emails.

[d]eHEALS: eHealth Literacy Scale.

Figure 2. Participant responses to attitudinal statements related to telehealth use.



As shown in Table 2, eHEALS was no longer significantly associated with the intention to use telehealth (*P*=.68). The telehealth attitudes variable maintained significance, indicating that a 1-point increase in telehealth attitudes reduced the odds of lower versus higher response in the intention to use telehealth services by 23% (OR 0.77, 95% CI 0.73-0.81). In addition, US-born participants, compared with foreign-born participants, had 2.20 (95% CI 1.35-3.58) times higher odds of lower versus higher response in the intention to use telehealth services.

Table 2. Multivariate proportional odds regression model of factors associated with lower versus higher response to the intention to use telehealth services (n=294).

Variable	Coefficient estimate (SE)	OR[a] (95% CI)	*P* value[b]
eHEALS[c]	0.01 (0.02)	1.01 (0.97-1.04)	.68
Telehealth attitudes	–0.28 (0.03)	0.77 (0.73-0.81)	<.001
US-born	0.79 (0.25)	2.20 (1.35-3.58)	.002

[a]OR: odds ratio.

[b]Wald chi-square test.

[c]eHEALS: eHealth Literacy Scale.

Mediation Analysis

Bivariate analysis (Table 1) revealed that eHealth literacy, as measured by eHEALS, was significantly associated with the intention to use telehealth services (*P*=.03). After controlling for the effect of country of birth, eHEALS remained significantly associated with the outcome (*P*=.02). The assumptions of the linear regression model (telehealth attitudes regressed on eHEALS and country of birth) for normal (Shapiro-Wilk test *P* value=.07) and homoscedastic errors (White test *P* value=.16) were satisfied. Linear regression analysis showed that eHEALS was significantly associated with the hypothesized mediator, telehealth attitudes (*P*=.002), controlling for the effect of country of birth (Table 3). Based on the proportional odds regression model, eHEALS was no longer associated (*P*=.68) with the intention of telehealth use after adjusting for telehealth attitudes and country of birth (Table 3). This indicated that telehealth attitudes fully mediated the association between eHEALS and intention to use telehealth services. Using the Delta method, the estimated NIE of eHEALS on the intention to use telehealth was significant (OR 0.95, 95% CI 0.93-0.98; *P*<.001). For a 1-point increase in eHEALS, the odds of lower use of telehealth services decreased by a factor of 0.95 (5%) via the increase in the score of telehealth attitudes, controlling for the effect of country of birth. The estimated proportion of mediated effect of eHEALS in the total effect was 117.87% (Table 3). The fact that the direct effect *c′* was opposite in sign to the indirect *ab* is known as inconsistent mediation [46] because the mediator acts like a suppressor variable. For the same reason, the estimated proportion of the mediated effect was greater than 1 [46].

Table 3. Adjusted estimated effects based on mediation analysis conducted with linear and proportional odds regression models (n=294).[a]

Variable	Coefficient estimate (SE)	OR[b] (95% CI)	*P* value
eHealth literacy (*a* coefficient)	0.18 (0.05)	N/A[c]	.002
Telehealth attitudes (*b* coefficient)	–0.28 (0.03)	0.77 (0.73-0.81)	<.001
eHealth literacy (direct effect, *c′* coefficient)	0.01 (0.02)	1.01 (0.97-1.04)	.68
eHealth literacy (indirect effect, *ab* coefficients)	–0.05 (0.01)	0.95 (0.93-0.98)	<.001

[a]All effects are adjusted for country of birth.

[b]OR: odds ratio.

[c]Coefficient was estimated using linear regression.

Discussion

Principal Findings

This study examined the attitudes of vulnerable minority groups toward telehealth and assessed the factors that shape their intention to use telehealth services. Findings revealed that marginalized Hispanic communities along the Texas-Mexico border have had limited exposure to telehealth. Despite that, participants exhibited generally positive attitudes toward telehealth which, in turn, were associated with a higher likelihood of using telehealth services if offered by one's health care provider. Additionally, these positive attitudes mediated the relationship between eHealth literacy and the intention to use telehealth, highlighting the important role that eHealth literacy plays in shaping attitudes and, ultimately, telehealth acceptance.

The positive association between attitudes toward telehealth and intention to use telehealth services is in line with the basic concept underlying the different technology acceptance frameworks [7,47,48], where reactions to using a certain information technology (attitudes, perceived usefulness, perceived ease of use) impact the behavioral intention to use that technology. Studies specific to HIT also support that pathway [10,25,26].

While health information technology acceptance and adoption models have evolved over time to integrate end-user characteristics, these models have rarely considered eHealth literacy. Kim and Park [10] include the technological/computer literacy domain of eHealth literacy [49,50], which they term as HIT self-efficacy. In their model, HIT self-efficacy is conceptualized to affect perceived usefulness and ease of use, which in turn shape attitudes and the behavioral intention to use HIT. Our findings are in line with that causal pathway, although

eHealth literacy is a much broader concept that encompasses many other skills beyond technological literacy and includes functional, communicative, critical, and transactional eHealth literacy skills [49]. This highlights the need to integrate eHealth literacy, with its multidimensional characteristics, in HIT conceptual frameworks. This is especially relevant as the evidence is mounting on eHealth literacy's role in extending the digital divide to health care [51,52] and, just as importantly, in facilitating the adoption of various eHealth apps. For example, our finding of the significance of eHealth literacy to telehealth use intention mirrors other findings where eHealth literacy has been found to have a significant association with the use of or the intention to use other consumer eHealth platforms such as patient portals [13,14] and mHealth [15,16].

Limitations

There are several limitations to this study. First, our outcome variable, the behavioral intention to use telehealth, was assessed by the question, "How likely are you to use telehealth services if they were offered by your provider?" This may be interpreted to assume that a participant has a health care provider, a questionable premise given the high rate of uninsurance in our sample. However, it is worth noting that despite being uninsured, participants interact with health care providers in a variety of traditional and nontraditional settings. Some have providers through local safety net clinics; data collected at the same event in 2015 from participants with an almost identical sociodemographic profile showed that 54% of the uninsured had received health care services at the local county health clinic [53]. Area residents also seek health care services across the border in Mexico; a population-based survey of 1405 Texas border county residents revealed that respondents, especially the uninsured, regularly sought health care services in Mexico, with 38% of respondents reporting a doctor's visit within the past 12 months [54]. Furthermore, many uninsured community members interact with health care providers for their children's health care services; while 43% of Hidalgo County's population between the ages of 18 and 64 years did not have health care coverage in 2018, the corresponding rate for those under age 19 years was 13.6% [32], reflecting the higher insurance rates among children through the Children's Health Insurance Program (CHIP) or children's Medicaid.

Second, similar to many health information technology acceptance and adoption studies, data were collected at a single time point, with no experimental manipulation or random assignment, resulting in the inability to establish causality.

Third, to our knowledge, there are no previous studies that evaluate the causal relationship between eHealth literacy, telehealth attitudes, and intention to use telehealth services. Therefore, our mediation analysis was not based on a theoretically defined causal chain of variables. However, some models depicting the causal pathway between end-user characteristics and the intention to use HIT include variables that can serve as proxies for eHealth literacy such as HIT self-efficacy [10]; the depicted causal pathways in these models support our findings. It is worth noting that these studies also use cross-sectional data.

Fourth, our introduction of telehealth to participants did not take into account the cost of telehealth services, a highly relevant factor for poor, uninsured communities. Another limitation is that most participants had not heard of telehealth. Thus, attitudes and the intention to use telehealth reflected a hypothetical scenario to most participants. Nevertheless, such positive attitudes point to a window of opportunity that can be reinforced by well-designed virtual platforms that take into account the eHealth literacy skills of the target population.

Finally, the sociodemographic homogeneity of the sample did not allow for capturing the impact of ethnicity, educational attainment, socioeconomic status, and other sociodemographic factors on the intention to use telehealth; such factors have been shown to exhibit varying levels of influence on the use of eHealth [55]. However, given our focus on vulnerable minority groups, this homogeneity allows us to better control for such effects. One exception is that the overrepresentation of females in our sample, a reflection of event attendance, may limit the generalizability of our findings to males. While the higher representation of women relative to men at the event may be an indication of working schedules and women assuming responsibilities for child immunizations and physicals, it may also be reflective of the well-documented gender differences in accessing health care and adhering to preventive care guidelines [56,57]. Such differences may potentially extend to the intentions to use telehealth services.

Future Research

Multiple venues exist to expand our knowledge on telehealth use. First, it is important to strengthen the evidence on the causal pathways leading to telehealth adoption by incorporating longitudinal research designs [9]. Although our mediation analysis only showed mathematically that telehealth attitudes were a significant mediator of the effect of eHealth literacy on the intention to use telehealth services, this finding provides a strong basis for testing the causal link through future longitudinal data or experimental designs. In addition, it is important to expand the literature on the intention to use telehealth to include actual adoption and use. The recent adoption of telehealth due to the COVID-19 pandemic provides a fertile ground for research on the facilitators and barriers to telehealth use in vulnerable communities. Future research should also address the development of attitudinal measures toward telehealth that exhibit good psychometric properties. Most studies in the literature, similar to our study, use various sets of statements that reflect the particular research question and setting [28,58,59]. Furthermore, it is important to incorporate additional factors such as social influence [9] and privacy and data security issues [60] and explore their impact on telehealth acceptance and adoption. The latter is especially important since perceived privacy and security concerns have been shown to impact the intention to use telehealth services [61]. Finally, it is essential to explore telehealth's role in extending services to the uninsured and investigate delivery and reimbursement models specifically targeting this population.

Conclusions

To our knowledge, this study is among the first to focus primarily on a vulnerable Hispanic population and use mediation

analysis to explore the role of eHealth literacy on the behavioral intention to use telehealth. Our findings contribute significantly to promoting telehealth adoption in communities where access to health care services is a major challenge and an understanding of end-user characteristics is key to successful intervention design and adoption.

Telehealth promises to address many of the access challenges facing ethnic and racial minorities, rural communities, and low-income populations. Understanding the factors that influence its acceptance and, subsequently, its adoption is an essential first step to designing culturally relevant platforms that take into account key characteristics of these communities. Raising awareness about telehealth and developing interventions that target eHealth literacy skills promise more positive attitudes and more willingness to engage in telehealth use.

Acknowledgments

We would like to thank the following individuals and organizations for their help and support, without which this study would not have been possible: Hidalgo County Health and Human Services Department and Pharr-San Juan-Alamo (PSJA) Independent School District leadership for facilitating and supporting our participation at OLS at the PSJA Early College High School site; UTRGV Health and Human Performance (HLTH 3305, Summer III) students for their assistance with data collection and data entry; Ms Alejandra Hernandez, HLTH 3305 student, for securing promotional items for distribution to OLS participants; Ms Yajaira Ayala, graduate research assistant, for assistance with coordination of OLS participation logistics, data collection, and data entry; and Mr Juan Carlos Ayala and Ms Sylvia Hernandez, departmental administrative assistants, for administrative support. We would also like to thank Ms Reem Ghaddar for proofreading multiple versions of the manuscript. This study was partially funded by the College of Health Professions at UTRGV and by UTRGV's Collaborative on Population Health Innovation and Improvement (CoPHII). The College and CoPHII played no role in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

Authors' Contributions

SG led the design of the study, data collection, and manuscript preparation. KV conducted the statistical analyses and drafted the statistical methodology and results. SGA assisted with the literature review, coordination of study logistics, data collection, data cleaning, and descriptive analyses. LM assisted with the design of the survey and with manuscript preparation. All authors reviewed multiple drafts of the manuscript and edited and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Berchick E, Barnett J, Upton R. Health Insurance Coverage in the United States. 2018. URL: https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-267.pdf [accessed 2020-05-07]
2. Dall T, Reynolds R, Jones K, Chakrabati R, Iacobucci W. The complexities of physician supply and demand: projections from 2017 to 2032. 2019. URL: https://aamc-black.global.ssl.fastly.net/production/media/filer_public/31/13/3113ee5c-a038-4c16-89af-294a69826650/2019_update_-_the_complexities_of_physician_supply_and_demand_-_projections_from_2017-2032.pdf [accessed 2020-05-07]
3. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
4. Steinberg EM, Valenzuela-Araujo D, Zickafoose JS, Kieffer E, DeCamp LR. The "battle" of managing language barriers in health care. Clin Pediatr (Phila) 2016 Dec;55(14):1318-1327 [FREE Full text] [doi: 10.1177/0009922816629760] [Medline: 26896341]
5. DuBard CA, Gizlice Z. Language spoken and differences in health status, access to care, and receipt of preventive services among US Hispanics. Am J Public Health 2008 Nov;98(11):2021-2028. [doi: 10.2105/AJPH.2007.119008] [Medline: 18799780]
6. Glossary. Health Resources and Services Administration. URL: https://www.hrsa.gov/about/strategic-plan/glossary [accessed 2020-05-07]
7. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
8. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
9. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]

10. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
11. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jan;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
12. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
13. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
14. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth literacy: patient engagement in identifying strategies to encourage use of patient portals among older adults. Popul Health Manag 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
15. James DC, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mhealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Dec 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
16. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons J, Kuhlmey A, Gellert P. Associations of health app use and perceived effectiveness in people with cardiovascular diseases and diabetes: population-based survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179 [FREE Full text] [doi: 10.2196/12179] [Medline: 30920383]
17. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
18. Strudwick G. Predicting nurses' use of healthcare technology using the technology acceptance model: an integrative review. Comput Inform Nurs 2015 May;33(5):189-198. [doi: 10.1097/CIN.0000000000000142] [Medline: 25974361]
19. Glaser M, Winchell T, Plant P, Wilbright W, Kaiser M, Butler MK, et al. Provider satisfaction and patient outcomes associated with a statewide prison telemedicine program in Louisiana. Telemed J E Health 2010 May;16(4):472-479. [doi: 10.1089/tmj.2009.0169] [Medline: 20438385]
20. Martin AB, Probst JC, Shah K, Chen Z, Garr D. Differences in readiness between rural hospitals and primary care providers for telemedicine adoption and implementation: findings from a statewide telemedicine survey. J Rural Health 2012 Jan;28(1):8-15. [doi: 10.1111/j.1748-0361.2011.00369.x] [Medline: 22236310]
21. Seto E, Smith D, Jacques M, Morita PP. Opportunities and challenges of telehealth in remote communities: case study of the Yukon telehealth system. JMIR Med Inform 2019 Nov 01;7(4):e11353 [FREE Full text] [doi: 10.2196/11353] [Medline: 31682581]
22. Zanaboni P, Wootton R. Adoption of routine telemedicine in Norwegian hospitals: progress over 5 years. BMC Health Serv Res 2016 Dec 20;16:496 [FREE Full text] [doi: 10.1186/s12913-016-1743-5] [Medline: 27644324]
23. Park J, Erikson C, Han X, Iyer P. Are state telehealth policies associated with the use of telehealth services among underserved populations? Health Aff (Millwood) 2018 Dec;37(12):2060-2068. [doi: 10.1377/hlthaff.2018.05101] [Medline: 30633679]
24. Trout KE, Rampa S, Wilson FA, Stimpson JP. Legal mapping analysis of state telehealth reimbursement policies. Telemed J E Health 2017 Oct;23(10):805-814. [doi: 10.1089/tmj.2017.0016] [Medline: 28430029]
25. Cimperman M, Makovec BM, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
26. Edwards L, Thomas C, Gregory A, Yardley L, O'Cathain A, Montgomery AA, et al. Are people with chronic diseases interested in using telehealth? A cross-sectional postal survey. J Med Internet Res 2014;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]
27. Valikodath NG, Leveque TK, Wang SY, Lee PP, Newman-Casey PA, Hansen SO, et al. Patient attitudes toward telemedicine for diabetic retinopathy. Telemed J E Health 2017 Mar;23(3):205-212 [FREE Full text] [doi: 10.1089/tmj.2016.0108] [Medline: 27336678]
28. Gurupur V, Shettian K, Xu P, Hines S, Desselles M, Dhawan M, et al. Identifying the readiness of patients in implementing telemedicine in northern Louisiana for an oncology practice. Health Informatics J 2017 Sep;23(3):181-196. [doi: 10.1177/1460458216639740] [Medline: 27102886]
29. Woo K, Dowding DW. Decision-making factors associated with telehealth adoption by patients with heart failure at home: a qualitative study. Comput Inform Nurs 2020 Apr;38(4):204-214. [doi: 10.1097/CIN.0000000000000589] [Medline: 31929355]
30. 2014-18 American Community Survey 5-year estimates. URL: https://www.census.gov/programs-surveys/acs [accessed 2020-05-07]
31. Rogers E. Diffusion of Innovations. 4th edition. New York: Free Press; 1995.
32. Small Area Health Insurance Estimates. URL: https://www.census.gov/programs-surveys/sahie.html [accessed 2020-05-07]
33. Chau PY, Hu PJ. Examining a model of information technology acceptance by individual professionals: an exploratory study. J Manag Inf Syst 2014 Dec 23;18(4):191-229. [doi: 10.1080/07421222.2002.11045699]

34. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]
35. Nunnally JC. Psychometric Theory. New York: McGraw-Hill Higher; 1994.
36. Paramio PG, Almagro BJ, Hernando G, Aguaded GJI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
37. Hennessy CH, Moriarty DG, Zack MM, Scherr PA, Brackbill R. Measuring health-related quality of life for public health surveillance. Public Health Rep 1994;109(5):665-672 [FREE Full text] [Medline: 7938388]
38. Chandola T, Jenkinson C. Validating self-rated health in different ethnic groups. Ethn Health 2000 May;5(2):151-159. [doi: 10.1080/713667451] [Medline: 10984833]
39. Agresti A. Analysis of Ordinal Categorical Data. 2nd edition. Hoboken: Wiley; 2010.
40. Liu H, Zhang Y, Luo F. Mediation analysis for ordinal outcome variables. In: Milsap RE, Bolt DM, van der Ark LA, editors. Quantitative Psychology Research. Cham: Springer; 2015:429-450.
41. Alwin DF, Hauser RM. The decomposition of effects in path analysis. Am Sociol Rev 1975 Feb;40(1):37-47. [doi: 10.2307/2094445]
42. VanderWeele TJ, Zhang Y, Lim P. Brief report: mediation analysis with an ordinal outcome. Epidemiology 2016 Sep;27(5):651-655. [doi: 10.1097/EDE.0000000000000510] [Medline: 27258323]
43. Bollen KA, Stine R. Direct and indirect effects: classical and bootstrap estimates of variability. Sociol Methodol 1990;20:115-140. [doi: 10.2307/271084]
44. 2017 Behavioral Risk Factor Surveillance System Survey Data. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/brfss/ [accessed 2020-05-25]
45. Stronks K, Kulu-Glasgow I, Agyemang C. The utility of 'country of birth' for the classification of ethnic groups in health research: the Dutch experience. Ethn Health 2009 Jun;14(3):255-269. [doi: 10.1080/13557850802509206] [Medline: 19052941]
46. McKinnon D. Introduction to Statistical Mediation Analysis. New York: Lawrence Erlbaum Associates; 2008.
47. Venkatesh V, Davis FD. A theoretical extension of the Technology Acceptance Model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
48. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
49. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of ehealth literacy: concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
50. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
51. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
52. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
53. Ghaddar S, Byun J, Krishnaswami J. Health insurance literacy and awareness of the Affordable Care Act in a vulnerable Hispanic population. Patient Educ Couns 2018 Dec;101(12):2233-2240. [doi: 10.1016/j.pec.2018.08.033] [Medline: 30220598]
54. Su D, Richardson C, Wen M, Pagán JA. Cross-border utilization of health care: evidence from a population-based study in south Texas. Health Serv Res 2011 Jun;46(3):859-876 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01220.x] [Medline: 21158855]
55. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of ehealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
56. Vaidya V, Partha G, Karmakar M. Gender differences in utilization of preventive care services in the United States. J Womens Health (Larchmt) 2012 Feb;21(2):140-145. [doi: 10.1089/jwh.2011.2876] [Medline: 22081983]
57. Leone JE, Rovito MJ, Mullin EM, Mohammed SD, Lee CS. Development and testing of a conceptual model regarding men's access to health care. Am J Mens Health 2017 Mar;11(2):262-274 [FREE Full text] [doi: 10.1177/1557988316671637] [Medline: 27698256]
58. Call VRA, Erickson LD, Dailey NK, Hicken BL, Rupper R, Yorgason JB, et al. Attitudes toward telemedicine in urban, rural, and highly rural communities. Telemed J E Health 2015 Aug;21(8):644-651. [doi: 10.1089/tmj.2014.0125] [Medline: 25839334]
59. Tonn P, Reuter SC, Kuchler I, Reinke B, Hinkelmann L, Stöckigt S, et al. Development of a questionnaire to measure the attitudes of laypeople, physicians, and psychotherapists toward telemedicine in mental health. JMIR Ment Health 2017 Oct 03;4(4):e39 [FREE Full text] [doi: 10.2196/mental.6802] [Medline: 28974485]

60. Vega-Barbas M, Seoane F, Pau I. Characterization of user-centered security in telehealth services. Int J Environ Res Public Health 2019 Feb 26;16(5):693 [FREE Full text] [doi: 10.3390/ijerph16050693] [Medline: 30813642]
61. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding older people's readiness for receiving telehealth: mixed-method study. J Med Internet Res 2018 Apr 06;20(4):e123 [FREE Full text] [doi: 10.2196/jmir.8407] [Medline: 29625950]

Abbreviations

CDC: Centers for Disease Control and Prevention
CoPHII: Collaborative on Population Health Innovation and Improvement
eHEALS: eHealth Literacy Scale
eHealth: electronic health
HIT: health information technologies
MAR: missing at random
mHealth: mobile health
NDE: natural direct effect
NIE: natural indirect effect
OLS: Operation Lone Star
OR: odds ratio
PSJA: Pharr-San Juan-Alamo
UTRGV: University of Texas Rio Grande Valley

Edited by G Eysenbach; submitted 09.06.20; peer-reviewed by L Marceau; comments to author 10.07.20; revised version received 15.07.20; accepted 26.07.20; published 03.09.20.

Please cite as:
Ghaddar S, Vatcheva KP, Alvarado SG, Mykyta L
Understanding the Intention to Use Telehealth Services in Underserved Hispanic Border Communities: Cross-Sectional Study
J Med Internet Res 2020;22(9):e21012
URL: https://www.jmir.org/2020/9/e21012
doi:10.2196/21012
PMID:32880579

Original Paper

Teleconsultation Between Patients and Health Care Professionals in the Catalan Primary Care Service: Message Annotation Analysis in a Retrospective Cross-Sectional Study

Francesc López Seguí[1,2], MSc; Sandra Walsh[3], PhD; Oscar Solans[4], MD; Cristina Adroher Mas[5], MSc; Gabriela Ferraro[6], PhD; Anna García-Altés[7], PhD; Francesc García Cuyàs[5], PhD, MD; Luis Salvador Carulla[8], PhD, MD; Marta Sagarra Castro[9], MD; Josep Vidal-Alaball[10,11], PhD, MD

[1]TIC Salut Social, Ministry of Health, Barcelona, Spain

[2]Centre de Recerca en Economía de la Salut, Pompeu Fabra University, Barcelona, Spain

[3]Institut de Biologia Evolutiva (CSIC), Pompeu Fabra University, Barcelona, Spain

[4]Health Department, Catalan Ministry of Health, Barcelona, Spain

[5]Sant Joan de Déu Hospital, Barcelona, Spain

[6]CSIRO, Australian National University, Canberra, Australia

[7]Agency for Healthcare Quality and Evaluation of Catalonia, Catalan Ministry of Health, Barcelona, Spain

[8]Research School of Population Health, Australian National University, Canberra, Australia

[9]Centre d'Atenció Primària Capellades, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Barcelona, Spain

[10]Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[11]Unitat de Suport a la Recerca de la Catalunya Central, Fundacio Institut Universitari per a la recerca a l'Atencio Primaria de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

Corresponding Author:
Josep Vidal-Alaball, PhD, MD
Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central
Institut Català de la Salut
Carrer Pica d'Estats, 36
Sant Fruitós de Bages, 08272
Spain
Phone: 34 936 93 00 40
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: Over the last decade, telemedicine services have been introduced in the public health care systems of several industrialized countries. In Catalonia, the use of eConsulta, an asynchronous teleconsultation service between primary care professionals and citizens in the public health care system, has already reached 1 million cases. Before the COVID-19 pandemic, the use of eConsulta was growing at a monthly rate of 7%, and the growth has been exponential from March 15, 2020 to the present day. Despite its widespread usage, there is little qualitative evidence describing how this tool is used.

Objective: The aim of this study was to annotate a random sample of teleconsultations from eConsulta, and to evaluate the level of agreement between health care professionals with respect to the annotation.

Methods: Twenty general practitioners retrospectively annotated a random sample of 5382 cases managed by eConsulta according to three aspects: the type of interaction according to 6 author-proposed categories, whether the practitioners believed a face-to-face visit was avoided, and whether they believed the patient would have requested a face-to-face visit had eConsulta not been available. A total of 1217 cases were classified three times by three different professionals to assess the degree of consensus among them.

Results: The general practitioners considered that 79.60% (4284/5382) of the teleconsultations resulted in avoiding a face-to-face visit, and considered that 64.96% (3496/5382) of the time, the patient would have made a face-to-face visit in the absence of a service like eConsulta. The most frequent uses were for management of test results (26.77%, 1433/5354), management of repeat prescriptions (24.30%, 1301/5354), and medical enquiries (14.23%, 762/5354). The degree of agreement among professionals

as to the annotations was mixed, with the highest consensus demonstrated for the question "Has the online consultation avoided a face-to-face visit?" (3/3 professionals agreed 67.95% of the time, 827/1217), and the lowest consensus for the type of use of the teleconsultation (3/3 professionals agreed 57.60% of the time, 701/1217).

Conclusions: This study shows the ability of eConsulta to reduce the number of face-to-face visits for 55% (79% × 65%) to 79% of cases. In comparison to previous research, these results are slightly more pessimistic, although the rates are still high and in line with administrative data proxies, showing that 84% of patients using teleconsultations do not make an in-person appointment in the following 3 months. With respect to the type of consultation performed, our results are similar to the existing literature, thus providing robust support for eConsulta's usage. The mixed degree of consensus among professionals implies that results derived from artificial intelligence tools such as message classification algorithms should be interpreted in light of these shortcomings.

(J Med Internet Res 2020;22(9):e19149) doi:10.2196/19149

KEYWORDS

teleconsultation; primary care; remote consultation; message annotation; face-to-face visits

Introduction

Five Years of Teleconsultation in the Catalan Public Health System

The Catalan health care System dispenses services for 7.6 million inhabitants, providing universal coverage through a tax-based system. Administratively, it is composed of a single public payer and multiple service providers that are publicly or privately owned, with an integrated system, a major role in community and primary health care, and the use of information technologies and digital health [1].

The Catalan Health Institute (Institut Catala de la Salut) is the provider of primary health care services to three-quarters of the population of Catalonia [2]. Over the last decade, telemedicine services have been introduced in the public health care systems of several industrialized countries. In Catalonia, eConsulta is an asynchronous teleconsultation service that has been integrated into the program of computerized medical records of primary care of the public health system [3]. The service was introduced in 2015 and was gradually phased in until 2017, when it became established as a service available to all primary care teams as a complement to face-to-face care. Although initially intended as a tool solely for general practitioners, eConsulta can now also be used by pediatricians, gynecologists, midwives, and nurses. At present, 94.9% (353/372) of primary care teams have used the tool, although this type of consultation still accounts for less than 1% of the total in primary care [4]. From the patients' point of view, eConsulta is one of the services forming part of their personal health folder, and is considered to be a tool that allows the user to access aspects of their health information, make enquiries, and carry out certain administrative procedures. In this space, which is accessed via a secure authentication process, the eConsulta interface allows users to choose which health professional to direct their enquiry to and attach files, while keeping a record of previous interactions.

As of June 8, 2020, eConsulta collected 1,630,881 messages corresponding to 1,030,536 conversations between 322,897 unique users and 11,620 primary care professionals (59% doctors, 41% nurses). Before the COVID-19 pandemic, eConsulta was already growing at a rate of 24,000 conversations, 44,000 messages, 5500 new users, and 140 new professionals per month, representing a monthly growth rate of 7%, 6%, 5%, and 2%, respectively. These rates have increased exponentially from March 15, 2020 to the present day (Figure 1). The impact of the pandemic on the Catalan digital health ecosystem has already been analyzed in a previous study [5], demonstrating that the average number of messages per conversation has been stable at 1.71, suggesting that many conversations contain a single message (Figure 2). Professionals and users have used the tool an average of 17.18 and 3.47 times, respectively. However, these figures continue to grow, indicating that users are satisfied with the experience.

Figure 1. Numbers of messages and conversations (left axis), and patients and health care professional users (right axis) of eConsulta.

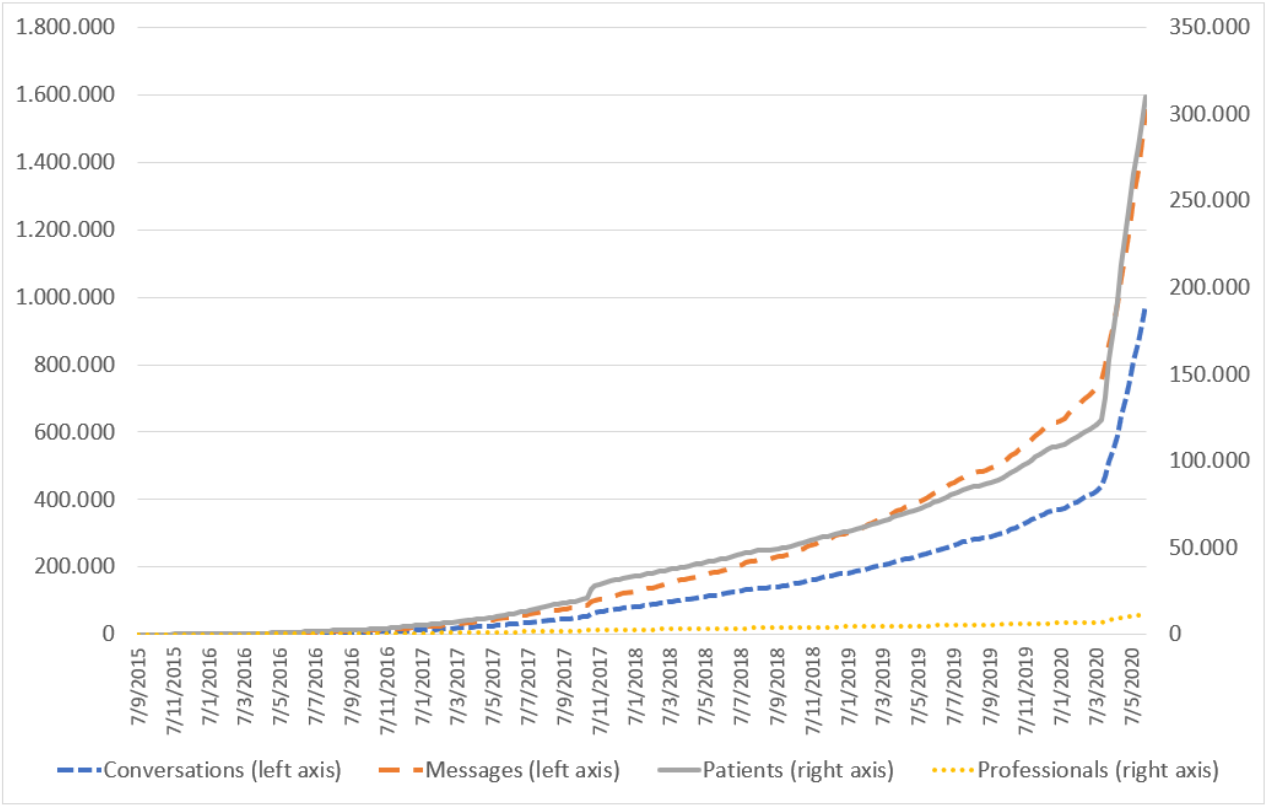
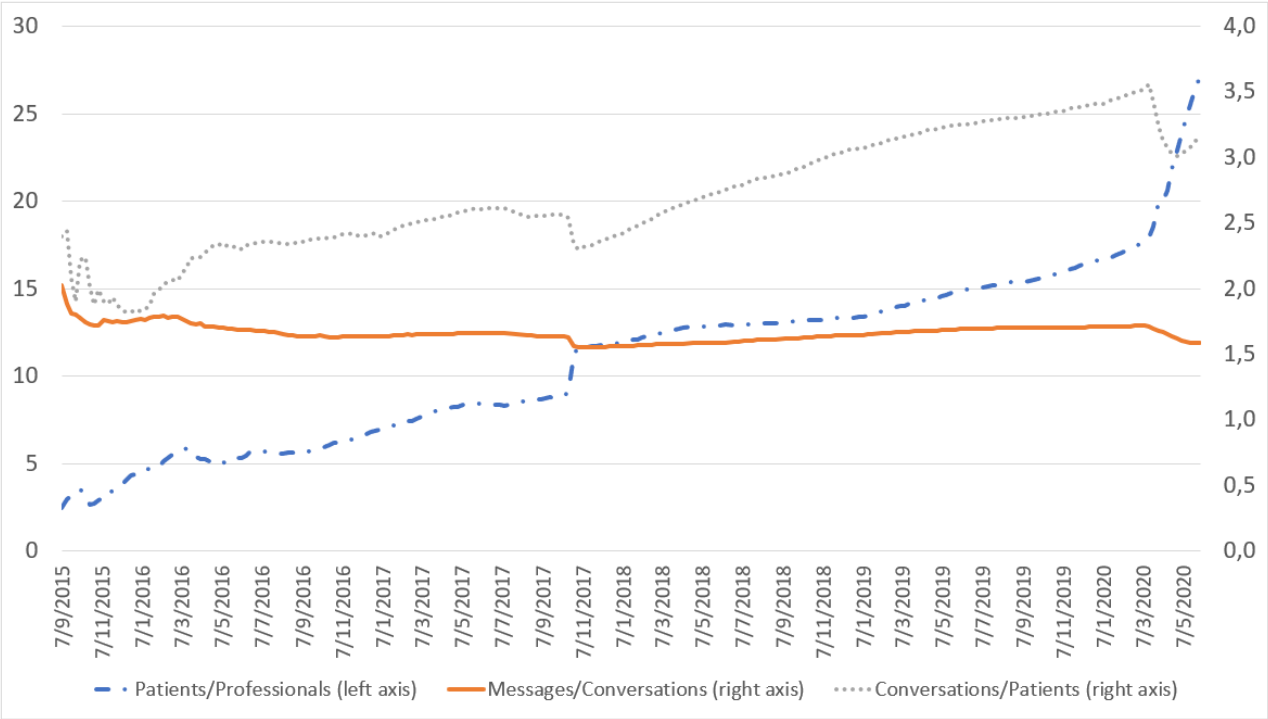


Figure 2. Patients per professional (left axis), messages per conversation and conversation per patient (right axis) in eConsulta.



In terms of the content and impact of eConsulta on face-to-face visits, a previous study [6] suggested that there was broad consensus among general practitioners that eConsulta has potential to resolve patient queries (avoiding the need for a face-to-face visit in 88% of cases), but that it induced demand (queries which otherwise would not have been made) in 28% of cases, and the most common use was for the management of test results (35%), clinical enquiries (16%), and the management of repeat prescriptions (12%). However, this study was based on a very small sample of cases involving general practitioners who use the tool in a semirural geographical area. Another study analyzed the predictive ability of machine-learning algorithms to identify types of conversations based on body text, but obtained very few conclusive results [7]. Understanding the

nature of the interaction (ie, the main topics of conversation) can help to manage such an online health care consultation tool more effectively and efficiently, which is essential considering the rapid rate at which its usage is growing.

Study Aim

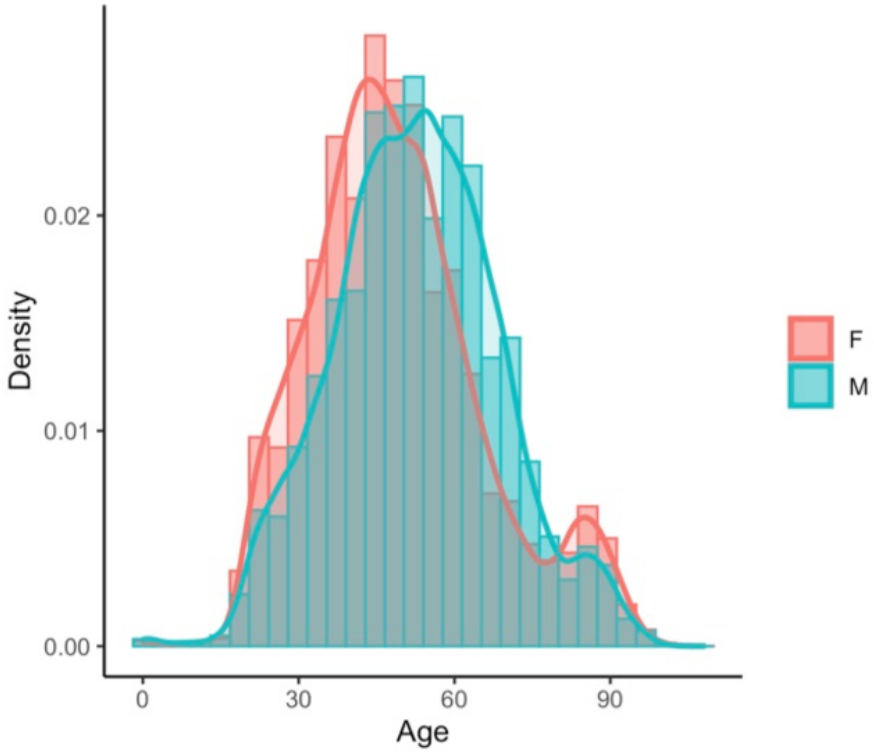
The aim of the present study was to classify a random sample of teleconsultations on eConsulta according to three types of information: the type of consultation based on 6 categories, professionals' judgement on whether or not a face-to-face visit was avoided, and whether professionals believe that the patient would have requested a face-to-face visit had eConsulta not been available. As a secondary objective, we evaluated the level of agreement among the health care professionals with respect to their annotation.

Methods

Sample

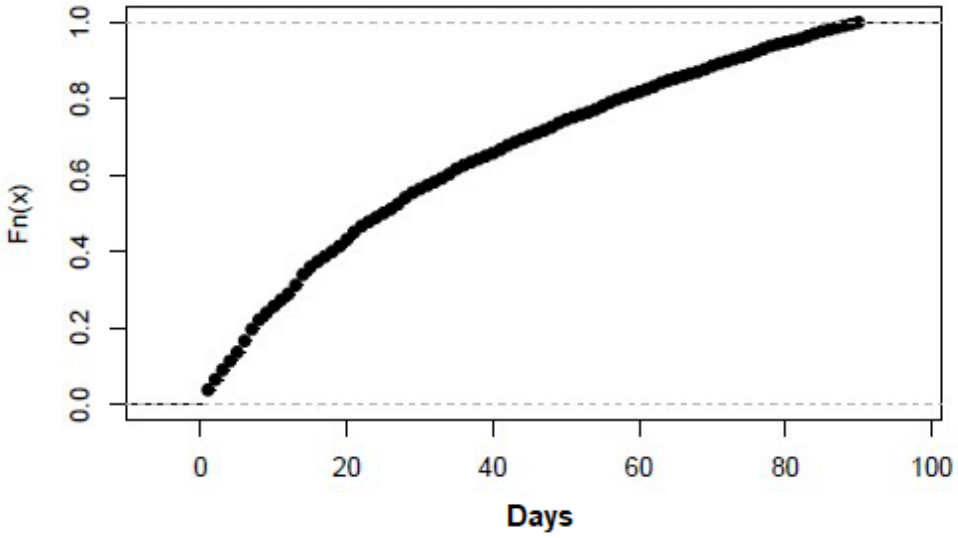
The sample analyzed included teleconsultations involving Catalan Health Institute centers initiated before September 22, 2019, comprising 403,274 messages, 236,178 conversations, and 69,111 unique users from all over Catalonia. Among the total conversations, 71.46% (168,772/236,178) were initiated by the patient, with the others initiated by the health care professional. In addition, 84.88% (187,569/220,981) of the conversations had a general practitioner as the clinical interlocutor, with the others being nurses (14.31%, 31,622/220,981) or other professionals such as pediatricians and gynecologists. The characteristics of professionals that use eConsulta were previously reported, demonstrating an age range of the general practitioners of 45-54 years, who scored higher than the 80th percentile on the quality of care index, had a high degree of accessibility, involved in teaching, and work on a health team in a high socioeconomic urban setting [3]. Messages and their respective subject lines had an average length of 231 and 18 characters, respectively. The users of the service were mostly women (57.00%, 39,393/69,111) and their average age was 50.62 years (SD 16.59) (Figure 3).

Figure 3. Age profile of patient users of the teleconsultation service, according to gender.



With regard to the use of face-to-face resources of the sample analyzed, the date of the next face-to-face visit and the total number of visits registered to the primary care team during the current year are available. Since the teleconsultation, 72.21% (150,135/207,910) of the cases did not subsequently visit the primary care team for a face-to-face consultation (for any reason), whereas 15.78% (32,802/207,910) had a consultation within the following 3 months and 12.01% (24,973/207,910) did so after 90 days. Assuming that visits made after 3 months may not be associated with the reason for the initial teleconsultation, 84.22% (175,108/207,910) of the online consultations may not be related to a subsequent face-to-face visit, suggesting that it is an effective tool for avoiding face-to-face visits. These results are in line with a previous study [6] and did not depend on the initiator of the conversation (Figure 4). Moreover, among users of teleconsultations who recorded at least one face-to-face visit, use of the in-person service occurred an average of 2.3 times a year.

Figure 4. Days until face-to-face visit for eConsultations that resulted in a face-to-face visit in the following 90 days (accumulated density).



Finally, in most of the online consultations that were associated with a subsequent face-to-face visit, the visit took place within the first 20 days after the teleconsultation, indicating the degree of accessibility to treat the patient in person.

Annotation Procedure

A random sample of 5460 unique conversations on eConsulta was generated, 4200 of which were selected to be annotated only once and the remaining 1260 were selected to be annotated three times by each general practitioner to evaluate the degree of consensus among professionals. Thus, a total of 7980 (4200+1260×3) conversations were randomly distributed among 20 health care professionals linked to the eConsulta Research Group (eHealth Office, Catalan Ministry of Health). The text in the subject line and message body was analyzed once the data were anonymized. Following the methods of López et al [6], each general practitioner recorded three pieces of information for each of their interactions: the type of consultation according to 6 author-proposed categories (see Multimedia Appendix 1), which is a modified version of our previously used classification system [6]; whether they believed a face-to-face visit was avoided, which was defined as the absence of the need for a face-to-face visit following the consultation; and whether they believed the patient would have requested a face-to-face visit had eConsulta not been available. The first aspect is an indicator of the effectiveness of the intervention, whereas the latter two questions were used as an approximate measure of the possible increased demand resulting from the ease of access to a general practitioner.

When agreeing to participate in the analysis, the participating health care professionals received a guide to standardize the annotation criteria as much as possible. The annotators were unaware that some of the conversations were being annotated three times and that the consensus among them was going to be evaluated. The study was approved by the Ethical Committee for Clinical Research at the Foundation University Institute for Primary Health Care Research Jordi Gol i Gurina (registration no. P19/096-P).

Data Analysis

The raw data were processed and analyzed with Bash and R languages (using Rstudio version 1.2.1335 and R version 3.5.3) with the packages data.table, dplyr, and ggplot2.

Results

Annotation

Overall, 98.57% (5382/5460) of the unique conversations in the sample were classified. Among these, for the subset of messages that were annotated three times (22.61%, 1217/5382), the annotation with the highest consensus was selected. Table 1 shows the results of the three main annotated variables.

Table 1. Results of the annotation.

Variable	n (%)
Has the online consultation avoided a face-to-face visit? (N=5382)	
Yes	4284 (79.60)
No	740 (13.75)
Not sure	360 (6.69)
In the absence of a service like eConsulta, would the patient have had a face-to-face consultation? (N=5382)	
Yes	3496 (64.96)
No	1626 (30.21)
Not sure	260 (4.83)
Type of teleconsultation (N=5354)	
Management of test results	1433 (26.77)
Temporary disability management	299 (5.58)
Management of visits/referrals	536 (10.01)
Repeat prescriptions	1301 (24.30)
Medical enquiries	762 (14.23)
Other	1023 (19.11)

According to the health professionals' annotation, the majority of the teleconsultations would have avoided a face-to-face visit. In addition, in the absence of the service, the patient would have arranged a face-to-face consultation in most cases (Table 1). Specifically, this study shows that eConsulta has the ability to decrease the number of face-to-face visit in between 55% (79%×65%) and 79% of cases. The most frequent uses were for the management of test results, the management of repeat prescriptions, and for general medical enquiries (Table 1).

Table 2 shows the results of the response to the questions "Has the online consultation avoided a face-to-face visit?" and "In the absence of a service like eConsulta, would the patient have had a face-to-face consultation?" according to the type of consultation. The ability of the teleconsultation to avoid a face-to-face visit was relatively higher for consultations related to the management of test results and repeat prescriptions, as the most frequent types of consultation, which are also those with lower induced demand. This implies that health professionals are using the tool in the most effective circumstances.

Table 2. Relationship between the type of consultation and the two other variables recorded.

Type of teleconsultation	N	Has the online consultation avoided a face-to-face visit? n (%)			In the absence of a service like eConsulta, would the patient have had a face-to-face consultation? n (%)		
		Yes	No	Not sure	Yes	No	Not sure
Management of test results	1433	1319 (92.04)	71 (4.95)	43 (3.00)	1105 (77.11)	323 (22.54)	5 (0.35)
Temporary disability management	299	265 (88.6)	27 (9.0)	7 (2.3)	217 (72.6)	82 (27.4)	0 (0.0)
Management of visits/referrals	536	397 (74.1)	120 (22.4)	19 (3.5)	341 (63.6)	187 (34.9)	8 (1.5)
Repeat prescriptions	1301	1223 (94.00)	50 (3.84)	28 (2.15)	988 (75.94)	311 (23.90)	2 (0.15)
Medical enquiries	762	586 (76.9)	146 (19.2)	30 (3.9)	520 (68.24)	226 (29.7)	16 (2.1)
Other	834[a]	468 (56.1)	315 (37.8)	51 (6.1)	311 (37.3)	488 (58.5)	35 (4.2)

[a]Although the total N of the "Other" category was 1023, only 834 observations were classified into the subcategories for these two variables.

Consensus Among Professionals

Table 3 shows the degree of consensus among professionals for the 1217 conversations that were annotated three times (96.58% of the target of 1260), according to three levels for each of the variables. Level 1 indicates that all professionals agree on the annotation, level 2 indicates that two of the three professionals agree on the annotation, and level 3 indicates that all three professionals responded differently to the same annotation.

Table 3. Consensus among three annotators for annotated conversations (N=1217).

Annotated variables	Level 1 (3/3 agreement), n (%)	Level 2 (2/3 agreement), n (%)	Level 3 (no agreement), n (%)
Has the online consultation avoided a face-to-face visit? (Yes/No/Not sure)	827 (67.95)	356 (29.25)	34 (2.79)
In the absence of a service like eConsulta, would the patient have had a face-to-face consultation? (Yes/No/Not sure)	477 (39.19)	713 (58.59)	27 (2.21)
Type of teleconsultation (Categories 1-6[a]/Not sure)	701 (57.60)	438 (35.99)	78 (6.41)

[a]See Multimedia Appendix 1 for category numbers and descriptions.

Notably, the lowest level of agreement (level 3) was negligible for all three variables. Responses to the question on whether eConsulta avoided a face-to-face visit generated the highest consensus, whereas the responses to whether eConsulta resulted in an induced demand generated the least consensus. The latter can be attributed to confusion as to the annotation process. Finally, it is necessary to keep in mind that the variable "type of teleconsultation" has more possible answers than offered; thus, absolute consensus among annotators (3/3) is less likely. Nevertheless, in 93.6% (1139/11217) of cases, at least 2/3 of the annotators coincided at the time of classifying the query.

Discussion

Principal Findings

This study focused on the use of teleconsultation in primary care in the Catalan health system by recruiting 20 health care professionals to annotate a total of 5382 cases. In line with the results of an earlier study [6], we found that eConsulta is a tool that primary care professionals mainly use to inform patients of their test results and for aspects related to the medication plan (eg, repeat prescription, expiration), for which it shows a high rate of resolution (avoiding a face-to-face visit). "Crossvalidation" among the professionals' annotations showed a high degree of consensus.

A major use of eConsulta appears to be for administrative or logistical reasons. This suggests that if the type of task to be performed by each professional was reformulated, some queries could be directed to nursing or even administrative staff. Given the growing pressures on primary care in Catalonia and the difficulty in recruiting general practitioners and pediatricians in some areas of the country, eConsulta could serve as a tool to empower other professional profiles in the management of certain cases. Although the results showed that health care professionals are able to resolve such cases efficiently through the use of this tool, freeing them from tasks that have little clinical value could increase their satisfaction and boost the efficiency of the primary health care system. Despite the relatively high rate (76%), medical enquiry–based consultations showed reduced potential of online resolution compared to other type of consultations, suggesting that policymakers should be cautious about the use of this kind of telemedicine service for this specific type of consultation. The types of consultations that are more likely to have increased demand should also be taken into account so as to control the amount of queries that might be received in the teleconsultation service.

From the user's point of view, the high virtual resolution of the queries also suggests a possible increase in their satisfaction [8,9]. The present results show that certain queries that are made telematically would not have been made in the absence of this type of service, thus improving people's access to the health care system and empowering them in the management of their own health and care.

For future developments aimed at increasing the capacity and optimizing the use of eConsulta, the reasons for the interactions should be systematically parameterized, which can facilitate the use of the system by the public and the development of apps based on trustable, transparent, and interpretable artificial intelligence. Simultaneously, each case should be directed to the most appropriate professional to obtain a response, leading to better case management and more efficient use of resources from the first interaction, while generating good habits in the relationship between the patient and the health care professional.

To conclude, it should be mentioned that if virtual consultations are to be used more frequently as a tool for clinical communication between health professionals and patients for health care reasons, other synchronous communication tools such as video conferencing ought to be developed [10]. These additional tools could be useful for patients with mild acute pathologies as well as for patients with chronic diseases with a high need for follow up, which can help to avoid successive face-to-face visits and improve the continuity of care [11].

Limitations

The high monthly growth of new users suggests that there are many first conversations, implying that there could be a large number of experimental queries. This would explain the fact that some of the conversations were annotated in the "Other" category. However, the fact that information systems do not record the reason for the teleconsultation makes it very difficult to analyze their association with the use of face-to-face resources.

Finally, subsequent studies should further confirm the degree of consensus with respect to the question "In the absence of a service like eConsulta, would the patient have had a face-to-face consultation?"

Acknowledgments

We are grateful for the collaboration of the health care professionals who participated in the annotation. This study was conducted with the support of the Secretary of Universities and Research of the Department of Business and Knowledge of the Government of Catalonia.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Categories of patient-physician consultation types.
[DOCX File , 15 KB - jmir_v22i9e19149_app1.docx]

References

1. Baltaxe E, Cano I, Herranz C, Barberan-Garcia A, Hernandez C, Alonso A, et al. Evaluation of integrated care services in Catalonia: population-based and service-based real-life deployment protocols. BMC Health Serv Res 2019 Jun 11;19(1):370-319 [FREE Full text] [doi: 10.1186/s12913-019-4174-2] [Medline: 31185997]
2. Catalan Health Institute Activities Report. 2018. URL: http://ics.gencat.cat/ca/lics/memories-dactivitat/memoria-de-lics-2018/ [accessed 2020-08-13]
3. Fernández OS, López Seguí F, Vidal-Alaball J, Bonet Simo JM, Vian OH, Cabo PR, et al. Primary Care Doctor Characteristics That Determine the Use of Teleconsultations in the Catalan Public Health System: Retrospective Descriptive Cross-Sectional Study. JMIR Med Inform 2020 Jan 31;8(1):e16484 [FREE Full text] [doi: 10.2196/16484] [Medline: 32012061]
4. Vidal-Alaball J, López Seguí F. Ha llegado para quedarse: beneficios e inconvenientes de la eConsulta. Atención Primaria Práctica 2020 Jan;2(1-2):100043. [doi: 10.1016/j.appr.2020.100043]
5. Vidal-Alaball J, Acosta-Roja R, Pastor Hernández N, Sanchez Luque U, Morrison D, Narejos Pérez S, et al. Telemedicine in the face of the COVID-19 pandemic. Aten Primaria 2020 Jun;52(6):418-422 [FREE Full text] [doi: 10.1016/j.aprim.2020.04.003] [Medline: 32402477]
6. López Seguí F, Vidal-Alaball J, Sagarra Castro M, García-Altés A, García Cuyàs F. General Practitioners' Perceptions of Whether Teleconsultations Reduce the Number of Face-to-face Visits in the Catalan Public Primary Care System: Retrospective Cross-Sectional Study. J Med Internet Res 2020 Mar 16;22(3):e14478 [FREE Full text] [doi: 10.2196/14478] [Medline: 32175914]
7. López Seguí F, Ander Egg Aguilar R, de Maeztu G, García-Altés A, García Cuyàs F, Walsh S, et al. Teleconsultations between Patients and Healthcare Professionals in Primary Care in Catalonia: The Evaluation of Text Classification Algorithms Using Supervised Machine Learning. Int J Environ Res Public Health 2020 Feb 09;17(3):1093 [FREE Full text] [doi: 10.3390/ijerph17031093] [Medline: 32050435]
8. von Wangenheim A, de Souza Nobre LF, Tognoli H, Nassar SM, Ho K. User satisfaction with asynchronous telemedicine: a study of users of Santa Catarina's system of telemedicine and telehealth. Telemed J E Health 2012 Jun;18(5):339-346. [doi: 10.1089/tmj.2011.0197] [Medline: 22703378]
9. Eedy DJ, Wootton R. Teledermatology: a review. Br J Dermatol 2001 Apr;144(4):696-707. [doi: 10.1046/j.1365-2133.2001.04124.x] [Medline: 11298526]
10. Mallow JA, Petitte T, Narsavage G, Barnes E, Theeke E, Mallow BK, et al. The Use of Video Conferencing for Persons with Chronic Conditions: A Systematic Review. Ehealth Telecommun Syst Netw 2016 Jun;5(2):39-56 [FREE Full text] [doi: 10.4236/etsn.2016.52005] [Medline: 29104830]
11. Dixon RF, Rao L. Asynchronous virtual visits for the follow-up of chronic conditions. Telemed J E Health 2014 Jul;20(7):669-672. [doi: 10.1089/tmj.2013.0211] [Medline: 24784174]

Edited by G Eysenbach; submitted 06.04.20; peer-reviewed by E Da Silva, MA Bahrami; comments to author 12.06.20; revised version received 18.06.20; accepted 22.06.20; published 17.09.20.

Please cite as:
López Seguí F, Walsh S, Solans O, Adroher Mas C, Ferraro G, García-Altés A, García Cuyàs F, Salvador Carulla L, Sagarra Castro M, Vidal-Alaball J
Teleconsultation Between Patients and Health Care Professionals in the Catalan Primary Care Service: Message Annotation Analysis in a Retrospective Cross-Sectional Study
J Med Internet Res 2020;22(9):e19149
URL: https://www.jmir.org/2020/9/e19149
doi:10.2196/19149
PMID:32687477

Original Paper

Patients' and Nurses' Experiences and Perceptions of Remote Monitoring of Implantable Cardiac Defibrillators in Heart Failure: Cross-Sectional, Descriptive, Mixed Methods Study

Maria Liljeroos[1,2*], PhD; Ingela Thylén[1,3*], PhD; Anna Strömberg[1,3*], Prof Dr

[1]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[2]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

[3]Department of Cardiology, Linköping University, Linköping, Sweden

[*]all authors contributed equally

Corresponding Author:
Maria Liljeroos, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Campus US
Linköping, 581 83
Sweden
Phone: 46 703728329
Fax: 46 16104878
Email: maria.liljeroos@liu.se

Abstract

Background: The new generation of implantable cardioverter-defibrillators (ICDs) supports wireless technology, which enables remote patient monitoring (RPM) of the device. In Sweden, it is mainly registered nurses with advanced education and training in ICD devices who handle the arrhythmias and technical issues of the remote transmissions. Previous studies have largely focused on the perceptions of physicians, and it has not been explored how the patients' and nurses' experiences of RPM correspond to each other.

Objective: Our objective is to describe, explore, and compare the experiences and perceptions, concerning RPM of ICD, of patients with heart failure (HF) and nurses performing ICD follow-up.

Methods: This study has a cross-sectional, descriptive, mixed methods design. All patients with HF and an ICD with RPM from one region in Sweden, who had transitioned from office-based visits to implementing RPM, and ICD nurses from all ICD clinics in Sweden were invited to complete a purpose-designed, 8-item questionnaire to assess experiences of RPM. The questionnaire started with a neutral question: "What are your experiences of RPM in general?" This was followed by one positive subscale with three questions (score range 3-12), with higher scores reflecting more positive experiences, and one negative subscale with three questions (score range 3-12), with lower scores reflecting more negative experiences. One open-ended question was analyzed with qualitative content analysis.

Results: The sample consisted of 175 patients (response rate 98.9%) and 30 ICD nurses (response rate 60%). The majority of patients (154/175, 88.0%) and nurses (23/30, 77%) experienced RPM as very good; however, the nurses noted more downsides than did the patients. The mean scores of the negative experiences subscale were 11.5 (SD 1.1) for the patients and 10.7 (SD 0.9) for the nurses (P=.08). The mean scores of the positive experiences subscale were 11.1 (SD 1.6) for the patients and 8.5 (SD 1.9) for the nurses (P=.04). A total of 11 out of 175 patients (6.3%) were worried or anxious about what the RPM entailed, while 15 out of 30 nurses (50%) felt distressed by the responsibility that accompanied their work with RPM (P=.04). Patients found that RPM increased their own (173/175, 98.9%) and their relatives' (169/175, 96.6%) security, and all nurses (30/30, 100%) answered that they found RPM to be necessary from a safety perspective. Most patients found it to be an advantage with fewer office-based visits. Nurses found it difficult to handle different systems with different platforms, especially for smaller clinics with few patients. Another difficulty was to set the correct number of alarms for the individual patient. This caused a high number of transmissions and a risk to miss important information.

Conclusions: Both patients and nurses found that RPM increased assurance, reliance, and safety. Few patients were anxious about what the RPM entailed, while about half of the nurses felt distressed by the responsibility that accompanied their work with

RPM. To increase nurses' sense of security, it seems important to adjust organizational routines and reimbursement systems and to balance the workload.

(J Med Internet Res 2020;22(9):e19550) doi:10.2196/19550

KEYWORDS

heart failure; remote patient monitoring; implantable cardioverter-defibrillator

Introduction

According to clinical guidelines, an implantable cardioverter-defibrillator (ICD), with or without cardiac resynchronization therapy (CRT), is recommended in selected patients with heart failure (HF) to reduce the risk of sudden cardiac death. CRT uses a special type of pacemaker called a biventricular pacemaker to treat HF. The CRT pacemaker is placed under the skin of the chest and connects to three leads in the heart: one in the right atrium, one in the right ventricle, and one in the left ventricle. The pacemaker sends electrical pulses to make the ventricles pump at the same time. The pacemaker can also speed up the heartbeats if the heart is beating too slowly. When the lead in the right ventricle is combined with a shock lead, the device is called an ICD, which continuously monitors the heart rhythm and, in case of abnormal ventricular rhythm, can shock the heart back to a normal rhythm. The indication could be either secondary (ie, in patients recovering from ventricular arrhythmia causing hemodynamic instability) or primary (ie, in patients who are at high risk for life-threatening ventricular arrhythmias in the future: primary prevention) [1]. The number of patients treated with a primary ICD is increasing and account for about 80% of all implants worldwide today [2].

Traditionally, ICD follow-ups have required in-person assessment, with quarterly to yearly office-based visits and with an increased rate when the device approaches its end of service, in case of advisories, or when the patient's health deteriorates. During these follow-up visits, a cardiologist, a registered nurse specialized in the care of ICD patients, and/or a technician noninvasively uses a programmer to gather data from the device. However, in the last decade the new generation of ICD devices that supports wireless technology has enabled remote monitoring of the device [3-6]. Consequently, in order to improve clinical practice of ICD follow-ups and to provide earlier detection of clinical problems, a wider use of remote patient monitoring (RPM) has been recommended by scientific associations in both Europe [5] and the United States [7], and has become the new standard of care for patient follow-up [8-10]. When a patient is remotely followed, the office-based visits can be limited to the initial postimplantation period and annual follow-ups, unless alerts or the patient's health requires an urgent in-person check [6]. In practice, the data transceiver is typically situated in the patient's bedroom to automatically receive data from the implant, using wireless Bluetooth technology, during the night, usually between 1 AM and 5 AM. In case of an alert (ie, arrhythmia or technical problem), the data are automatically transmitted to the manufacturer's central repository using a mobile network link or a landline where the personnel at the device clinic have access to the data on a secure, dedicated website.

Remote monitoring of implanted devices has been found to have a number of advantages for both patients and health care personnel, compared to office-based visits, when the connectivity and transmissions work properly [11]. For example, RPM of the ICD reduces health care costs, time consumption, as well as transportation costs for patients [12,13]. It also provides a feeling of security for the patients knowing that their device is constantly being monitored and if there are some functional problems, those will be detected without delay [14]. Previous studies also indicate that ICD patients are generally satisfied with RPM [15,16], and comprehension of the usefulness of RPM has been positively associated with the acceptance of being monitored remotely [15,16]. Physicians also regard RPM as a clinically useful technology that affords significant benefits for patients and health care organizations, with the most significant benefit being the early detection of atrial arrhythmias, lead failure, and worsening of HF in CRT patients [17]. The EVOLVO (Evolution of Management Strategies of Heart Failure Patients With Implantable Defibrillators) study demonstrated a reduction of 35% in urgent admissions and 21% in urgent office-based visits for worsening HF in the RPM arm, even though this study was not powered to demonstrate clinical benefit [18]. Further, the REMOTE-CIED (Remote patient management for Cardiac Implantable Electronic Devices) trial showed that patient-reported health status and ICD acceptance did not differ between patients on RPM and patients receiving in-clinic check-ups alone in the first 2 years after ICD implantation [19].

However, disadvantages have also been described. For example, some patients (5%-22%) do not feel comfortable with RPM and report a strong preference for regular office-based visits to feel secure [15,16]. Furthermore, RPM has been described as system centered, providing patients with little or no data from their device [20]. Patients missed receiving feedback via their monitor, 84% wished for a more detailed response, and 21% wished for faster feedback after scheduled transmissions [16,21]. Yet another disadvantage is the lack of data integration with electronic medical record platforms and other systems in the current RPM systems, leading to the device diagnostics being underutilized by the health care personnel [20]. Despite a reduction of office-based follow-ups, RPM is perceived as increasing workload for the staff involved [17]. Furthermore, the American PREDICT-RM (Patient-Related Determinants of ICD Remote Monitoring) registry has also found that about every fourth patient chose not to activate their RPM system at home, leading to extra time consumption for the staff in identifying and contacting those patients. Younger age, racial and ethnic minorities, having no health insurance, shorter travel

distance to the hospital, and the presence of comorbidities or procedure-related adverse events have been found to be associated with a lower likelihood of RPM activation [22]. Finally, the EDUCAT study showed that a high overall understanding of RPM was related to patient age, where younger patients had a better comprehension of home monitoring but the number of data transmissions were unrelated to comprehension, which confirms the importance of training in patients' acceptance of the system [23].

In Europe with its different health care and reimbursement systems, the heterogeneity of follow-up appointments is quite substantial when it comes to the accumulative time spent and the number of health care personnel involved in the visits, as described in a survey by the European Heart Rhythm Association in 2011. This *snapshot* survey included 26 *real-word practice* centers from seven European countries—the United Kingdom, Spain, Switzerland, France, Germany, Italy, and Greece—and showed that the mean duration of a visit was 27 minutes (25^{th} and 75^{th} percentiles were 15 and 35) for scheduled office-based visits and that most visits involved a cardiologist and a nurse simultaneously (59%). Nurses alone did 4% of the face-to-face visits [24]. However, in Sweden, it is mainly registered nurses with advanced education and training in ICD devices who handle the arrhythmias and technical issues of the ICDs and who consult a cardiologist when needed. There is no uniform education and training on ICD monitoring. Some of the manufacturers offer product-specific education online repeatedly every year; others offer short written instructions. It is usually up to the different clinics to educate and train new nurses by introducing them to the task assignment. Previous studies have largely focused on the perceptions of physicians, and it has not been explored how the patients' and nurses' experiences and perceptions of RPM compare to each other. Since nurses worldwide have become more involved in the care for ICD patients with RPM, and as the number of patients with RPM of their devices increases, it is important to explore their experiences.

Nurses in Sweden usually work with one platform per manufacturer, and each nurse has to handle three to four different platforms, each covering about five to ten different ICD-CRT-D (cardiac resynchronization therapy defibrillator) models or systems, depending on the procurement in the respective region.

Therefore, the aim of the study was to describe, explore, and compare experiences and perceptions of cardiac nurses performing ICD follow-up, concerning RPM of ICD in patients with HF.

Methods

Design and Setting

This study had a cross-sectional, descriptive, mixed methods design. All patients with HF and an ICD with RPM from one region in Sweden were invited to participate in the study. In addition, all ICD nurses working at all the ICD clinics in Sweden were invited to complete a survey.

The region of Sörmland has a land area of 6103 km^2 and about 300,000 inhabitants. There are three hospitals with a total of 270 hospital beds on medical wards, but only one hospital has an in-hospital device clinic. The travel distance for patients with an implanted device could be up to 100 km.

Over 1 year, from October 2015 until October 2016, RPM was introduced and started for all ICD recipients during their scheduled visits at the device clinic, and all new ICD recipients received RPM directly after the implant. Due to various reasons, 4 patients declined RPM, so at the end of 2018 there were 310 ICD recipients on RPM in Sörmland County.

Before implementing RPM, ICD recipients had scheduled office-based visits every 3 to 6 months. After the implementation of RPM, patients routinely visit the clinic once a year and, in between, a scheduled transmission is performed. The ICD recipients can call the ICD nurse on weekdays when needed.

Sample and Procedure

All adult ICD recipients having a verified HF diagnosis according to the European Society of Cardiology guidelines [1] (N=177) were invited to participate in the study during their yearly follow-up visit at the in-hospital device clinic, from January to December 2018. Exclusion criteria were being less than 18 years old and not being able to understand Swedish.

Patients interested in participating provided written informed consent. They were thereafter given additional written information and questionnaire packets to complete at home and return by mail in prepaid envelopes.

The ICD nurses were identified by contacting the National Swedish Pacemaker and ICD Registry, which provided names and email addresses for all ICD nurses (N=50) working at an ICD clinic in Sweden at the time. An electronic survey was sent out and two reminders were posted.

Measures and Instruments

Overview

Demographic data and data on comorbidities were self-reported by the patients. These data included gender, age, living arrangements, place of birth, and educational level. The patients also self-reported validated measures concerning the level of ICD-related concerns [25], symptoms of depression and anxiety [26-28], and perceived control [29,30]. Demographic data from the nurses were also self-reported and included gender, age, work experience, number of patients at the clinic, number of patients on RPM, and time spent per week working with RPM.

Experiences of Remote Monitoring

Two 8-item questionnaires to assess patients' and nurses' experiences of remote monitoring were developed by the study team (see Table 1). Patients rated the items (eg, "RPM is technically difficult for me") on a 4-point Likert scale from 1 (Totally agree) to 4 (Do not agree).

The nurses answered similar questions on a 4-point Likert scale, but their questions also concerned how comfortable they were handling RPM and the responsibility this brings (eg, "The responsibility that accompanies my work with RPM worries

me" and "The responsibility that accompanies my work with RPM increases my security").

For both patients and nurses, there was one final open-ended question: "I experience these advantages or disadvantages of RPM"; it was possible to write as many comments as one liked when answering the question.

Table 1. Questionnaire items and responses to assess patients' and nurses' experiences of remote patient monitoring (RPM).

Questions or statements	Responses
Patient questionnaire	
1. What are your experiences of RPM in general?	1 (Bad), 2 (Fairly bad), 3 (Fairly good), or 4 (Good)
2. RPM is unnecessary.	1 (Totally agree), 2 (Mostly agree), 3 (Partly agree), or 4 (Do not agree)
3. RPM is technically difficult for me.	1 (Totally agree), 2 (Mostly agree), 3 (Partly agree), or 4 (Do not agree)
4. RPM makes me worried.	1 (Totally agree), 2 (Mostly agree), 3 (Partly agree), or 4 (Do not agree)
5. RPM increases my security.	4 (Totally agree), 3 (Mostly agree), 2 (Partly agree), 1 (Do not agree)
6. RPM makes me feel safe.	4 (Totally agree), 3 (Mostly agree), 2 (Partly agree), 1 (Do not agree)
7. RPM provides increased security and safety for my relatives.	4 (Totally agree), 3 (Mostly agree), 2 (Partly agree), 1 (Do not agree)
8. I experience these advantages or disadvantages of RPM.	Open-ended question
Nurse questionnaire	
1. What are your experiences of RPM in general?	1 (Bad), 2 (Fairly bad), 3 (Fairly good), or 4 (Good)
2. RPM is unnecessary.	1 (Totally agree), 2 (Mostly agree), 3 (Partly agree), or 4 (Do not agree)
3. RPM is technically difficult for my patients.	1 (Totally agree), 2 (Mostly agree), 3 (Partly agree), or 4 (Do not agree)
4. The responsibility that accompanies my work with RPM worries me.	1 (Totally agree), 2 (Mostly agree), 3 (Partly agree), or 4 (Do not agree)
5. The responsibility that accompanies my work with RPM increases my security.	4 (Totally agree), 3 (Mostly agree), 2 (Partly agree), 1 (Do not agree)
6. RPM is necessary from a patient safety perspective.	4 (Totally agree), 3 (Mostly agree), 2 (Partly agree), 1 (Do not agree)
7. RPM gives the patient increased security and safety.	4 (Totally agree), 3 (Mostly agree), 2 (Partly agree), 1 (Do not agree)
8. I experience these advantages or disadvantages of RPM.	Open-ended question

Data Analysis

Descriptive statistics were used to present sample characteristics for all study variables. Demographic and clinical variables were compared using chi-square statistics or the Student *t* test.

Regarding experiences of remote monitoring, for both patients and nurses, the items were divided into two subscales based on an exploratory factor analysis: one with negative experiences (items 2-4) and one with positive experiences (items 5-7). The items in each subscale were summed to a total score. For the negative experiences (score range 3-12), lower scores reflected more negative experiences with RPM. For the positive experiences (score range 3-12), higher scores reflected higher levels of more positive experiences with RPM. Item 1 was considered neutral and was calculated separately for each group.

The level of statistical significance was set to $P<.05$. The statistical analyses were conducted using SPSS Statistics for Windows, version 25.0 (IBM Corp).

The final open-ended question was analyzed with manifest qualitative content analysis [31]. The three authors who developed the survey and analyzed the qualitative data are researchers with years of experience as clinical nurses, two working in HF care (ML and AS) and one in ICD care (IT).

Ethical Considerations

The study was approved by the Regional Ethics Committees for Human Research in Linköping, Sweden (ref 2017/441-31), and was conducted in accordance with the World Medical Association Declaration of Helsinki and the Code of Ethics for Nurses [32,33].

Results

Patients' and ICD Nurses' Characteristics

The sample consisted of 175 patients (response rate 98.9%) and 30 nurses (response rate 60%). The patients' mean age was 69.9 years (SD 9.7); 138 out of 175 patients (78.9%) were males, 144 were retired (82.3%), and 109 were married (62.3%). A total of 128 out of 175 participants (86.0%) reported no symptoms of anxiety or depression, and 129 out of 175 patients (73.7%) scored a low level of ICD-related concerns (see Table 2).

The nurses' mean age was 52.7 years (SD 8.7) and 26 out of 30 (86%) were females. They had been working as nurses for a mean of 26 years (range 5-47) and as ICD nurses for 7.6 years (range 1-14). They spent a mean of 7.5 hours/week (range 1-30) working with RPM (see Table 2).

Table 2. Characteristics of participating patients and nurses.

Characteristics	Value, n (%) or mean (SD)
Gender (male), n (%)	
Patients (N=175)	138 (78.9)
Nurses (N=30)	4 (13)
Age (years), mean (SD)	
Patients (N=175)	69.9 (9.7)
Nurses (N=30)	52.7 (8.7)
Patient origin of birth (N=175), n (%)	
Sweden	153 (87.4)
Other Nordic country	17 (9.7)
Other part of Europe	5 (2.9)
Patient living conditions (N=175), n (%)	
Single	57 (32.6)
Married	109 (62.3)
Living with spouse and child	7 (4.0)
Living with relative	2 (1.2)
Patient education (N=175), n (%)	
Elementary school	90 (51.4)
Education after elementary school	63 (36.0)
University or higher education	22 (12.6)
Patient main occupation (N=175), n (%)	
Employed	18 (10.3)
Self-employed	9 (5.1)
Retired	144 (82.3)
Sick leave	4 (2.3)
Patient HADS[a] score, anxiety (N=175)	
Total score, mean (SD)	3.6 (3.7)
No symptoms (0-7), n (%)	128 (86.0)
Mild symptoms (8-10), n (%)	10 (7.0)
Moderate-to-severe symptoms (>10), n (%)	10 (7.0)
Patient HADS score, depression (N=175)	
Total score, mean (SD)	3.6 (3.9)
No symptoms (0-7), n (%)	124 (84.4)
Mild symptoms (8-10), n (%)	19 (12.4)
Moderate-to-severe symptoms (>10), n (%)	4 (2.8)
Patient CAS[b] total score, mean (SD)	19.3 (5.2)
Patient 8-item ICDC[c] score (N=175)	
Total score, mean (SD)	6.3 (6.6)
Low level of ICD[d] concerns (0-10), n (%)	129 (73.7)
High level of ICD concerns (11-28), n (%)	46 (26.3)
Nurses only (N=30), mean (SD)	
Years since nurse exam	26.0 (9.4)

Characteristics	Value, n (%) or mean (SD)
Years working at device clinic	7.6 (3.7)
ICD patients at the clinic	345 (205)
ICD patients on remote monitoring	233 (290)
Hours/week working with remote monitoring	7.5 (4.4)

[a]HADS: Hospital Anxiety and Depression Scale.

[b]CAS: Control Attitude Scale.

[c]ICDC: Patient Implantable Cardioverter-Defibrillator Concerns Questionnaire.

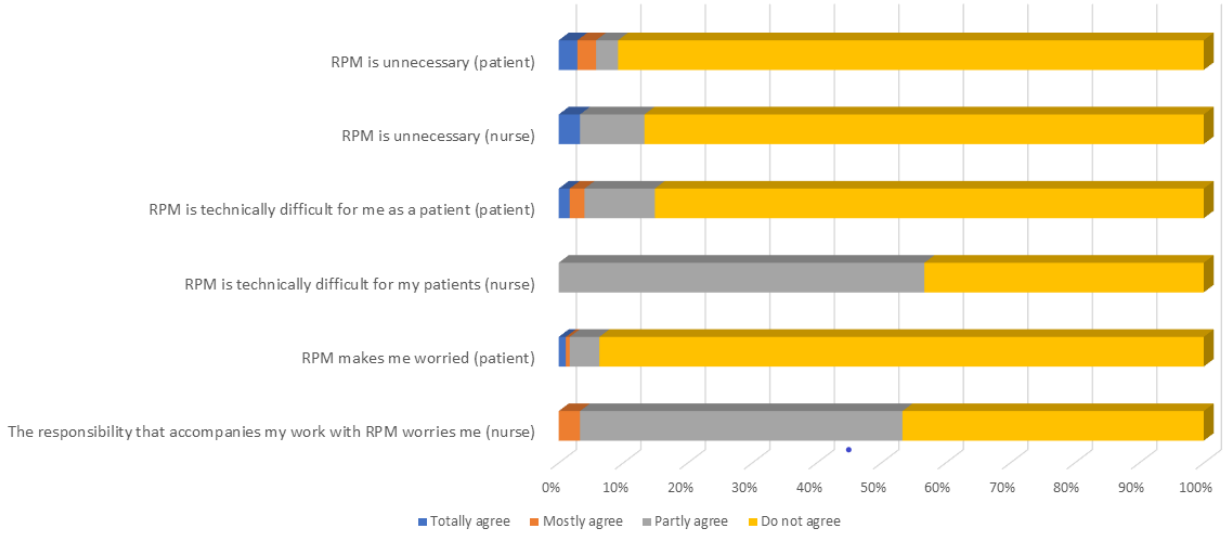
[d]ICD: implantable cardioverter-defibrillator.

Experiences of Remote Monitoring

The majority of patients (154/175, 88.0%) as well as nurses (23/30, 77%) experienced RPM in general as very good. The mean scores of the negative experiences subscale were 11.5 (SD 1.1) for the patients and 10.7 (SD 0.9) for the nurses (P=.08), with a trend for the nurses being more negative than for the patients, although this was not statistically significant.

A total of 16 out of 175 patients (9.1%) and 4 out of 30 nurses (13%) found RPM unnecessary or partly unnecessary (see Figure 1). A total of 17 out of 30 nurses (57%) found the technical equipment somewhat difficult for the patients to handle; in contrast, 149 of the 175 patients (85.1%) answered that they did not experience any technical difficulties in handling RPM (P=.04). Only 11 out of 175 patients (6.3%) were worried or anxious about what the RPM entailed, while 15 out of 30 nurses (50%) felt distressed by the responsibility that accompanied their work with RPM (P=.04).

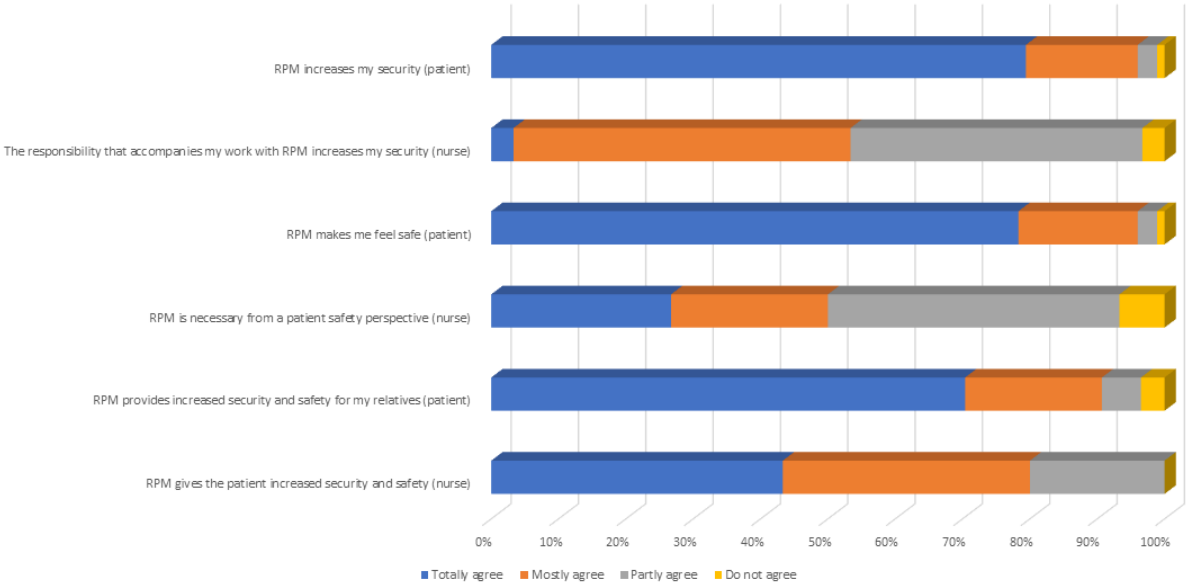
Figure 1. Negative experiences regarding remote patient monitoring (RPM) perceived by patients and nurses.



The mean scores of the positive experiences subscale were 11.1 (SD 1.6) for the patients and 8.5 (SD 1.9) for the nurses (P=.03), meaning that patients were more positive toward RPM than were the nurses.

The majority of the patients found that RPM increased their own (173/175, 98.9%) and their relatives' (169/175, 96.6%) security. All nurses (30/30, 100%) answered that RPM increased patient security and safety, and 28 out of 30 nurses (93%) found it necessary from a patient safety perspective (see Figure 2). Also, 15 out of 30 nurses (50%) answered that the responsibility that accompanied working with RPM increases their security, since they knew there would be an alert in case of malfunction or arrhythmias.

Figure 2. Positive experiences regarding remote patient monitoring (RPM) perceived by patients and nurses.



Patients' Perceptions About Being Monitored With RPM

Overview

In total, 101 out of 175 patients (57.7%) provided responses to the open-ended question regarding perceived advantages and disadvantages with RPM. The analysis resulted in 110 meaningful units, whereof 94 were described as advantages and 16 as disadvantages, which were further analyzed and categorized into two categories: *security and safety* and *organization of care*. Each category reflects both advantages and disadvantages with RPM.

Security and Safety

The most prominent advantage described was that RPM increased the sense of security and safety, not only for the patient but also for their relatives. Nevertheless, some patients highlighted the fact that atrial fibrillation was not automatically reported by one specific RPM manufacturer, which affected the feeling of security. Several patients expressed that they now could live a more "normal daily life," not having to think about the ICD when knowing that someone was watching over them. However, misconceptions were also described where the patients believed that they were continuously monitored in real time, 24/7. Likewise, it was considered as a lack of safety when traveling and not bringing the home monitor, when patients were used to be monitored, which could cause worries.

Organization of Care

When it comes to the new way of organizing care, most of the patients described that it was positive not needing to travel to hospital for follow-up as often as before. In contrast, for some patients, the fewer number of office-based visits to the ICD clinic was considered a disadvantage since they appreciated the face-to-face interaction. The patients felt that they had no one to talk to when worries or questions arose and did not want to bother the nurse with a phone call. Still, others appreciated the possibility to send data to the ICD clinic, and it was described as reassuring to have the possibility to call the ICD nurse when needed. Finally, some patients emphasized the size of the home monitor and wished it to be smaller, while others reported that the control light shined too brightly during the night.

ICD Nurses' Perceptions About Working With RPM

Overview

A total of 23 out of 30 nurses (77%) responded to the open-ended question; this resulted in 76 meaningful units, whereof 31 were described as advantages and 45 as disadvantages, which were further analyzed and categorized into three categories: *security and safety*, *organization of care*, and *managing technology*. Each category reflects both advantages and disadvantages with RPM.

Security and Safety

In accordance with the patients' perceptions, the nurses also described that RPM increased the safety and sense of security for the patients with early detection of arrhythmias, device malfunctions, low battery status, and decompensation of HF status. Early detection led to more immediate actions, and some alerts could be handled by phone instead of through an office-based visit. However, it was considered difficult to customize the correct alert limits for the individual patient, which often resulted in keeping the default alerts. This caused a high, and sometimes unnecessary, number of transmissions with a risk to miss severe arrhythmias and/or malfunction in the high flow of information.

Organization of Care

The nurses described that fewer office-based visits were an advantage for both the health care personnel and the patients.

Fewer regular visits made it possible to schedule patients with short notice when needed. Nevertheless, the nurses also found the task assignments surrounding RPM, including all documentation in the medical records, to be burdensome. Some nurses stated that handling the transmission with interpretation of data, getting in contact with the patient in case of arrhythmia or device malfunction, consulting a physician if necessary, and finally documenting the alert took the same amount of time or more compared to an office-based visit. Some hospitals had not planned ahead before implementing RPM and, therefore, had no routines on how to document transmissions and lacked action plans on how to handle alerts. In these cases, RPM caused stress for the nurses, since they did not have any time set aside to handle all transmissions; they were then worried about missing important information. It was also described that hospital managers and heads of departments did not recognize the time-consuming work nurses did when handling daily transmissions. They just noticed that there were fewer patients at the clinic and tried to give the nurses other work tasks instead, which led to frustration and dissatisfaction.

Managing Technology

The nurses also described that having to learn all the different systems with different platforms and log-ins was difficult and stressful, especially for smaller clinics with few ICD patients, and the nurses wished for a joint platform for all manufacturers. They also highlighted how time-consuming it was to trace and handle different technical problems and the time it takes trying to reach the patient by phone. Extra stress was caused by all the time it took trying to reach the patient when the monitor lost contact with the server, which was described as a common technical problem. When there is no contact, patients experience a false sense of security, thinking that the nurses know about arrythmias when they do not.

Patients' and Nurses' Weighted Results

Based on patients' and nurses' experiences of RPM, Table 3 presents weighted results with clinical implications and needed interventions for changed practice.

Table 3. Weighted results based on patients' and nurses' experiences of remote patient monitoring (RPM).

Weighted results	Clinical implications	Needed interventions for changed practice
Experiences and satisfaction		
Both patients and nurses had good experiences of RPM, but patients were more positive than the nurses.	Acknowledge dissatisfaction among nurses and identify obstacles to work with RPM.	Engage device manufactures to arrange online seminars and support for nurses involved in RPM.
A few patients and nurses found RPM unnecessary.	Continue to offer RPM to patients with implantable cardioverter-defibrillator (ICD).	Provide education and motivational support for RPM to every patient that receives an ICD.
Some nurses found it challenging to customize the correct alert settings for the individual patient, which resulted in a high number of transmissions.	Provide practical information to nurses about how and when to individualize alert settings.	Engage device manufacturers to arrange online seminars and support for nurses involved in RPM.
Security and safety		
Most patients experienced that RPM increased security, and this was in line with the nurses' perceptions, since they knew there would be an alert in case of malfunction or arrhythmias.	Provide targeted written information about RPM to patients and nurses and highlight the security aspect.	Distribute a pamphlet with appropriate local information to patients and nurses new in RPM positions, in addition to the specific information from the manufacturer.
Some patients highlighted that atrial fibrillation was not automatically reported by one specific RPM manufacturer, which affected the feeling of security.	Provide targeted written information about RPM to patients and give information about the data collection.	Distribute a pamphlet with information from the specific manufacturer and inform the patient about the possibility to perform a patient-initiated transmission in case of tachycardia.
Technical aspects		
Nurses found the technical equipment difficult for the patients to handle more often than did the patients.	Identify the patient's perceptions about how the RPM operates.	Offer technical support given by the manufacturer and provide the patients with written contact information. Arrange for timely and repeated group information targeting technical issues for patients.
Some patients had misconceptions about being continuously monitored in real time, 24/7.	Identify the patient's perceptions about how the RPM operates.	Proactively bring up how and when the data are being transferred to the clinic and the importance of calling the emergency service center in case of a life-threatening illness.
Having to learn all the different systems with different platforms and log-ins was difficult and stressful for nurses.	Offer nurses new in the RPM position a mentorship program covering technical aspects and solutions.	Engage device manufacturers to arrange online seminars and support for nurses involved in RPM. Ask device manufacturers to help set up networks with nurses working with the same platform.
Nurses highlighted how time-consuming it was to trace and handle different technical problems and the time it takes trying to reach the patient by phone (ie, when the monitor lost contact with the server).	Provide targeted written information about RPM to patients and nurses and highlight the technical aspects.	Offer technical support given by the manufacturer and provide the patients with written contact information. Arrange for timely and repeated group information targeting technical issues for patients.
Emotional aspects		
Only few patients were worried or anxious about what the RPM entailed, while half of the nurses felt distressed by the responsibility that accompanied their work with RPM.	Identify the nurses' perceptions about their workload when handling RPM.	Contact the device manufacturers and ask them to arrange online seminars and support for nurses involved in RPM. Define a written decision algorithm for the clinic in order to standardize the handling of transmissions.
It was considered as a lack of safety by some patients when traveling and not bringing the home monitor.	Provide targeted written information about RPM and traveling routines to patients.	Distribute a pamphlet with appropriate local information, including a clear *travel plan* and how the individual patient is recommended to act during travel.
Some patients felt that they had no one to talk to when questions arose and did not want to bother the nurse with a phone call, while others found it reassuring to have the possibility to call the ICD nurse when needed.	Acknowledge the emotional aspect of being an ICD recipient and identify those in need of extended support.	Proactively bring up the emotional aspect and offer emotional support. Provide written contact information for the clinic.

Weighted results	Clinical implications	Needed interventions for changed practice
Organization of care		
Most of the patients described that it was positive not needing to travel to hospital for follow-up as often as before, but some patients considered the fewer number of office-based visits to the ICD clinic as a disadvantage, since they appreciated the face-to-face interaction.	Identify the patient's needs and wishes for follow-up of the device.	Offer person-centered care with individual follow-up appointments and/or provide telephone-based support in between the office-based follow-ups. Provide repeated patient education about RPM and its management using a person-centered approach and by applying, for example, teach-back methodology.
Patients wanted to receive information directly from their remote monitored device.	Use automated direct call messaging for follow-up in patients with a device. Encourage patients to log in to their medical record electronically (when appropriate) to access the notes from the latest remote follow-up.	Using specific apps in smartphones, patients may have the possibility to check the website with information about their own device and to communicate or chat online with the health care personnel involved in the care of the patient in the future.
Some hospitals had not planned ahead before implementing RPM; they had no routines on how to document transmissions and lacked action plans on how to handle alerts.	Identify the workflow and perform meticulous care planning before the implementation of RPM.	Define a written decision algorithm for the clinic in order to standardize the handling of transmissions.
Hospital managers and heads of departments did not recognize the time-consuming work nurses did when handling daily transmissions.	Acknowledge nurses' workloads by giving the heads of departments insight into how the remote transmissions impact on the regular appointments.	Provide a supportive environment for RPM (ie, activities that do not involve direct patient interaction), since the most frequently reported barrier for not implementing RPM is found to be lack of reimbursement. To prevent dissatisfaction by the nurses, new working structures might be necessary.

Discussion

Principal Findings

A main finding in this study was that both patients and nurses found several positive aspects with RPM, but the nurses noted more downsides than did the patients. There could be several explanations to this. Two survey studies exploring telemonitoring in HF in various countries found that health care professionals described several patient-related barriers due to physical or mental conditions. Further, the nurses have seen a range of different technical problems in different patients over time, while the individual patients only have their own experiences and, therefore, a more optimistic and positive view on the technology [34,35].

Patients expressed that they could live a more normal life after receiving RPM. This confirms results in previous research showing that patients were content with RPM and did not feel like patients as much anymore [16,36]. This is probably related to the fact that patients felt that RPM increased security and safety for both themselves and their relatives. Also, for the vast majority, it was considered as an advantage that RPM lead to fewer office-based visits. The advantage that RPM is less time-consuming than in-hospital follow-ups has also been seen in other studies [16]; however, both in this study and in previous research, the lack of direct face-to-face contact was missed by some patients [37].

Even as patients received information and education about RPM, some patients believed that RPM included live transmissions, both day and night, and that someone continuously watched their electrocardiograms as when being monitored with telemetry during a hospital stay. Ottenberg et al found that when patients were prescribed RPM for their ICD but then did not install the home monitor, it was largely attributed to not understanding the purpose of the RPM system or being unsure whether their system was correctly transmitting information [37]. This highlights the need for repeated patient education about RPM and its management using a person-centered approach and by applying, for example, teach-back methodology [38]. High-quality training has been found to improve patients' understanding and comprehension and has been positively associated with anxiety and acceptance levels [23].

Downsides described by the nurses were often related to organizational issues; for example, a clear description of their own responsibilities handling alerts. Further, about half of the nurses also found the responsibility associated with managing RPM stressful due to different technologies, and they had concerns related to patient safety. A recent review describes the requirement of the referring nurse to be an expert in cardiac pacing and device follow-up, and a daily connection with a website should be performed to evaluate received alerts. To avoid stress and worries for the nurses, every center must define a written decision algorithm in order to standardize the handling of alerts [13].

It was also referred to as "invisible work" that is not recognized as time-consuming by managers and without proper reimbursement systems. It is important that health systems provide a supportive environment for RPM (ie, activities that do not involve direct patient interaction), since the most frequently reported barrier for not implementing RPM was found to be lack of reimbursement [17,39]. In the eHealth era,

with increasing remote monitoring of various treatments, symptoms, and devices, adapting the organization of care is key and organizations need to adapt to best make use of remote monitoring.

Nurses also found handling the daily transmissions as burdensome and time-consuming since it was difficult to tailor the alarm limits for each patient. It is known, as shown previously, that many transmissions are patient initiated without any event. In a recent study by Ninni et al [40] that included 1423 transmissions, it was found that as many as 77% were initiated by the patients, and only about 3% of the transmissions per patient led to actions or interventions by the health care personnel. The authors stressed the need to optimize automatic transmissions and focus on patient education to reduce the workload at the device clinic.

In addition, technical concerns took a lot of time, with nurses trying to locate the problem and get in contact with the patient. This led to stress and an increased workload, and nurses also felt that it was a false safety for the patient who did not always realize that the transmission failed. A recent study highlights the encounter through telephone calls that took place related to home monitoring. Five types of clinical work were performed that may also refer to RPM: inclusion work, coordination work, diagnostic work, education work, and comfort work [41]. The authors found that telephone calls increased time spent in telemonitoring, and most telephone calls contained more than two issues [41].

Currently, patients do not receive information directly from their remote monitored device, which patients in this study pointed out as a downside. However, findings from a feasibility [42] and evaluation [20] study in the United States suggest that it is not only feasible to deliver data from remote monitoring directly to patients, but also that this data sharing does not adversely impact clinic workflow and that patients perceive a benefit from having access to their remote monitoring data. The same result was found by Mirro et al when evaluating the impact of sharing ICD data summaries through a patient portal. At the end of the study, two-thirds of patients were satisfied with the amount of information received through the electronic or paper ICD data summary. Further, providing patients with their device data did not increase ICD-specific clinical workload [43].

In the future, by using specific apps in smartphones, patients may have the ability to check the website with information about their own devices and to communicate or chat online with the health care personnel involved in the care of the patient.

Conclusions

Both patients and ICD nurses found RPM to be safe and to increase the sense of security for patients and caregivers. There was a discrepancy between nurses and patients with regard to the technical equipment, where very few patients, but every other nurse, stated that the technology was difficult for the patients to handle.

Few patients were worried or anxious about the RPM, while half of the nurses felt distressed by the responsibility that accompanied their work. Nurses also described it as time-consuming to contact patients in case of alerts. To improve RPM from the perspective of the nurses, the organizational routines, reimbursement systems, and the balance of responsibilities and workloads need to be reviewed.

Acknowledgments

We would like to thank all patients and nurses who participated in the study. We also thank Lotta Altis and Johan Lyckström, ICD nurse specialists, for their help during data collection.

Conflicts of Interest

None declared.

References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Authors/Task Force Members, Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2016 Aug;18(8):891-975 [FREE Full text] [doi: 10.1002/ejhf.592] [Medline: 27207191]
2. Wijers SC, van der Kolk BYM, Tuinenburg AE, Doevendans PAF, Vos MA, Meine M. Implementation of guidelines for implantable cardioverter-defibrillator therapy in clinical practice: Which patients do benefit? Neth Heart J 2013 Jun;21(6):274-283 [FREE Full text] [doi: 10.1007/s12471-013-0407-x] [Medline: 23572330]
3. Lampert R. Managing with pacemakers and implantable cardioverter defibrillators. Circulation 2013 Oct 01;128(14):1576-1585. [doi: 10.1161/CIRCULATIONAHA.113.001555] [Medline: 24081953]
4. Ploux S, Varma N, Strik M, Lazarus A, Bordachar P. Optimizing implantable cardioverter-defibrillator remote monitoring: A practical guide. JACC Clin Electrophysiol 2017 Apr;3(4):315-328 [FREE Full text] [doi: 10.1016/j.jacep.2017.02.007] [Medline: 29759443]
5. Brignole M, Auricchio A, Baron-Esquivias G, Bordachar P, Boriani G, Breithardt O, ESC Committee for Practice Guidelines (CPG), Document Reviewers, et al. 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: The Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in

collaboration with the European Heart Rhythm Association (EHRA). Eur Heart J 2013 Aug;34(29):2281-2329. [doi: 10.1093/eurheartj/eht150] [Medline: 23801822]
6. European Society of Cardiology (ESC), European Heart Rhythm Association (EHRA), Brignole M, Auricchio A, Baron-Esquivias G, Bordachar P, et al. 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: The Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). Europace 2013 Aug;15(8):1070-1118. [doi: 10.1093/europace/eut206] [Medline: 23801827]
7. Slotwiner D, Varma N, Akar JG, Annas G, Beardsall M, Fogel RI, et al. HRS Expert Consensus Statement on remote interrogation and monitoring for cardiovascular implantable electronic devices. Heart Rhythm 2015 Jul;12(7):e69-e100. [doi: 10.1016/j.hrthm.2015.05.008] [Medline: 25981148]
8. Guédon-Moreau L, Lacroix D, Sadoul N, Clémenty J, Kouakam C, Hermida J, ECOST Trial Investigators. A randomized study of remote follow-up of implantable cardioverter defibrillators: Safety and efficacy report of the ECOST trial. Eur Heart J 2013 Feb;34(8):605-614 [FREE Full text] [doi: 10.1093/eurheartj/ehs425] [Medline: 23242192]
9. Calò L, Gargaro A, De Ruvo E, Palozzi G, Sciarra L, Rebecchi M, et al. Economic impact of remote monitoring on ordinary follow-up of implantable cardioverter defibrillators as compared with conventional in-hospital visits. A single-center prospective and randomized study. J Interv Card Electrophysiol 2013 Jun;37(1):69-78. [doi: 10.1007/s10840-013-9783-9] [Medline: 23515883]
10. Braunschweig F, Anker SD, Proff J, Varma N. Remote monitoring of implantable cardioverter-defibrillators and resynchronization devices to improve patient outcomes: Dead end or way ahead? Europace 2019 Jun 01;21(6):846-855 [FREE Full text] [doi: 10.1093/europace/euz011] [Medline: 30903152]
11. Bertini M, Marcantoni L, Toselli T, Ferrari R. Remote monitoring of implantable devices: Should we continue to ignore it? Int J Cardiol 2016 Jan 01;202:368-377. [doi: 10.1016/j.ijcard.2015.09.033] [Medline: 26432486]
12. Fauchier L, Sadoul N, Kouakam C, Briand F, Chauvin M, Babuty D, et al. Potential cost savings by telemedicine-assisted long-term care of implantable cardioverter defibrillator recipients. Pacing Clin Electrophysiol 2005 Jan;28 Suppl 1:S255-S259. [doi: 10.1111/j.1540-8159.2005.00071.x] [Medline: 15683509]
13. Lucà F, Cipolletta L, Di Fusco SA, Iorio A, Pozzi A, Rao CM, et al. Remote monitoring: Doomed to let down or an attractive promise? Int J Cardiol Heart Vasc 2019 Sep;24:100380 [FREE Full text] [doi: 10.1016/j.ijcha.2019.100380] [Medline: 31193998]
14. Heidbüchel H, Lioen P, Foulon S, Huybrechts W, Ector J, Willems R, et al. Potential role of remote monitoring for scheduled and unscheduled evaluations of patients with an implantable defibrillator. Europace 2008 Mar;10(3):351-357. [doi: 10.1093/europace/eun010] [Medline: 18245771]
15. Watanabe E, Kasai A, Fujii E, Yamashiro K, Brugada P. Reliability of implantable cardioverter defibrillator home monitoring in forecasting the need for regular office visits, and patient perspective. Japanese HOME-ICD study. Circ J 2013;77(11):2704-2711 [FREE Full text] [doi: 10.1253/circj.cj-13-0387] [Medline: 23903000]
16. Petersen HH, Larsen MCJ, Nielsen OW, Kensing F, Svendsen JH. Patient satisfaction and suggestions for improvement of remote ICD monitoring. J Interv Card Electrophysiol 2012 Sep;34(3):317-324. [doi: 10.1007/s10840-012-9675-4] [Medline: 22418649]
17. Mairesse GH, Braunschweig F, Klersy K, Cowie MR, Leyva F. Implementation and reimbursement of remote monitoring for cardiac implantable electronic devices in Europe: A survey from the health economics committee of the European Heart Rhythm Association. Europace 2015 May;17(5):814-818. [doi: 10.1093/europace/euu390] [Medline: 25713012]
18. Landolina M, Perego GB, Lunati M, Curnis A, Guenzati G, Vicentini A, et al. Remote monitoring reduces healthcare use and improves quality of care in heart failure patients with implantable defibrillators: The evolution of management strategies of heart failure patients with implantable defibrillators (EVOLVO) study. Circulation 2012 Jun 19;125(24):2985-2992. [doi: 10.1161/CIRCULATIONAHA.111.088971] [Medline: 22626743]
19. Versteeg H, Timmermans I, Widdershoven J, Kimman G, Prevot S, Rauwolf T, et al. Effect of remote monitoring on patient-reported outcomes in European heart failure patients with an implantable cardioverter-defibrillator: Primary results of the REMOTE-CIED randomized trial. Europace 2019 Sep 01;21(9):1360-1368. [doi: 10.1093/europace/euz140] [Medline: 31168604]
20. Daley C, Toscos T, Mirro M. Data integration and interoperability for patient-centered remote monitoring of cardiovascular implantable electronic devices. Bioengineering (Basel) 2019 Mar 17;6(1):25 [FREE Full text] [doi: 10.3390/bioengineering6010025] [Medline: 30884894]
21. Pedersen SS, Knudsen C, Dilling K, Sandgaard NCF, Johansen JB. Living with an implantable cardioverter defibrillator: Patients' preferences and needs for information provision and care options. Europace 2017 Jun 01;19(6):983-990. [doi: 10.1093/europace/euw109] [Medline: 27267553]
22. Akar JG, Bao H, Jones P, Wang Y, Chaudhry SI, Varosy P, et al. Use of remote monitoring of newly implanted cardioverter-defibrillators: Insights from the patient related determinants of ICD remote monitoring (PREDICT RM) study. Circulation 2013 Nov 26;128(22):2372-2383. [doi: 10.1161/CIRCULATIONAHA.113.002481] [Medline: 24043302]
23. Laurent G, Amara W, Mansourati J, Bizeau O, Couderc P, Delarche N, EDUCAT Registry Investigators. Role of patient education in the perception and acceptance of home monitoring after recent implantation of cardioverter defibrillators: The

EDUCAT study. Arch Cardiovasc Dis 2014 Oct;107(10):508-518 [FREE Full text] [doi: 10.1016/j.acvd.2014.06.009] [Medline: 25218008]
24. Boriani G, Auricchio A, Klersy C, Kirchhof P, Brugada J, Morgan J, European Heart Rhythm Association, Eucomed. Healthcare personnel resource burden related to in-clinic follow-up of cardiovascular implantable electronic devices: A European Heart Rhythm Association and Eucomed joint survey. Europace 2011 Aug;13(8):1166-1173. [doi: 10.1093/europace/eur026] [Medline: 21345922]
25. Pedersen SS, van Domburg RT, Theuns DAMJ, Jordaens L, Erdman RAM. Concerns about the implantable cardioverter defibrillator: A determinant of anxiety and depressive symptoms independent of experienced shocks. Am Heart J 2005 Apr;149(4):664-669. [doi: 10.1016/j.ahj.2004.06.031] [Medline: 15990750]
26. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
27. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
28. Lisspers J, Nygren A, Söderman E. Hospital Anxiety and Depression Scale (HAD): Some psychometric data for a Swedish sample. Acta Psychiatr Scand 1997 Oct;96(4):281-286. [doi: 10.1111/j.1600-0447.1997.tb10164.x] [Medline: 9350957]
29. Moser DK, Riegel B, McKinley S, Doering LV, Meischke H, Heo S, et al. The Control Attitudes Scale-Revised: Psychometric evaluation in three groups of patients with cardiac illness. Nurs Res 2009;58(1):42-51 [FREE Full text] [doi: 10.1097/NNR.0b013e3181900ca0] [Medline: 19092554]
30. Årestedt K, Ågren S, Flemme I, Moser DK, Strömberg A. A psychometric evaluation of the four-item version of the Control Attitudes Scale for patients with cardiac disease and their partners. Eur J Cardiovasc Nurs 2015 Aug;14(4):317-325. [doi: 10.1177/1474515114529685] [Medline: 24671774]
31. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
32. The ICN Code of Ethics for Nurses. Geneva, Switzerland: International Council of Nurses (ICN); 2012. URL: https://www.icn.ch/sites/default/files/inline-files/2012_ICN_Codeofethicsfornurses_%20eng.pdf [accessed 2020-09-16]
33. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
34. Aamodt IT, Lycholip E, Celutkiene J, Strömberg A, Atar D, Falk RS, et al. Health care professionals' perceptions of home telemonitoring in heart failure care: Cross-sectional survey. J Med Internet Res 2019 Feb 06;21(2):e10362 [FREE Full text] [doi: 10.2196/10362] [Medline: 30724744]
35. Kato NP, Johansson P, Okada I, de Vries AE, Kinugawa K, Strömberg A, et al. Heart failure telemonitoring in Japan and Sweden: A cross-sectional survey. J Med Internet Res 2015 Nov 13;17(11):e258 [FREE Full text] [doi: 10.2196/jmir.4825] [Medline: 26567061]
36. Timmermans I, Meine M, Zitron E, Widdershoven J, Kimman G, Prevot S, et al. The patient perspective on remote monitoring of patients with an implantable cardioverter defibrillator: Narrative review and future directions. Pacing Clin Electrophysiol 2017 Jul;40(7):826-833. [doi: 10.1111/pace.13123] [Medline: 28612995]
37. Ottenberg AL, Swetz KM, Mueller LA, Gerhardson S, Mueller PS. "We as Human Beings Get Farther and Farther Apart": The experiences of patients with remote monitoring systems. Heart Lung 2013;42(5):313-319 [FREE Full text] [doi: 10.1016/j.hrtlng.2013.03.002] [Medline: 23582212]
38. Ha Dinh TT, Bonner A, Clark R, Ramsbotham J, Hines S. The effectiveness of the teach-back method on adherence and self-management in health education for people with chronic disease: A systematic review. JBI Database System Rev Implement Rep 2016 Jan;14(1):210-247. [doi: 10.11124/jbisrir-2016-2296] [Medline: 26878928]
39. Dubner S, Auricchio A, Steinberg JS, Vardas P, Stone P, Brugada J, et al. ISHNE/EHRA expert consensus on remote monitoring of cardiovascular implantable electronic devices (CIEDs). Europace 2012 Feb;14(2):278-293. [doi: 10.1093/europace/eur303] [Medline: 22232544]
40. Ninni S, Delahaye C, Klein C, Marquie C, Klug D, Lacroix D, et al. A report on the impact of remote monitoring in patients with S-ICD: Insights from a prospective registry. Pacing Clin Electrophysiol 2019 Mar;42(3):349-355. [doi: 10.1111/pace.13598] [Medline: 30609077]
41. Andersen TO, Nielsen KD, Moll J, Svendsen JH. Unpacking telemonitoring work: Workload and telephone calls to patients in implanted cardiac device care. Int J Med Inform 2019 Sep;129:381-387 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.06.021] [Medline: 31445281]
42. Daley CN, Chen EM, Roebuck AE, Ghahari RR, Sami AF, Skaggs CG, et al. Providing patients with implantable cardiac device data through a personal health record: A qualitative study. Appl Clin Inform 2017 Oct;8(4):1106-1116 [FREE Full text] [doi: 10.4338/ACI-2017-06-RA-0090] [Medline: 29241248]
43. Mirro M, Daley C, Wagner S, Rohani Ghahari R, Drouin M, Toscos T. Delivering remote monitoring data to patients with implantable cardioverter-defibrillators: Does medium matter? Pacing Clin Electrophysiol 2018 Nov;41(11):1526-1535. [doi: 10.1111/pace.13505] [Medline: 30225880]

Abbreviations

CRT: cardiac resynchronization therapy
CRT-D: cardiac resynchronization therapy defibrillator
EVOLVO: Evolution of Management Strategies of Heart Failure Patients With Implantable Defibrillators
HF: heart failure
ICD: implantable cardioverter-defibrillator
PREDICT-RM: Patient-Related Determinants of ICD Remote Monitoring
REMOTE-CIED: Remote patient management for Cardiac Implantable Electronic Devices
RPM: remote patient monitoring

Edited by G Eysenbach; submitted 22.04.20; peer-reviewed by C Daley, J Guzman-Clark, P Wonggom; comments to author 26.06.20; revised version received 08.07.20; accepted 26.07.20; published 28.09.20.

Please cite as:
Liljeroos M, Thylén I, Strömberg A
Patients' and Nurses' Experiences and Perceptions of Remote Monitoring of Implantable Cardiac Defibrillators in Heart Failure: Cross-Sectional, Descriptive, Mixed Methods Study
J Med Internet Res 2020;22(9):e19550
URL: http://www.jmir.org/2020/9/e19550/
doi:10.2196/19550
PMID:32985997

Original Paper

Client Satisfaction and Experience With Telepsychiatry: Development and Validation of a Survey Using Clinical Quality Domains

Eva Serhal[1], MBA, PhD; Anne Kirvan[1], MSW; Marcos Sanches[2], MSc; Allison Crawford[1], PhD, MD

[1]Virtual Mental Health and Outreach, Centre for Addiction and Mental Health, Toronto, ON, Canada

[2]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Allison Crawford, PhD, MD
Virtual Mental Health and Outreach
Centre for Addiction and Mental Health
825-250 College St
Toronto, ON, M5T1R8
Canada
Phone: 1 4165358501 ext 30318
Email: allison.crawford@utoronto.ca

Abstract

Background: Telepsychiatry is an increasingly used model of mental health care that connects patients with psychiatrists at a distance via videoconference. Telepsychiatry is an effective clinical intervention that improves access to quality care in regions with limited resources or in clinical situations where in-person care is unavailable.

Objective: This study aims to develop a validated survey tool to measure patient experience and satisfaction with telepsychiatry based on the quality of care domains. This study also seeks to understand which health service outcomes were most strongly correlated with overall satisfaction in the context of telepsychiatry.

Methods: The survey created in this study was developed and validated with a panel of subject matter and process experts and was piloted with 274 patients who received clinical consultations through the TeleMental Health Program at the Centre for Addiction and Mental Health. Factor analysis was used to determine correlations between questions and quality of care domains and was also used to assess model fit.

Results: The study provides a validated survey to measure patient satisfaction and experience with telepsychiatry across 4 domains: access and timeliness, appropriateness, effectiveness, and safety. Both safety and access and timeliness were found to be statistically significant predictors of satisfaction in our sample.

Conclusions: By situating patient satisfaction and experience within this framework, the survey facilitates patient data collection and interpretation through a clinical quality lens.

(J Med Internet Res 2020;22(9):e19198) doi:10.2196/19198

KEYWORDS

telemedicine; psychiatry; mental health; patient satisfaction; quality of health care

Introduction

Telepsychiatry is an increasingly common modality of mental health care that connects patients with psychiatrists at a distance via videoconference [1]. This mental health care delivery model is an effective clinical intervention that reduces geographical barriers and improves access to care in regions with limited resources [2]. Although satisfaction with telepsychiatry is commonly reported as being high among service users and service providers alike [3], opportunities for research exist in the development and validation of quantitative and qualitative indices to measure clinical service, satisfaction, and experience and thus to ensure access to quality care [4].

Patient satisfaction and experience are widely accepted as gauges of health service performance and are commonly used as measures of overall quality of care in health systems [5-9]. Keeping patient experience at the forefront of quality measurement ensures that services remain acceptable and

appropriate to patients by being responsive to their needs [10]. Several patient satisfaction surveys for telepsychiatry have been developed and used in research and practice [3,11,12]. Two patient surveys that we were able to locate in the literature have been validated—one that assesses patient experiences of privacy and security [13] and another designed to measure the attitudes of laypeople and providers toward telepsychiatry [14]—yet no comprehensive satisfaction measures have been validated to date.

Validated measures of patient satisfaction have, however, been developed in the broader field of telemedicine [15,16]. Yip et al [16] identified a method of validating telemedicine tools and provided validated questions in the following domains: audiovisual quality, general satisfaction, accessibility, use of equipment and correctness of vital signs being transferred, level of comfort, and satisfaction with the telemedicine encounter.

The literature points to a number of factors to consider in the development and validation of patient experience surveys. Litwin's [17] literature on psychometrics highlights the importance of assessing reliability criteria, selecting appropriate validity criteria, and scaling or scoring. Mazor et al [5] recommend collecting a large sample of surveys when evaluating patient experience to increase reliability of the ratings, as individual patients are more likely to provide similar ratings across items. It is also important to recognize the possibility of biases inherent to data collection caused by nonresponse. For example, satisfied patients are more likely to respond than unsatisfied patients [5,18-22].

To gain a more comprehensive understanding of patient satisfaction and experience with telepsychiatry, this research study had two primary objectives: (1) to develop a validated survey tool that measures different dimensions of client experience and satisfaction with telepsychiatry and (2) to examine the relationship between dimensions of clinical quality, as measured in the survey, and overall satisfaction.

Methods

Conceptual Framework

This study links several health service outcomes (HSOs) commonly used to describe patient experience and overall quality of health interventions [23-26]. The survey was developed with items relating to HSO domains to obtain a robust assessment of the overall patient experience of telepsychiatry. HSO domains included the following:

1. *Safety*: "Avoiding injuries to patients from the care that is intended to help them" [23]
2. *Effectiveness*: "Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse)" [23]
3. *Efficiency*: "Avoiding waste, in particular waste of equipment, supplies, ideas, and energy" [23]
4. *Access and timeliness*: Improving the "fit between the patient and the health care system" [24], and "reducing waits and sometimes harmful delays for those who receive and those who give care" [23]
5. *Appropriateness*: Perceived fit, relevance, and compatibility of the intervention [26].

Patient-centeredness, a commonly used metric of service quality, was not explicitly included as a domain in the survey, as the survey itself is intended to provide a patient-centered measure of service quality.

Survey Development

The survey questions were generated in 2 steps. First, we identified relevant questions derived from the literature, previously developed questions drawn from historical program surveys, and any additional questions we felt were necessary to understand the quality of a patient's experience with telepsychiatry. Next, to ensure that there was representation of questions across the HSO domains, we mapped the total set of 21 questions to the domains of safety, efficiency, access and timeliness, appropriateness, and effectiveness. We included access with timeliness because, for many patients living in rural areas, lack of timeliness is a direct result of a lack of access to local care. These domains align with other proposed domains for the measurement of telehealth, such as access, cost, experience, and effectiveness [27]. Research ethics board review was not required for survey development and use, as this is a program evaluation and quality improvement project.

Question Validation

Content Validity

Following a similar validation model recommended by Yip et al [16], a panel of 7 subject matter experts were engaged, including 4 psychiatrists that deliver telepsychiatry, a social worker, a research coordinator, and an administrative director, as well as 3 process experts in survey design and marketing. This panel was asked to review the survey for relevance, clarity, plain language, and consistent distribution of questions across HSO domains. The panel was pulled together as a group to meet and discuss the questions. Those who were not able to participate in person were emailed with written instructions that asked them to do the following:

1. Rate on a scale of 1-5 how important the question is to include in the survey
2. Rate on a scale of 1-5 the suitability of each question within its HSO domain
3. Rate on a scale of 1-5 how well the survey measures each HSO domain
4. Recommend new questions if something important is missing
5. Flag awkward or unclear questions for discussion by the survey design panel.

Where there were discrepancies, there was discussion between the team members about the nature of the discrepancy until consensus was reached.

Testing the Validity of the Survey

This survey was piloted with patients who received clinical consultations through our telepsychiatry program. Although Yip et al [16] validated their patient survey with 38 patients, our study included a sample size of 274 patients for validation

to ensure appropriate representation and sufficient power to conduct analysis of all 21 questions.

The 21-question survey was printed and sent to telemedicine coordinators from 25 referring primary care sites throughout Ontario. Site selection was based on primary care teams that regularly referred to the program and that had dedicated telemedicine coordinators to support anonymous survey collection. The sites were primarily in rural or underserved areas. Coordinators were provided with instructions that requested that they provide a survey to each telepsychiatry patient after the patient completed their telepsychiatry appointment. It is important to note, however, that 8 of the sites did not return any surveys, and there was no opportunity to determine the overall response rate.

This study used convenience sampling, a nonprobability sampling method, whereby a survey was provided to all patients who completed a telepsychiatry appointment. Random sampling was not possible based on the nature of the clinical service. Each survey had a unique identifier and a code for the primary care organization name. No personal identifiers were collected.

Each site was sent 10 paper-based surveys and 10 letter-sized envelopes, an instruction page, a schedule for collection of surveys, and an introductory letter to the site's executive director and telemedicine coordinator. More surveys and envelopes were sent on an as-needed basis. Instructions asked the telemedicine coordinators to provide each patient, at the beginning of their telepsychiatry consultation, with a copy of the survey, an envelope, and a pen. Telemedicine coordinators then returned to collect the completed and sealed surveys at the end of the consultation. Each mailed survey package included 4 postage-paid manila envelopes for quarterly collection. Quarterly emails asking each site to send us all of their collected anonymous patient satisfaction surveys were sent as a reminder.

Factorial Structure

To test the structure of the scale, with previously defined 5 constructs, a confirmatory factor analysis (CFA) model was adjusted to the data using Mplus 7.11 [28]. The 5-point ordinal Likert scale was accommodated by the weighted least squares with mean and variance-adjusted chi-square statistic (WLSMV) estimator in Mplus [29]. The WLSMV estimator, which is robust to non-normality, uses polychoric correlation and provides adjusted chi-squared statistics. Simulations conducted by Flora and Curran [30] show that the WLSMV estimator is likely to work well with our sample size. The WLSMV estimator uses all available data through pair-wise correlation calculation and assumes missing completely at random. Goodness of fit for the CFA models were assessed using the chi-square statistic (for which a significant value is considered to be evidence of lack of fit), the root mean square error of approximation (RMSEA), and the confirmatory fit index (CFI). The chi-square statistic is sensitive to sample size and departures from multivariate normality, which justifies a focus also on the other fit indices [31]. All the inferential statistical analyses described in this study accounted for the clustering of patients within each of the 17 primary care sites, using an extension of the pseudo-likelihood method, described by Asparouhov and Muthem [32].

Once the initial CFA was conducted, model respecification was conducted based on the interpretation of the model according to the meaning of items, original model coefficients, and exploratory factor analysis (EFA), which used Geomin Rotation with 4 and 5 factors to be consistent with the theory underlying the development of the tool and with results from the initial CFA. The EFA was also conducted in Mplus 7.11, with the WLSMV estimator for ordinal data. A final test of the structure was conducted using CFA. Given that the respecification and test of the final model used the same data used to fit the original CFA, we recognize that replication with independent data would add evidence to our results.

Reliability

Once all client experience scales were obtained, their reliability was assessed. Reliability of a scale has to do with the extent to which the scale produces similar results under similar conditions and is measured via the internal consistency of the items used. This type of reliability has been popularly estimated by Cronbach alpha [33]; however, this method has been criticized [34,35]. The reported reliability estimates in this study are, therefore, calculated by the method outlined by Raykov [35] and can be interpreted in the same way as the coefficient alpha: the closer to 1 its value is, the higher the reliability, with values higher than 0.7 being considered acceptably reliable [32,36].

Impact on Overall Satisfaction

The client satisfaction and experience survey is useful on its own by providing a tested tool that measures aspects of patient experience with telepsychiatry. Under the assumption that these experiences are important drivers of overall attitudes of patients toward health care services and institutions and that these attitudes can be important summary indicators of overall performance, we extended our final CFA model by specifying causal paths from the validated experiences to the overall satisfaction, also included in our questionnaire as a 5-point Likert scale question. As this is just an extension of our final CFA, its technical details are the same as described above for the CFA model.

Results

Descriptive Analysis

Of the 274 patients who returned the questionnaire in the 2016/2017 and 2017/2018 fiscal years, 147 (53.6%) reported their gender as female, 118 (43.1%) reported their gender as male, 3 reported their gender as transgender, and 1 reported their gender as intersex. Moreover, 4 participants reported their gender as other and 3 preferred not to answer. With respect to age, 25.5% of the participants were aged less than 30 years, 36.2% were between 30 and 49 years, 36.9% were above 50 years, and 5 respondents reported that they preferred not to answer. A total of 175 patients (63.9%) stated this to be their first experience with telepsychiatry, and 26 (6.5%) patients had been hospitalized for mental health issues in the previous 12 months. At the end of the study, 274 surveys were completed in 17 different primary care sites. Items missing values were observed in 57 surveys, 72% of which had 3 or fewer missing values.

In general, patient-reported experiences tended to be very positive; 58% of the patients strongly agree with the overall satisfaction question and 91% either strongly agree or agree. The complete results for all survey questions are shown in Table 1.

Table 1. Client satisfaction survey results.

Questions	Number of survey responses, n	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
Q1. I am satisfied with the length of time I had to wait between my referral and the Telepsychiatry appointment.	269	5 (1.9)	27 (10.0)	40 (14.9)	107 (39.8)	90 (33.5)
Q2. It was easy to book my Telepsychiatry appointment.	267	3 (1.1)	11 (4.1)	26 (9.7)	116 (43.4)	111 (41.6)
Q3. During my Telepsychiatry appointment, I was able to see the psychiatrist clearly.	270	6 (2.2)	1 (0.4)	12 (4.4)	79 (29.3)	172 (63.7)
Q4. During my Telepsychiatry appointment, I was able to hear the psychiatrist clearly.	271	3 (1.1)	4 (1.5)	13 (4.8)	81 (29.9)	170 (62.7)
Q5. I am confident that the psychiatrist and my health care providers are working as a team.	270	4 (1.5)	1 (0.4)	17 (6.3)	100 (37.0)	148 (54.8)
Q6. I feel that there was an adequate amount of time allotted for the Telepsychiatry appointment.	270	5 (1.9)	8 (3.0)	14 (5.2)	97 (35.9)	146 (54.1)
Q7. I felt comfortable during my Telepsychiatry appointment.	270	5 (1.9)	5 (1.9)	45 (16.7)	96 (35.6)	119 (44.1)
Q8. I believe Telepsychiatry is just as effective as an in-person psychiatry appointment.	269	11 (4.1)	20 (7.4)	55 (20.4)	86 (32.0)	97 (36.1)
Q9. I was able to get an appointment through Telepsychiatry sooner than an in-person psychiatry appointment.	257	7 (2.7)	10 (3.9)	51 (19.8)	84 (32.7)	105 (40.9)
Q10. I felt that confidentiality was protected throughout my Telepsychiatry appointment.	268	4 (1.50)	2 (0.7)	10 (3.7)	107 (39.9)	145 (54.1)
Q11. The psychiatrist understood my concerns.	266	6 (2.3)	1 (0.4)	18 (6.8)	102 (38.3)	139 (52.3)
Q12. The psychiatrist treated me with courtesy and respect.	269	4 (1.5)	1 (0.4)	4 (1.5)	62 (23.0)	198 (73.6)
Q13. The psychiatrist explained my diagnosis in a way that I could understand.	259	4 (1.5)	2 (0.8)	31 (12.0)	87 (33.6)	135 (52.1)
Q14. The psychiatrist involved me in decisions about my treatment plan.	260	3 (1.2)	1 (0.4)	33 (12.7)	100 (38.5)	123 (47.3)
Q15. The psychiatrist explained the benefits and risks of any medications he/she recommended.	254	2 (0.8)	11 (4.3)	48 (18.9)	81 (31.9)	112 (44.1)
Q16. I am confident that I will be able to follow the psychiatrist's recommendations.	264	4 (1.5)	5 (1.9)	43 (16.3)	93 (35.2)	119 (45.1)
Q17. I understand what to do if I have a mental health emergency following this appointment.	265	4 (1.5)	14 (5.3)	32 (12.1)	98 (37.0)	117 (44.2)
Q18. The physical location of my Telepsychiatry appointment was convenient for me to get to.	269	4 (1.5)	5 (1.9)	11 (4.1)	90 (33.5)	159 (59.1)
Q19. I experienced a significant improvement in my mental health while I was waiting for my Telepsychiatry appointment.	260	54 (20.8)	71 (27.3)	79 (30.4)	27 (10.4)	29 (11.2)
Q20. I experienced a significant decline in my mental health while I was waiting for my Telepsychiatry appointment.	253	40 (15.8)	54 (21.3)	108 (42.7)	31 (12.3)	20 (7.9)
Q21. Overall, I am satisfied with the Telepsychiatry appointment.	266	5 (1.9)	1 (0.4)	18 (6.8)	88 (33.1)	154 (57.9)

Preliminary CFA

The initial CFA tested the theoretical factor structure defined in the earlier stages of the development of the scales. The model fit was fairly good, but it did indicate some inconsistencies with the factor structure as defined (how the questions related to the HSO domains). The chi-square was 373.34 with 160 degrees of freedom and $P<.001$. The RMSEA (0.07) and CFI (0.97) both indicated that the proposed model was acceptable overall; however, the factor loadings demonstrated that several of the questions did not align well with other factor items, as indicated by coefficient variance and nonsignificance, particularly with respect to effectiveness and efficiency (Table 2).

Table 2. Factor coefficients for the preliminary confirmatory factor analysis.

Survey questions and factors	Estimate[a]	SE[b]	Estimate/SE[c]	*P* value[d]
Factor 1: Access and timeliness				
Q1. I am satisfied with the length of time I had to wait between my referral and the Telepsychiatry appointment.	1.00[e]	N/A[f]	N/A	N/A
Q2. It was easy to book my Telepsychiatry appointment.	1.09	0.10	11.30	<.001
Q16. I am confident that I will be able to follow the psychiatrist's recommendations.	1.45	0.11	13.70	<.001
Q18. The physical location of my Telepsychiatry appointment was convenient for me to get to.	1.28	0.08	15.27	<.001
Factor 2: Appropriateness				
Q8. I believe Telepsychiatry is just as effective as an in-person psychiatry appointment.	1.00[e]	N/A	N/A	N/A
Q11. The psychiatrist understood my concerns.	1.32	0.05	25.02	<.001
Q14. The psychiatrist involved me in decisions about my treatment plan.	1.28	0.05	27.16	<.001
Factor 3: Effectiveness				
Q3. During my Telepsychiatry appointment, I was able to see the psychiatrist clearly.	1.00[e]	N/A	N/A	N/A
Q4. During my Telepsychiatry appointment, I was able to hear the psychiatrist clearly.	1.01	0.06	16.82	<.001
Q5. I am confident that the psychiatrist and my health care providers are working as a team.	1.18	0.03	38.14	<.001
Q13. The psychiatrist explained my diagnosis in a way that I could understand.	1.30	0.05	25.58	<.001
Q19. I experienced a significant improvement in my mental health while I was waiting for my Telepsychiatry appointment.	0.36	0.10	3.70	<.001
Q20. I experienced a significant decline in my mental health while I was waiting for my Telepsychiatry appointment.	−0.01	0.06	−0.24	.81
Factor 4: Efficiency				
Q6. I feel that there was an adequate amount of time allotted for the Telepsychiatry appointment.	1.00[e]	N/A	N/A	N/A
Q9. I was able to get an appointment through Telepsychiatry sooner than an in-person psychiatry appointment.	0.70	0.05	13.03	<.001
Factor 5: Safety				
Q7. I felt comfortable during my Telepsychiatry appointment.	1.00[e]	N/A	N/A	N/A
Q10. I felt that confidentiality was protected throughout my Telepsychiatry appointment.	1.10	0.05	23.60	<.001
Q12. The psychiatrist treated me with courtesy and respect.	1.19	0.04	27.43	<.001
Q15. The psychiatrist explained the benefits and risks of any medications he/she recommended.	0.97	0.05	20.22	<.001
Q17. I understand what to do if I have a mental health emergency following this appointment.	1.02	0.06	16.52	<.001

[a]Estimate, model estimate of factor loadings.

[b]SE: SE of estimates.

[c]Estimate/SE: ratio between estimates and their SE (under the assumption of normality of estimates, this ratio has a standard normal distribution, which is used to calculate the *P* value).

[d]*P* value: estimate of the probability of a coefficient equal or larger than that found under the null hypothesis.

[e]Coefficients were fixed at 1.00 to allow model identification.

[f]N/A: not appliable.

Model Respecification

Besides not showing very good fit indices, the CFA showed some other issues that led us to respecify the model. This respecification took into account the correction of the issues in the initial model, the theoretical basis for the constructs, the modification indices from the CFA, and insights from an EFA. We detail the changes made to the model in the following section.

The efficiency factor showed different signs of poor fit. It had inconsistent correlations above 1 with other factors. Items Q6 ("I feel that there was an adequate amount of time allotted for the telepsychiatry appointment") and Q9 ("I was able to get an appointment through telepsychiatry sooner than an in-person psychiatry appointment") showed poor variance. On the basis of the reinterpretation of these items and the fact that they did not come together in the EFA, Q6 was moved to factor effectiveness and Q9 to factor access and timeliness. The efficiency factor was removed, as none of the questions lined up well with this domain.

Items Q19 ("I experienced a significant improvement in my mental health while I was waiting for my telepsychiatry appointment") and Q20 ("I experienced a significant decline in my mental health while I was waiting for my telepsychiatry appointment") in the effectiveness factor did not fit well, as can be seen in Table 2. They had almost none of their variance explained, and the EFA showed them together, but with low loadings and separated from other items. Upon reviewing the interpretation of these items, it was decided that they should be removed from the analysis because they represented an effect of the treatment received while awaiting a telepsychiatry consultation rather than an experience with the telepsychiatry service itself. Item Q16 ("I am confident that I will be able to follow the psychiatrist's recommendations") was moved from the access and timeliness domain to the appropriateness domain, and item Q13 ("The psychiatrist explained my diagnosis in a way that I could understand") was moved from the effectiveness domain to the safety domain as first suggested by the largest modification indices and confirmed by the EFA and theory-related considerations. Despite other significant modification indices and some differences in the EFA structure, we did not make any further changes to the model as they were not warranted by theoretical considerations.

Final CFA

A second CFA was run to test the structure of the revised model, using the same measures of fit as outlined above. Chi-square was 288.19 with 129 degrees of freedom and $P<.001$. The RMSEA (0.07) and CFI (0.98) continued to suggest a good fit. Overall, we observed remarkable improvement in the model fit. The reliability was lower but still acceptable for access and timeliness (0.72) and good for the other 3 factors (appropriateness=0.81, effectiveness=0.83, and safety=0.86). The final model is shown in Table 3.

Table 3. Factor coefficients for the final confirmatory factor analysis.

Survey questions and factors	Estimate[a]	SE[b]	Estimate/SE[c]	*P* value[d]
Factor 1: Access and timeliness				
Q1. I am satisfied with the length of time I had to wait between my referral and the Telepsychiatry appointment.	1.00[e]	N/A[f]	N/A	N/A
Q2. It was easy to book my Telepsychiatry appointment.	1.10	0.10	11.23	<.001
Q9. I was able to get an appointment through Telepsychiatry sooner than an in-person psychiatry appointment.	0.98	0.08	11.90	<.001
Q18. The physical location of my Telepsychiatry appointment was convenient for me to get to.	1.29	0.08	15.63	<.001
Factor 2: Appropriateness				
Q8. I believe Telepsychiatry is just as effective as an in-person psychiatry appointment.	1.00[e]	N/A	N/A	N/A
Q11. The psychiatrist understood my concerns.	1.31	0.05	24.56	<.001
Q14. The psychiatrist involved me in decisions about my treatment plan.	1.27	0.05	27.54	<.001
Q16. I am confident that I will be able to follow the psychiatrist's recommendations.	1.18	0.05	22.10	<.001
Factor 3: Effectiveness				
Q3. During my Telepsychiatry appointment, I was able to see the psychiatrist clearly.	1.00[e]	N/A	N/A	N/A
Q4. During my Telepsychiatry appointment, I was able to hear the psychiatrist clearly.	1.01	0.06	16.60	<.001
Q5. I am confident that the psychiatrist and my health care providers are working as a team.	1.17	0.03	36.97	<.001
Q6. I feel that there was an adequate amount of time allotted for the Telepsychiatry appointment.	1.09	0.06	18.63	<.001
Factor 4: Safety				
Q7. I felt comfortable during my Telepsychiatry appointment.	1.00[e]	N/A	N/A	N/A
Q10. I felt that confidentiality was protected throughout my Telepsychiatry appointment.	1.11	0.05	23.50	<.001
Q12. The psychiatrist treated me with courtesy and respect.	1.20	0.04	28.15	<.001
Q13. The psychiatrist explained my diagnosis in a way that I could understand.	1.20	0.03	40.02	<.001
Q15. The psychiatrist explained the benefits and risks of any medications he/she recommended.	0.97	0.05	19.94	<.001
Q17. I understand what to do if I have a mental health emergency following this appointment.	1.02	0.06	16.68	<.001

[a]Estimate: model estimate of factor loadings.

[b]SE: SE of estimates.

[c]Estimate/SE: ratio between estimates and their SE (under the assumption of normality of estimates, this ratio has *a standard normal distribution, which is used to calculate the P* value).

[d]*P* value: estimate of the probability of a coefficient equal or larger than that found under the null hypothesis.

[e]Coefficients were fixed at 1.00 to allow model identification.

[f]N/A: not appliable.

Patient Experience

Overall, the client satisfaction surveys demonstrated high ratings from patients across the 4 domains of access and timeliness (mean 16.7 out of 20, SD 2.8), appropriateness (mean 16.8 out of 20, SD 3.0), effectiveness (mean 17.9 out of 20, SD 2.6), and safety (mean 25.8 out of 30, SD 4.0), with the highest overall score for effectiveness (Table 4). The total factor score was calculated for each factor by summing the scores for each item. Subjects with missing values in at least one item of a factor did not have the total score calculated for that factor. We observe that for all factors, our sample shows total scores that are concentrated on the higher end of the scale, as can be seen in Table 4.

Table 4. Survey results stratified by factor.

Factor	Participants, n	Mean (SD)	SE	Minimum	First quartile	Median	Third quartile	Maximum
Access and timeliness	246	16.7 (2.8)	0.18	4	15	17	19	20
Appropriateness	252	16.8 (3.0)	0.19	4	15	17	20	20
Effectiveness	268	17.9 (2.6)	0.16	4	16	19	20	20
Safety	245	25.8 (4.0)	0.26	6	23	26	29	30

Table 5 demonstrates the coefficient of the factors as predictors of overall satisfaction. Access and timeliness, and safety were found to be statistically significant predictors of overall satisfaction based on the *P* value, appropriateness was almost significant, and effectiveness was not found to significantly predict overall satisfaction.

Table 5. Confirmatory factor analysis with overall satisfaction.

Factor	Estimate	SE	Estimate/SE[a]	*P* value
Access and timeliness	0.29	0.12	2.46	.01
Appropriateness	0.37	0.23	1.61	.11
Effectiveness	−0.02	0.10	−0.17	.86
Safety	0.62	0.21	2.93	.003

[a]Estimate/SE: ratio between estimates and their SE (under the assumption of normality of estimates, this ratio has a standard normal distribution, which is used to calculate the *P* value)

Discussion

Principal Findings

This study provides a validated survey tool to measure patient satisfaction and experience with telepsychiatry across 4 HSO domains: access and timeliness, appropriateness, effectiveness, and safety. This validated survey tool serves as a good model for future research and program evaluation that ensures patient experience is collected through a quality of care lens. By clustering items into HSO domains, survey results allow for targeted quality improvement efforts that address patient-identified gaps in service quality. This survey has global relevance in telepsychiatry as well as the broader fields of telemental health and telemedicine.

Consistent with previous research studies [3,11,12], patients' responses suggest high levels of satisfaction with telepsychiatry services. In this study, patient satisfaction was high across all 4 domains, with the highest overall score for the effectiveness domain, followed by safety, appropriateness, and access and timeliness. Our study also sought to understand which HSOs were most strongly correlated with overall satisfaction in the context of telepsychiatry and found access and timeliness, and safety to be statistically significant predictors. The term safety in this case represents both physical and psychological safety, which is evident in the way that the six individual items clustered together around issues of effective communication, comfort, confidentiality, respect, and patient knowledge about what steps to take in case of an emergency.

Although the effectiveness of telepsychiatry is well established in the literature [37] and is a key dimension of overall quality of care [23,25], it is important to note that effectiveness was not found to be a statistically significant predictor of patient satisfaction in this study. Appropriateness was found to be almost significant; however, only access and timeliness, and safety were found to statistically significantly predict satisfaction. A potential reason for this is that the majority of patients accessing telepsychiatry in Ontario are in rural areas and thus are likely to have fairly limited access to timely service locally, so timeliness and access are major components in their overall satisfaction [38]. This information has important clinical practice and policy implications when we consider the crucial role that access, timeliness, and safety play in shaping overall patient satisfaction with telepsychiatry services, including ensuring safe spaces for patients, safety protocols and guidelines for clinicians, and a focus on timely service that supports patients with significant barriers to access.

Another possible explanation for the lack of significance of effectiveness on satisfaction may be related to the temporal relationship of the survey with the appointment, especially given that a majority of the respondents (63.9%) were receiving telepsychiatry for the first time. In other words, judgments of effectiveness may not be as salient to satisfaction immediately following a first appointment. In the future, examining survey responses after repeat appointments will be an important direction.

Although telepsychiatry programs have increased the reach of mental health services, research suggests that telepsychiatry is still underused by patients and providers alike [38]. As work on this survey neared completion, we encountered an unexpected rapid increase in telepsychiatry use in the context of the pandemic caused by COVID-19. Many commentators predict a sustained increase in the use of telehealth and other digital health technologies [39,40]. Now more than ever, understanding the perspectives of end users and stakeholders, including patients and providers, is crucial in service planning, evaluation, and research. Patient satisfaction and experience surveys provide a useful avenue for feedback that can be used to guide

patient-centered quality improvement initiatives that are responsive to the needs of patients [10].

To ensure the acceptability of telepsychiatry services, it will be important to ensure that services address those factors that have been identified as the most significant contributors to overall satisfaction for patients, namely, access and timeliness, and safety. As telepsychiatry services continue to increase in use, we need to keep patient experiences and perspectives at the forefront of not only quality measurement but also program and systems planning.

Future Directions

With a survey tool now developed and adjusted based on our factor analysis findings and validated for use, we plan to utilize the survey to conduct a second analysis to assess the overall quality of our service and identify opportunities for program improvement within each domain. This will allow us to determine if the survey is helpful in gauging the impact of iterative quality improvement initiatives. Simultaneously, we will conduct additional consultations relating to new and emerging questions that arise in relation to patient experiences in telepsychiatry. Our team actively engages people with lived experience in the co-design of programs, and we will seek to involve patients in further improving the survey. Finally, we are always open to the feedback of health providers working in our program and those referring patients to our program. If changes are made as a result of these consultations, additional factor analysis and validation will be conducted.

After the advent of the COVID-19 pandemic, our team has also committed to taking a digital health equity perspective, working toward more equitable access to telehealth, while also recognizing that larger structural factors may impact access to technology and/or influence comfort with accessing care using technology [41]. In the future, we will ensure that the survey addresses equity, an important HSO identified by the Institute of Medicine [23]. We plan to include additional questions to assess equity as it relates to digital health, including questions on language, education, ethnicity, age, gender, and culturally safe and compassionate care.

In addition, qualitative research may help to illuminate some of the barriers and facilitators to the use of telepsychiatry by both patients and providers. Although measures of patient experience and satisfaction inform person-centered care, it is important to continue to be open to patient-derived expressions of experience and quality and to ensure that these are not overlooked.

Limitations

Self-selection bias is a possible limitation to this study, as the ability to self-select may affect the response rate, which is unknown for this study. For example, the most satisfied or unsatisfied patients are more likely to respond to the survey [42]. To minimize nonresponse, telemedicine coordinators at the different community provider sites were asked to give the patient experience survey to all patients at the beginning of their appointment and to collect them upon completion. Another possible bias highlighted by Williams et al [43] is that telemedicine survey respondents may be more likely to respond favorably to satisfaction surveys given the perceived potential impacts on health care service delivery and access. It could be argued, however, that the perceived pressure to respond favorably is diminished in the case of telepsychiatry consultations, as in the context of our service, the consulting provider does not provide ongoing care to the patient, and patient responses are anonymous. In future research, it would be beneficial to know the response rate to better understand the magnitude of response bias. To account for the potential bias whereby patients rate all HSOs highly based on their preference for a particular psychiatrist [42], rather than the overall process or experience of their telepsychiatry appointment, we looked at aggregate means instead of data at the individual psychiatrist level.

Revising the model based partly on information from the data itself lowers the level of evidence for our final CFA model. Despite the fact that we only addressed the most relevant fit issues and based model modifications on theoretical considerations, it is possible that our final model is affected by overfitting problems. The only way to address that, however, would be through the replication of the analysis with a different data set, which we plan to do in future iterations.

Conclusions

This study sought to address a notable gap in the literature with respect to validated measures of patient satisfaction with telepsychiatry. This study used HSOs as a guiding framework for the development and validation of a patient satisfaction and experience survey. By situating patient satisfaction and experience within this framework, the survey facilitates patient data collection and interpretation through a clinical quality lens. This study also illuminates the clinical quality domains that would benefit from targeted quality improvement initiatives to further improve overall patient experience and satisfaction with telepsychiatry.

Acknowledgments

The authors would like to acknowledge the members of the Virtual Mental Health team at the Centre for Addiction and Mental Health for supporting the development of the client satisfaction and experience survey as well as the ongoing dissemination and collection of surveys. In particular, the authors would like to thank Hamza Sandhu, Denise Canso, Aubrey Vigna, Amanda Arena, Dr David Goldbloom, Dr David Rodie, Dr Robert Cooke, and others. The authors would also like to acknowledge the contributions of the referring primary care sites throughout Ontario who have and continue to support this initiative and the patients who took the time to provide their valuable feedback.

Conflicts of Interest

None declared.

References

1. Conn D, Gajaria A, Madan R. Telepsychiatry: effectiveness and feasibility. Smart Homecare Technol Telehealth 2015 Apr:59. [doi: 10.2147/shtt.s45702]
2. Chakrabarti S. Usefulness of telepsychiatry: a critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015 Oct 22;5(3):286-304 [FREE Full text] [doi: 10.5498/wjp.v5.i3.286] [Medline: 26425443]
3. Schubert NJ, Backman PJ, Bhatla R, Corace KM. Telepsychiatry and patient-provider concordance. Can J Rural Med 2019;24(3):75-82 [FREE Full text] [doi: 10.4103/CJRM.CJRM_9_18] [Medline: 31249155]
4. Hilty D, Yellowlees P, Myers K, Parish M, Rabinowitz T. The effectiveness of e-mental healthvidence base, how to choose the model based on ease/cost/strength,future areas of research. In: E-Mental Health. New York, USA: Springer International Publishing; 2016.
5. Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. Health Serv Res 2002 Oct;37(5):1403-1417 [FREE Full text] [doi: 10.1111/1475-6773.11194] [Medline: 12479503]
6. Carlson MJ, Blustein J, Fiorentino N, Prestianni F. Socioeconomic status and dissatisfaction among HMO enrollees. Med Care 2000 May;38(5):508-516. [doi: 10.1097/00005650-200005000-00007] [Medline: 10800977]
7. Ford RC, Bach SA, Fottler MD. Methods of measuring patient satisfaction in health care organizations. Health Care Manage Rev 1997;22(2):74-89. [Medline: 9143904]
8. Rosenthal GE, Shannon SE. The use of patient perceptions in the evaluation of health-care delivery systems. Med Care 1997 Dec;35(11 Suppl):NS58-NS68. [doi: 10.1097/00005650-199711001-00007] [Medline: 9366880]
9. Young GJ, Meterko M, Desai KR. Patient satisfaction with hospital care: effects of demographic and institutional characteristics. Med Care 2000 Mar;38(3):325-334. [doi: 10.1097/00005650-200003000-00009] [Medline: 10718357]
10. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 3;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
11. Dham P, Gupta N, Alexander J, Black W, Rajji T, Skinner E. Community based telepsychiatry service for older adults residing in a rural and remote region- utilization pattern and satisfaction among stakeholders. BMC Psychiatry 2018 Sep 27;18(1):316 [FREE Full text] [doi: 10.1186/s12888-018-1896-3] [Medline: 30261845]
12. Hantke N, Lajoy M, Gould CE, Magwene EM, Sordahl J, Hirst R, et al. Patient satisfaction with geriatric psychiatry services via video teleconference. Am J Geriatr Psychiatry 2020 Apr;28(4):491-494. [doi: 10.1016/j.jagp.2019.08.020] [Medline: 31530457]
13. Zhou L, Thieret R, Watzlaf V, Dealmeida D, Parmanto B. A telehealth privacy and security self-assessment questionnaire for telehealth providers: development and validation. Int J Telerehabil 2019;11(1):3-14 [FREE Full text] [doi: 10.5195/ijt.2019.6276] [Medline: 31341542]
14. Tonn P, Reuter SC, Kuchler I, Reinke B, Hinkelmann L, Stöckigt S, et al. Development of a questionnaire to measure the attitudes of laypeople, physicians, and psychotherapists toward telemedicine in mental health. JMIR Ment Health 2017 Oct 3;4(4):e39 [FREE Full text] [doi: 10.2196/mental.6802] [Medline: 28974485]
15. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
16. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the telemedicine satisfaction questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
17. Litwin M. How to Measure Survey Reliability and Validity. Thousand Oaks, CA: Sage Publications; 1995.
18. Barkley WM, Furse DH. Changing priorities for improvement: the impact of low response rates in patient satisfaction. Jt Comm J Qual Improv 1996 Jul;22(6):427-433. [doi: 10.1016/s1070-3241(16)30245-0] [Medline: 8806045]
19. Etter JF, Perneger TV, Rougemont A. Does sponsorship matter in patient satisfaction surveys? A randomized trial. Med Care 1996 May;34(4):327-335. [doi: 10.1097/00005650-199604000-00004] [Medline: 8606557]
20. Lasek RJ, Barkley W, Harper DL, Rosenthal GE. An evaluation of the impact of nonresponse bias on patient satisfaction surveys. Med Care 1997 Jul;35(6):646-652. [doi: 10.1097/00005650-199706000-00009] [Medline: 9191708]
21. Pearson D, Maier ML. Assessing satisfaction and non-response bias in an HMO-sponsored employee assistance program. Empl Assist Q 1995 Jun 2;10(3):21-34. [doi: 10.1300/j022v10n03_02]
22. Woolliscroft JO, Howell JD, Patel BP, Swanson DB. Resident-patient interactions: the humanistic qualities of internal medicine residents assessed by patients, attending physicians, program supervisors, and nurses. Acad Med 1994 Mar;69(3):216-224. [doi: 10.1097/00001888-199403000-00017] [Medline: 8135980]

23. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
24. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. Med Care 1981 Mar;19(2):127-140. [doi: 10.1097/00005650-198102000-00001] [Medline: 7206846]
25. Mayberry RM, Nicewander DA, Qin H, Ballard DJ. Improving quality and reducing inequities: a challenge in achieving best care. Proc (Bayl Univ Med Cent) 2006 May;19(2):103-118 [FREE Full text] [doi: 10.1080/08998280.2006.11928138] [Medline: 16609733]
26. Robertson-Preidler J, Biller-Andorno N, Johnson TJ. What is appropriate care? An integrative review of emerging themes in the literature. BMC Health Serv Res 2017 Jun 30;17(1):452 [FREE Full text] [doi: 10.1186/s12913-017-2357-2] [Medline: 28666438]
27. Rural Telehealth Toolkit. Rural Health Information Hub. URL: https://www.ruralhealthinfo.org/toolkits/telehealth [accessed 2020-06-18]
28. Muthén L, Muthén B. Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén; 1998.
29. Brown T. Confirmatory Factor Analysis for Applied Research. New York, USA: Guilford Press; 2006.
30. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods 2004 Dec;9(4):466-491 [FREE Full text] [doi: 10.1037/1082-989X.9.4.466] [Medline: 15598100]
31. Wang J, Wang X. Structural Equation Modeling: Applications Using Mplus. Hoboken, NJ: John Wiley/Higher Education Press; 2012.
32. Asparouhov T, Muthem B. Multivariate Statistical Modeling With Survey Data. In: Federal Committee on Statistical Methodology. 2005 Presented at: FCSM'05; August 11-15, 2005; New York, USA.
33. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
34. Novick MR, Lewis C. Coefficient alpha and the reliability of composite measurements. Psychometrika 1967 Mar;32(1):1-13. [doi: 10.1007/bf02289400]
35. Raykov T. Alpha if item deleted: a note on loss of criterion validity in scale development if maximizing coefficient alpha. Br J Math Stat Psychol 2008 Dec;61(Pt 2):275-285. [doi: 10.1348/000711007X188520] [Medline: 17535482]
36. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993;78(1):98-104. [doi: 10.1037/0021-9010.78.1.98]
37. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jul 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
38. Serhal E, Crawford A, Cheng J, Kurdyak P. Implementation and utilisation of telepsychiatry in Ontario: a population-based study. Can J Psychiatry 2017 Oct;62(10):716-725 [FREE Full text] [doi: 10.1177/0706743717711171] [Medline: 28541753]
39. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jul 4;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
40. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and covid-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
41. Crawford A, Serhal E. Digital health equity and covid-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 2;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
42. Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. Eval Program Plann 1983;6(3-4):185-210. [doi: 10.1016/0149-7189(83)90002-2] [Medline: 10299618]
43. Williams TL, May CR, Esmail A. Limitations of patient satisfaction studies in telehealthcare: a systematic review of the literature. Telemed J E Health 2001;7(4):293-316. [doi: 10.1089/15305620152814700] [Medline: 11886667]

Abbreviations

CFA: confirmatory factor analysis
CFI: confirmatory fit index
EFA: exploratory factor analysis
HSO: health service outcome
RMSEA: root mean square error of approximation
WLSMV: weighted least squares with mean and variance-adjusted chi-square test statistic

Edited by G Eysenbach; submitted 08.04.20; peer-reviewed by D Hilty, P Yellowlees, N Hantke, C Kruse; comments to author 12.06.20; revised version received 19.07.20; accepted 25.07.20; published 29.09.20.

Please cite as:
Serhal E, Kirvan A, Sanches M, Crawford A
Client Satisfaction and Experience With Telepsychiatry: Development and Validation of a Survey Using Clinical Quality Domains
J Med Internet Res 2020;22(9):e19198
URL: http://www.jmir.org/2020/9/e19198/
doi:10.2196/19198
PMID:32755896

Original Paper

A Personalized Health Monitoring System for Community-Dwelling Elderly People in Hong Kong: Design, Implementation, and Evaluation Study

Hailiang Wang[1,2], PhD; Yang Zhao[1,3], PhD; Lisha Yu[3], PhD; Jiaxing Liu[3], BSc; Inez Maria Zwetsloot[4], PhD; Javier Cabrera[5], PhD; Kwok-Leung Tsui[1,3], PhD

[1]Centre for Systems Informatics Engineering, City University of Hong Kong, Hong Kong, China

[2]School of Design, The Hong Kong Polytechnic University, Hong Kong, China

[3]School of Data Science, City University of Hong Kong, Hong Kong, China

[4]Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China

[5]Department of Statistics and Biostatistics, Rutgers University, New Brunswick, NJ, United States

Corresponding Author:
Kwok-Leung Tsui, PhD
School of Data Science
City University of Hong Kong
Tat Chee Avenue
Hong Kong
China
Phone: 86 34422177
Email: kltsui@cityu.edu.hk

Abstract

Background: Telehealth is an effective means to assist existing health care systems, particularly for the current aging society. However, most extant telehealth systems employ individual data sources by offline data processing, which may not recognize health deterioration in a timely way.

Objective: Our study objective was two-fold: to design and implement an integrated, personalized telehealth system on a community-based level; and to evaluate the system from the perspective of user acceptance.

Methods: The system was designed to capture and record older adults' health-related information (eg, daily activities, continuous vital signs, and gait behaviors) through multiple measuring tools. State-of-the-art data mining techniques can be integrated to detect statistically significant changes in daily records, based on which a decision support system could emit warnings to older adults, their family members, and their caregivers for appropriate interventions to prevent further health deterioration. A total of 45 older adults recruited from 3 elderly care centers in Hong Kong were instructed to use the system for 3 months. Exploratory data analysis was conducted to summarize the collected datasets. For system evaluation, we used a customized acceptance questionnaire to examine users' attitudes, self-efficacy, perceived usefulness, perceived ease of use, and behavioral intention on the system.

Results: A total of 179 follow-up sessions were conducted in the 3 elderly care centers. The results of exploratory data analysis showed some significant differences in the participants' daily records and vital signs (eg, steps, body temperature, and systolic blood pressure) among the 3 centers. The participants perceived that using the system is a good idea (ie, attitude: mean 5.67, SD 1.06), comfortable (ie, self-efficacy: mean 4.92, SD 1.11), useful to improve their health (ie, perceived usefulness: mean 4.99, SD 0.91), and easy to use (ie, perceived ease of use: mean 4.99, SD 1.00). In general, the participants showed a positive intention to use the first version of our personalized telehealth system in their future health management (ie, behavioral intention: mean 4.45, SD 1.78).

Conclusions: The proposed health monitoring system provides an example design for monitoring older adults' health status based on multiple data sources, which can help develop reliable and accurate predictive analytics. The results can serve as a guideline for researchers and stakeholders (eg, policymakers, elderly care centers, and health care providers) who provide care for older adults through such a telehealth system.

(J Med Internet Res 2020;22(9):e19223) doi:10.2196/19223

KEYWORDS

telehealth monitoring; personalized health; technology acceptance; digital biomarkers; digital phenotyping; wearables; falls detection; fitness tracker; sensors; elderly population

Introduction

In Hong Kong, residents aged 65 years old and above will account for 33.7% of the total population in 2066, compared to 17.0% in 2018 [1]. Aging reduces the physical and cognitive capacities of older adults and affects their ability to live independently or perform daily activities. Older adults are also susceptible to chronic diseases (eg, hypertension, diabetes, and dementia). For example, approximately 73% of Hong Kong residents aged 75 and above have hypertension [2] and nearly 1 out of 10 community-dwelling residents aged 70 or above has dementia. Managing such chronic diseases or their exacerbations is associated with close to 80% of health care budgets [3]. Another critical health issue for older adults is falling, which can result in decreased mobility level, fear of falling, and even death [4]. Approximately 18% of Hong Kong community-dwelling older adults experience falls. Among fallers, approximately 10% incur bone fractures [5] and around 32% experience soft tissue injuries [6]. These falls are associated with increases of up to HKD 552 million (US$70 million) in extra annual health care costs, approximately 30% of which can be reduced through an effective fall prevention program [7].

Recently, the Hong Kong government proposed a policy of "aging in place," which encourages empowering older adults to remain in communities for long-term care services and to promote their well-being [8]. Such community-based services can ease the public financial burden as they are cheaper than public hospitals and can save costs that would be spent on misused health care resources (eg, unnecessary hospitalizations) [9]. Along with the shift from hospital care to community care, community-based health care systems are facing unprecedented challenges of limited capacity and resources to monitor older adults' health continuously. Moreover, the caregivers in communities may lack professional knowledge and thus cannot detect health anomalies or suggest the next treatment for the elderly when needed. Therefore, innovative solutions for continuous monitoring of the health of the community-dwelling elderly population and linking with health care professionals are needed in the Hong Kong health system.

Owing to rapid developments in information technology, telehealth monitoring has provided cost-effective and timely access to quality care [10-12]. Telehealth monitoring systems utilize telecommunication technologies (eg, digital monitoring sensors) to capture and deliver health data (eg, vital signs) and services between patients and health care professionals. In particular, such a system provides a feasible solution to the increasing demand for long-term care support and monitoring to community-dwelling elderly individuals who may have difficulties in accessing health services [13-16]. A review of the current literature showed that telehealth systems have been assisting older adults in specific health-related areas such as chronic conditions [17-22], falls [23,24], and general wellness [25-28]. For example, Or and Tao [19] developed a patient-centered, tablet computer-based self-monitoring system to enable older adults with type 2 diabetes and hypertension to measure and monitor their blood glucose and blood pressure. Sparks and colleagues [22] proposed a decision support system that seeks to help community nurses monitor the well-being of their chronically ill patients, using an all-in-one station-based health monitoring device. Doty and colleagues [24] developed a wearable multimodal monitoring system designed for the real-life long-term monitoring of patients susceptible to falls. In addition to trial studies on telehealth monitoring systems, we also found some national telehealth programs that have been implemented, such as the Whole Systems Demonstrator program of the UK Department of Health [29], the Care Coordination/Home Telehealth program introduced by Veterans Health Administration in the United States [10], and the Home Monitoring of Chronic Disease in Aged Care program funded by the Australian government [16]. Significant benefits have been reported, such as a 19% reduction in numbers of hospital admissions [10] and 45% reduction in mortality rates [29].

However, as reported in previous studies [16,28,30,31], there are still some restrictions and challenges that could block the timely detection of health deterioration when implementing continuous monitoring systems, such as offline data processing and analysis, usage of individual smart devices, and single-parameter measurement. Motivated by these challenges, we aimed to integrate some state-of-the-art techniques into these monitoring systems, such as advanced biomedical signal analysis, statistical data analysis, predictive analytics, and decision support, which can help provide efficient health care services [32-34]. In addition, we sought to design a system that can utilize various smart devices to collect different health-related measurements for providing an efficient and accurate approach to raising health awareness in community monitoring [35,36].

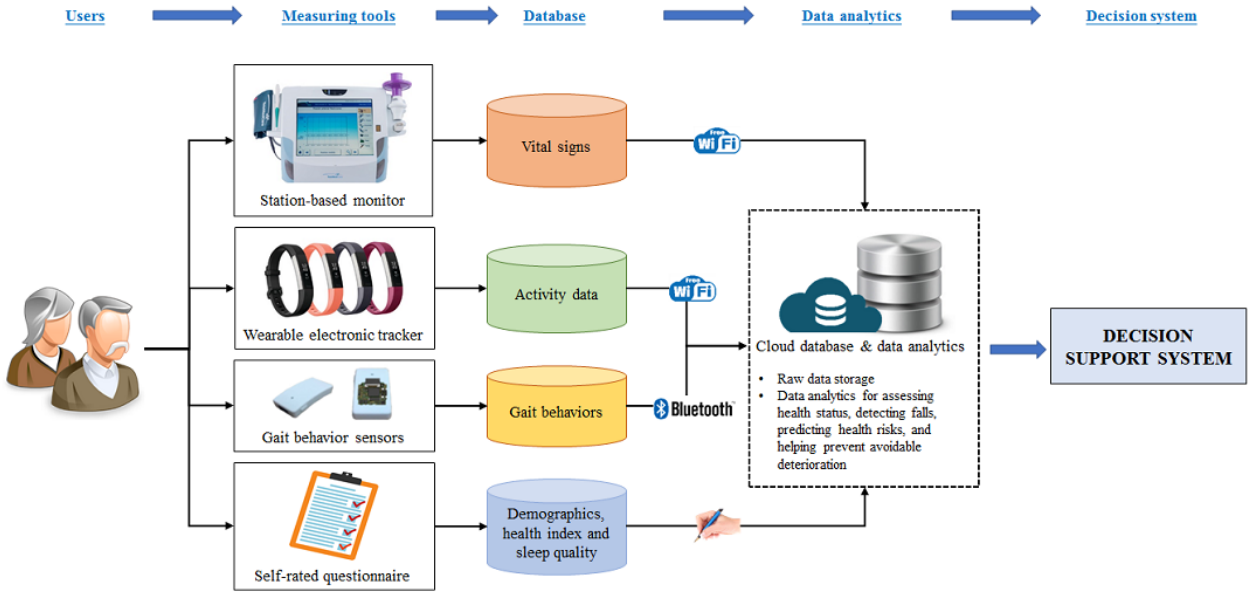
Methods

System Design

Schematic Diagram

Figure 1 shows the schematic diagram of our proposed personalized health monitoring system. The system captures and records older adults' health-related information such as continuous vital signs and gait behaviors through various measuring tools, and will be integrated and analyzed using state-of-the-art data mining techniques. When any statistically significant changes in daily records are detected, the decision support system will emit warnings to older adults, their family members, and their caregivers, who can then take appropriate interventions to prevent further health deterioration. The details of the measuring tools integrated into the system are described below.

Figure 1. Schematic diagram of the proposed personalized health monitoring system.



Vital Signs

Older adults' vital signs are measured and recorded using an all-in-one station-based telemonitoring device (TMC, Telemedcare Systems Pty Ltd, Sydney, Australia). This device was selected based on a comprehensive and independent technology assessment process as described previously [9,16,22]. The vital signs that a TMC unit can measure include body temperature, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate, and blood oxygen level (SpO_2). The vital sign data can be electronically sent to a centralized database for quality control and diagnostic purposes.

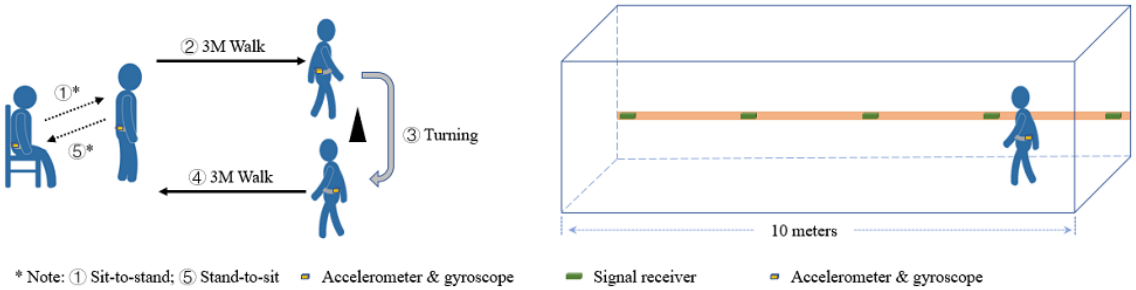
Daily Activity

A commercial device (Fitbit-Alta, Fitbit Inc, USA) is integrated into our proposed system to record older adults' steps and sleep data. All of these data will be synchronized and uploaded to the Fitbit cloud server for quality control and diagnoses.

Gait and Balance Sensor Signals

Older adults' gait and balance status are measured using wearable sensors that are cost-effective with few constraints on monitoring movements [37-39]. Older adults need to put on a sensor (ie, accelerometer and gyroscope) before performing a 3-meter timed up and go (3M-TUG) test and a 10-meter straight walking (10M-SW) test (see Figure 2). The 3M-TUG test is a well-known clinical test of gait mobility [40] with high reliability [41]. During the 3M-TUG test, older adults need to stand up from a chair, walk 3 meters, turn around, walk back 3 meters, and sit down on the chair. The completion time of the 3M-TUG test is recorded as it is associated with impaired mobility and increased fall risks [40]. Gait speed is cited as the "sixth vital sign" [42] to reflect functional and physiological changes [43,44], and can further reflect fall risk [45]. Gait speed can be calculated from the 10M-SW test with participants walking 10 meters in a straight line. Signal data of gait behaviors are collected from the wearable sensor during the two gait tests. In addition, the Berg Balance Scale score is collected by registered physiotherapists [46,47] to identify older adults who are prone to falls and in need of preventive treatments [48,49].

Figure 2. Illustrations of (left) the 3-meter timed up and go test and (right) the 10-meter straight walking test.



Demographic Information, Sleep Quality, and Wellness Status

Customized questionnaires were used to collect older adults' demographic information (eg, age, gender, and chronic disease history). Sleep quality was measured using the Pittsburgh Sleep Quality Index scale [50], one of the most commonly used clinical measures of sleep quality [51,52], with the score ranging from 0 to 21 (higher scores indicate worse sleep quality). Older adults were asked to self-report their daily wellness level, also called the Health Index, by rating on a 10-point scale from 1 ("feeling terrible") to 10 ("feeling terrific") [28].

System Implementation

During the implementation phase, we collected raw data for system algorithm development and examined users' acceptance of our proposed health monitoring system. A 3-month follow-up design was utilized in 3 centers that are part of a local nongovernment organization providing community services to the elderly [53]. Center A is a nursing home that provides 24-hour service to residents aged 60 or above and are mentally suitable for group living. Center B and Center C provide daycare services for residents aged 60 or above who live in the community.

Participants

Directors of the 3 centers first approached their center members, explained our study protocol, and invited older adults to participate. Based on the name lists from the directors, we recruited older adults who met all of the following inclusion criteria: (1) community-dwelling Hong Kong residents, (2) at least 60 years old, (3) willing to participate in the study, and (4) capable of cooperating in the assessment. Older adults with unstable or life-threatening illness were excluded. After completing all of the follow-up assessments, each participant was given a 50 HKD (US $6.50) supermarket coupon as a token of appreciation. The pilot study was approved by the Research Ethics Committee of City University of Hong Kong (reference number: 3-2-201803_02). All participants provided written informed consent before participating in the study.

Procedure

The implementation phase was scheduled from November 14, 2017 to February 13, 2018 in Center A; from December 17, 2017 to March 16, 2018 in Center B; and from March 1, 2018 to May 31, 2018 in Center C. We conducted follow ups every day during the 3-month period, excluding public holidays or special arrangements at the center (eg, special holiday leave in Center A; special training days in Center C). We ended up with a total of 58 follow ups for Center A, 63 follow ups for Center B, and 58 follow ups for Center C.

After obtaining participants' consent forms, trained research assistants collected participants' demographic and sleep quality information using questionnaires, and distributed each participant a Fitbit-Alta to wear 24 hours per day. The research assistants conducted the following operations in each follow-up visit, namely, every day during the 3-month period (except public holidays or special arrangements at the center).

First, the research assistants checked the battery of the Fitbit in use and, if needed, replaced it with a prepaired, fully charged Fitbit (each participant used two paired Fitbit devices during the pilot study). The research assistants then synchronized the Fitbit data with a tablet and uploaded it to the Fitbit server.

Second, the research assistants asked the participants to be seated in front of a TMC unit and guided each participant to use the TMC for measurements of vital signs. After all the measurements, the research assistants synchronized the vital sign data to the TMC server.

Third, the research assistants recorded the participants' self-rated health status (ie, health index).

In addition, the research assistants measured participants' body weight and performed a 3M-TUG test and a 10M-SW test once a week during the 3-month period (except for public holidays or special arrangements at the center).

System Evaluation

Exploratory data analysis was performed to summarize the collected datasets. After the 3-month pilot study, we conducted a survey through distributing a questionnaire to evaluate users' acceptance of the system. The participants were asked to rate their perceived acceptance of the system with respect to attitude (eg, "it is a wise idea to use this system") [54], self-efficacy (eg, "it is comfortable to perform self-monitoring via the system") [54], perceived usefulness (eg, "using this system for self-monitoring improves your health") [55], perceived ease of use (eg, "learning to perform self-monitoring via the system is easy for you") [55], and behavioral intention (eg, "you intend to perform self-monitoring using the system in the next 2 months") [56], using a 7-point Likert-type scale, ranging from 1 ("very strongly disagree") to 7 ("very strongly agree").

Results

Table 1 presents the demographic information, daily activities, and vital signs of the 45 participants. Exploratory data analysis showed significant differences among group means in age ($F_{2,42}$ =9.138, P=.001), steps ($F_{2,42}$=33.9, P<.001), body temperature ($F_{2,42}$=145.1, P<.001), and SBP ($F_{2,42}$=4.417, P=.02). No other significant group differences were found. Figure 3 shows the longitudinal variations of the vital signs among the 3 centers.

Table 1. Demographics, daily activities, and vital signs of the 45 participants stratified by center.

Variable	Center A (n=10)	Center B (n=24)	Center C (n=11)
Age (years), mean (SD), range	88.7 (3.7), 82-94	76.3 (7.8), 61-91	81.4 (9.9), 71-106
Gender, n (%)			
Female	8 (80)	21 (87)	4 (36.4)
Male	2 (20)	3 (13)	7 (63.6)
Chronic disease (self-reported), n (%)			
Hypertension	9 (90)	13 (54)	7 (58)
Heart disease	2 (20)	3 (13)	2 (17)
Stroke	0 (0)	1 (4)	3 (25)
Diabetes mellitus	2 (20)	5 (21)	3 (25)
Cancer	0 (0)	3 (13)	0 (0)
High cholesterol	0 (0)	8 (33)	4 (33)
Asthma	2 (20)	0 (0)	1 (8)
PSQI[a], mean (SD)	7.80 (3.01)	7.42 (3.99)	7.00 (1.91)
Health index, mean (SD)	7.06 (1.72)	8.35 (1.24)	6.33 (1.60)
Daily activities, mean (SD)			
Steps (number)	4751.8 (2481.7)	11817.5 (4089.1)	2751.5 (1718.5)
Sleep (hours)	7.08 (1.77)	7.35 (1.60)	6.76 (2.14)
Vital signs, mean (SD)			
Body temperature (°C)	36.45 (0.12)	35.89 (0.14)	36.72 (0.17)
DBP[b] (mmHg)	74.05 (7.58)	69.28 (8.37)	71.02 (4.28)
SBP[c] (mmHg)	142.20 (3.72)	130.10 (11.43)	130.63 (14.65)
Heart rate (beats/minute)	74.99 (5.92)	74.19 (10.73)	70.64 (11.28)
SpO_2[d] (%)	96.84 (1.62)	97.71 (1.29)	96.41 (1.90)

[a]PSQI: Pittsburgh Sleep Quality Index.

[b]DBP: diastolic blood pressure.

[c]SBP: systolic blood pressure.

[d]SpO_2: blood oxygen level.

Figure 4 shows an example of a segmented 3M-TUG task using accelerometer data and gyroscope data. Algorithms developed to segment the signal data into sit-to-stand, walking, and stand-to-sit are provided in our previous publication [57].

Overall, the participants strongly agreed that using the system is a good idea (mean 5.67, SD 1.06). The participants agreed that using the system is comfortable (mean 4.92, SD 1.11), useful to improve their health (mean 4.99, SD 0.91), and easy to use (mean 4.99, SD 1.00). In general, the participants showed a positive intention to use the first version of our personalized telehealth system in their future health management (mean 4.45, SD 1.78).

Figure 3. Vital signs of each participant from the 3 centers over time. The dashed lines represent mean values. BT: body temperature; DBP: diastolic blood pressure; SBP: systolic blood pressure; HR: heart rate; SpO_2: blood oxygen level.

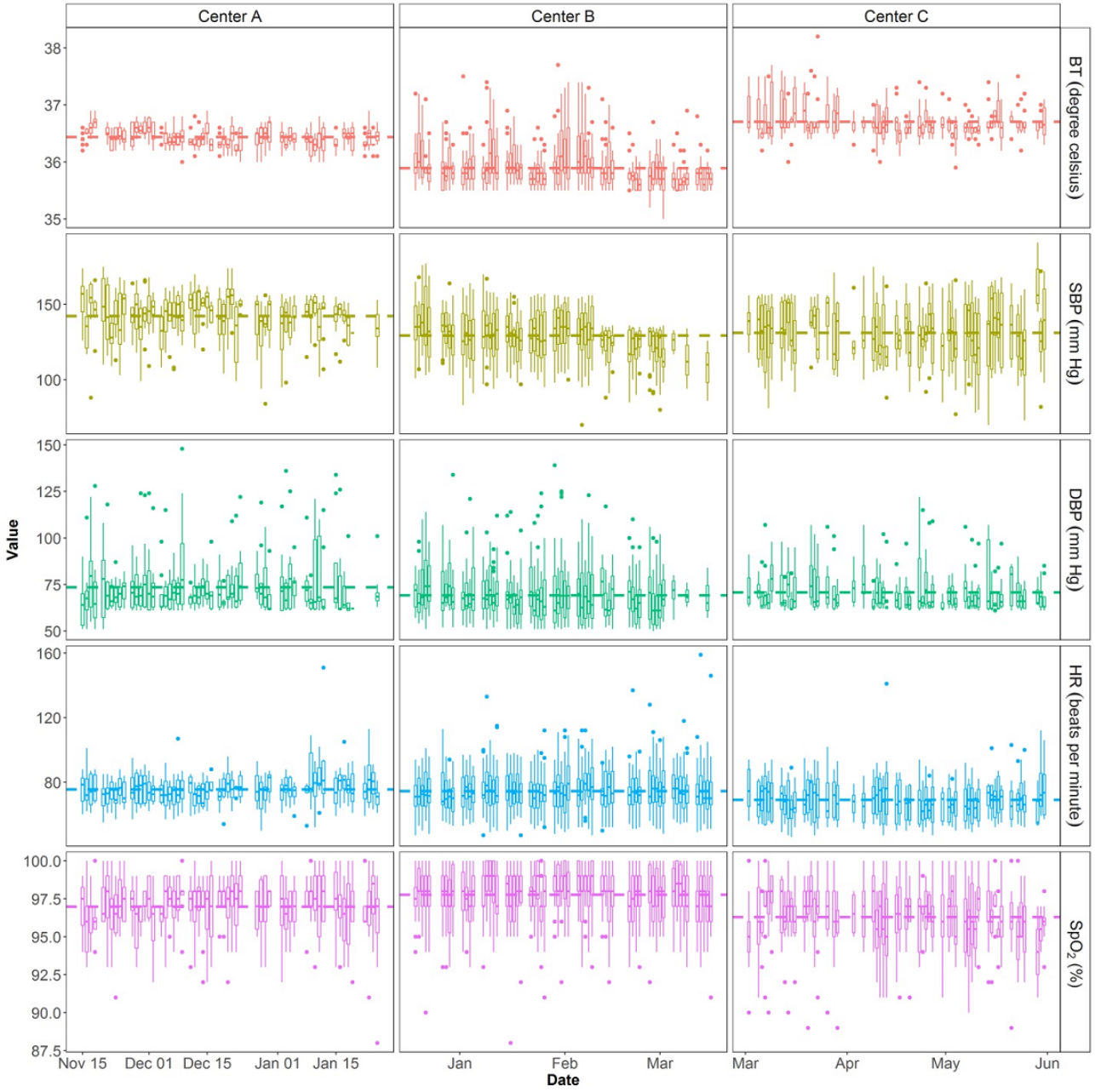
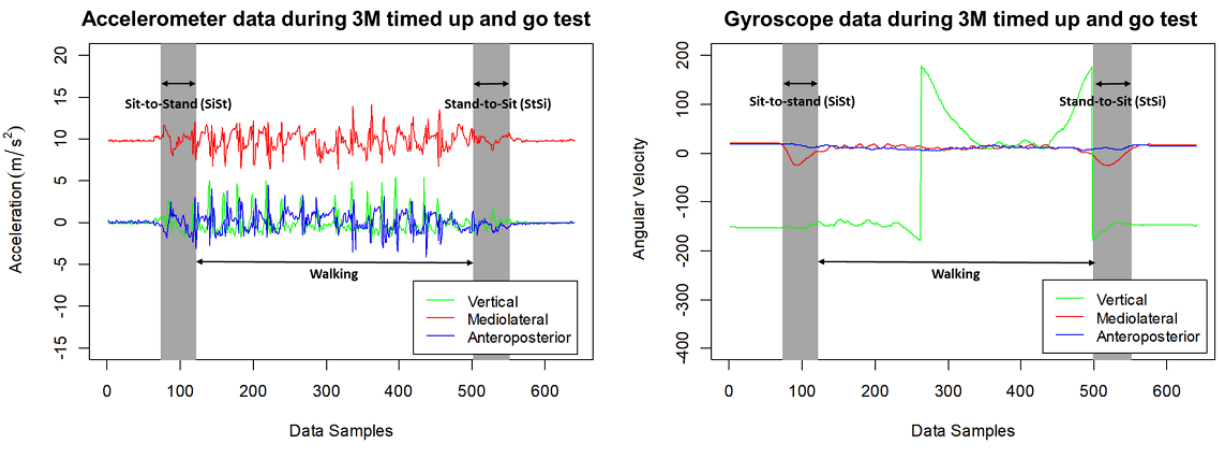


Figure 4. Example of segmented 3-meter (3M) timed up and go tasks using (left) accelerometer data and (right) gyroscope data.



Discussion

Principal Findings

Innovative health care solutions such as telehealth are a possible solution to support community caregivers to meet the increasing health services demand. In this paper, we proposed our first version of an integrated, personalized telehealth monitoring system and demonstrated its implementation for Hong Kong community-dwelling older adults. We further evaluated its user acceptance after 3 months. This system can help communicate and manage the data collected from different sources, detect health anomalies, provide wellness prediction, signal alerts on health risks, and propose health improvement advice for reduction of adverse health outcomes.

For elderly individuals with chronic illnesses or at high risk of falls, timely detection of health anomalies is critical in health management. Any adverse conditions remaining untreated could result in a higher chance of hospitalization and longer recovery times [25]. Existing monitoring devices or systems mainly focus on the monitoring of vital signs, which has limitations for wellness prediction. The raw data from various sources can be of multidimensions, multiscales, and of varying precision. It is therefore important to develop tools and protocols for integrating and mining the personalized health-related data collected from various devices. Our proposed system integrates continuous measurements of vital signs, daily activities, and gait behaviors, allowing for better analytics and interpretation of disease progression and fall risks. Forecasting the wellness of the elderly based on identified personalized rules may be a useful indicator in early anomaly detection and potential treatment [25]. One of the key observations of our findings is that health-related data (eg, number of steps, body temperature, and SBP) showed high variations among the 3 test centers. For example, older adults who lived in Center A commonly exhibited a lack of physical activities, whereas those recruited from Center B exhibited a more active lifestyle with more physical activities. Such differences among various populations (eg, older adults at different care centers) may or may not affect the development of accurate predictive analytics. Thus, future studies with a large sample size are needed to validate the existence of population variations in health-related data and to further examine how such variations affect health predictions.

The collected data provide a rich resource to develop models and algorithms for smart personalized health management through risk assessment, and disease and harm prevention. For example, the vital sign data can be analyzed and modeled to identify biomarkers for anomalies in health status. Based on the automated risk stratification, decision support systems can be developed for patients themselves, their family members, or clinicians who can review patients' health status and decide whether a health care service is needed.

Implications and Future Work

Theoretically, our study offers an example of a system that provides multidimension, multiscale, and multiprecision data for elderly health monitoring. The integrated use of such data from multiple sources can offer more reliable information as compared with single-source data [28,58]. Based on the data collected from the proposed system, we have developed methods for incorporating data from multiple sources for predictive modeling (eg, wellness prediction for community-dwelling elderly people) [28,59]. For elderly health monitoring, it is important to combine continuous health monitoring data with discrete demographic/medical data. The variables consist of outcomes from sensing devices that output a continuous stream of sensor information that are related to various activities. Demographics/medical data consist of only a few variables collected discretely over a fixed time period. Naïve integration may result in situations where high-dimensional data dominate, and simple data summaries may cause significant loss of key information. New methodologies will be needed to determine the scale and dimensionality for best performance under various predictive models.

Following Chow [60], we propose to utilize different data fusion techniques for integrating heterogeneous data in terms of dimension, scale, and precision by means of statistical modeling. First, it is necessary to process and reconstruct collected data on a unified coordinate or reference grid. Here, we propose adopting a kernel-based smoothing method [61]. Given a measurement taken from a specific source at a specific time, the measurement will be filtered over a user-defined space-time domain through a kernel regression function. Details of this method are provided by Chow [60], who applied the method to the fusion of road traffic data. Second, data from different sources could be combined using the voting technique [62,63], which is essentially a weighted linear combination of information from different sources, in which the weights are defined according to the accuracy or creditability of the associated data sources as determined in advance. Moving forward, we will use heterogeneous longitudinal methods with a dynamic risk adjustment scheme for monitoring individual risk, such as the DySS, RA-CUSUM, and RA-EWMA methods [64,65]. In addition, we will identify major vital signs that affect health conditions based on health monitoring and lifestyle data, and then develop modeling strategies that will take the vital signs features as input for state-of-the-art machine learning algorithms such as boosting, support vector machine, random forest, ensemble modeling, and neural networks for wellness forecasting. Correlations between adverse health outcomes and multiple risk factors (eg, poor gait and balance) will be investigated based on the collected health monitoring data.

Practically, our system is likely to be of interest to policymakers, elderly care centers, and health care providers, particularly given the urgent need to increase the capability to care for the elderly as the burden shifts from hospital care. The proposed system involves linking health care providers with their patients without spending unnecessary time on less productive aspects of community activities such as avoidable driving to and from communities and on-site measurements of vital signs to assess health condition. The system has the potential to detect significant changes in health condition and to flag these changes as more caregiver attention is required to keep older adults out of hospitals. In addition, when hospitalization is needed, such a system may help to automate risk stratification of patients, which may facilitate hospital resources allocation. Thus, it can be beneficial to improving the quality of health care services

provided, and ease the heavy burden on local health care systems. Moreover, end users may not initially accept and adopt a new technology after its introduction for a variety of reasons, and therefore will not experience the benefits. Our preliminary acceptance findings showed a relatively positive attitude for using our system (mean score=5.67) and slightly high levels of self-efficacy, perceived usefulness, perceived ease of use, and behavioral intention (with mean scores of 4-5). One possible reason could be that the participants may not have fully perceived the potential benefits of our system as the utilized system in the present study did not include any prediction algorithms for health management and timely communications between the system and the participants. Based on previous studies, there could be some other factors that significantly affect user acceptance of health information technology, such as social influence (eg, family/friends' opinions) [66,67], facilitating conditions (eg, organizational and technical infrastructure) [68], and technology anxiety (eg, anxiety in using technology) [69].

To help implement our system into practice, we will perform further longitudinal acceptance modeling studies on the full version of our system to focus on the factors that affect older adults' and health care professionals' acceptance [70]. Following that, targeted strategies (eg, community-based technology support services and training workshops) will be promoted to improve user acceptance on our smart system. In the long run, our proposed research is expected to develop effective ways to reduce the growing elderly care burden on health care systems.

Limitations

Missing data is a common problem encountered in most health care–related studies. Monitoring data quality in the presence of missing data is required, because the accuracy and reliability of measurements may be impaired when nonmedical experts perform the measurements [16]. One limitation of this study is the lack of data quality monitoring during implementation. Based on the data collected, we developed a data quality monitoring method to signal issues with the accuracy of the collected data quickly [71], which could be beneficial for further studies. Moreover, our study was based on a 3-month design and cannot evaluate the influences of season/time of the year on data variation. Future studies, in particular longitudinal studies of more than 6 months, are recommended to consider the evaluation of the effects of season/time of year.

Acknowledgments

This work was partly supported by the RGC Theme-Based Research Scheme (no. T32-102-14N), CityU Grant (no. 9610406), and the National Natural Science Foundation of China (no. 71901188). The authors wish to thank the HKCS for their support in participant recruitment and thank all of the participants of the pilot studies.

Authors' Contributions

Conceptualization: HW and YZ; Data collection: HW, LY, and JL; Data analysis: HW, LY, and IZ; Methodology: HW and YZ; Writing—original draft: HW; Writing—review and editing: HW, YZ, LY, and JC; Funding acquisition: KLT; Supervision: KLT.

Conflicts of Interest

None declared.

References

1. The Government of the Hong Kong Special Administrative Region. Hong Kong Population Projections 2017-2066. Census and Statistics Department. URL: https://www.statistics.gov.hk/pub/B1120015072017XXXXB0100.pdf [accessed 2020-09-22]
2. The Government of the Hong Kong Special Administrative Region. Elderly Health Service. Department of Health. URL: http://www.elderly.gov.hk/eindex.html [accessed 2020-09-22]
3. Anderson G, Horvath J. The growing burden of chronic disease in America. Public Health Rep 2004;119(3):263-270 [FREE Full text] [doi: 10.1016/j.phr.2004.04.005] [Medline: 15158105]
4. Tinetti ME. Clinical practice. Preventing falls in elderly persons. N Engl J Med 2003 Jan 02;348(1):42-49. [doi: 10.1056/NEJMcp020719] [Medline: 12510042]
5. Ho SC, Woo J, Chan SS, Yuen YK, Sham A. Risk factors for falls in the Chinese elderly population. J Gerontol A Biol Sci Med Sci 1996 Sep;51(5):M195-M198. [doi: 10.1093/gerona/51a.5.m195] [Medline: 8808988]
6. Chu LW, Pei CK, Chiu A, Liu K, Chu MM, Wong S, et al. Risk factors for falls in hospitalized older medical patients. J Gerontol A Biol Sci Med Sci 1999 Jan;54(1):M38-M43. [doi: 10.1093/gerona/54.1.m38] [Medline: 10026661]
7. Chu L, Chi I, Chiu A. Falls and fall-related injuries in community-dwelling elderly persons in Hong Kong: a study on risk factors, functional decline, and health services utilisation after falls. Hong Kong Med J 2007;13(1):S8-S12 [FREE Full text]
8. The Government of the Hong Kong Special Administrative Region. Services for the Elderly. Social Welfare Department. URL: https://www.swd.gov.hk/en/index/site_pubsvc/page_elderly/ [accessed 2020-09-22]
9. Celler BG, Sparks R, Nepal S, Alem L, Varnfield M, Li J, et al. Design of a multi-site multi-state clinical trial of home monitoring of chronic disease in the community in Australia. BMC Public Health 2014 Dec 15;14:1270 [FREE Full text] [doi: 10.1186/1471-2458-14-1270] [Medline: 25511206]

10. Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]
11. Lai AM, Kaufman DR, Starren J, Shea S. Evaluation of a remote training approach for teaching seniors to use a telehealth system. Int J Med Inform 2009 Nov;78(11):732-744 [FREE Full text] [doi: 10.1016/j.ijmedinf.2009.06.005] [Medline: 19620023]
12. Hebert MA, Korabek B, Scott RE. Moving research into practice: A decision framework for integrating home telehealth into chronic illness care. Int J Med Inform 2006 Dec;75(12):786-794. [doi: 10.1016/j.ijmedinf.2006.05.041] [Medline: 16872892]
13. Mabo P, Victor F, Bazin P, Ahres S, Babuty D, Da Costa A, COMPAS Trial Investigators. A randomized trial of long-term remote monitoring of pacemaker recipients (the COMPAS trial). Eur Heart J 2012 May;33(9):1105-1111 [FREE Full text] [doi: 10.1093/eurheartj/ehr419] [Medline: 22127418]
14. Crossley G, Boyle A, Vitense H, Sherfesee L, Mead RH. Trial design of the clinical evaluation of remote notification to reduce time to clinical decision: the Clinical evaluation Of remote NotificatioN to rEduCe Time to clinical decision (CONNECT) study. Am Heart J 2008 Nov;156(5):840-846. [doi: 10.1016/j.ahj.2008.06.028] [Medline: 19061696]
15. Brettle A, Brown T, Hardiker N, Radcliffe J, Smith C. Telehealth: The effects on clinical outcomes, cost effectiveness and the patient experience: A systematic overview of the literature. University of Salford. 2013. URL: http://usir.salford.ac.uk/id/eprint/29392/ [accessed 2020-09-22]
16. Celler BG, Sparks RS. Home telemonitoring of vital signs--technical challenges and future directions. IEEE J Biomed Health Inform 2015 Jan;19(1):82-91. [doi: 10.1109/JBHI.2014.2351413] [Medline: 25163076]
17. Kakria P, Tripathi NK, Kitipawang P. A Real-Time Health Monitoring System for Remote Cardiac Patients Using Smartphone and Wearable Sensors. Int J Telemed Appl 2015;2015:373474. [doi: 10.1155/2015/373474] [Medline: 26788055]
18. Etemadi M, Inan OT, Heller JA, Hersek S, Klein L, Roy S. A Wearable Patch to Enable Long-Term Monitoring of Environmental, Activity and Hemodynamics Variables. IEEE Trans Biomed Circuits Syst 2016 Apr;10(2):280-288 [FREE Full text] [doi: 10.1109/TBCAS.2015.2405480] [Medline: 25974943]
19. Or C, Tao D. A 3-Month Randomized Controlled Pilot Trial of a Patient-Centered, Computer-Based Self-Monitoring System for the Care of Type 2 Diabetes Mellitus and Hypertension. J Med Syst 2016 Apr;40(4):81. [doi: 10.1007/s10916-016-0437-1] [Medline: 26802011]
20. Megalingam R, Unnikrishnan U, Subash A, Pocklassery G, Thulasi A, Mourya G. Wearable medical devices in preventive health care: Cuffless blood pressure measurement. 2014 Presented at: Intelligent Computing, Communication and Devices; 2014; New Delhi, India. [doi: 10.1007/978-81-322-2012-1_80]
21. Spano E, Di Pascoli S, Iannaccone G. Low-Power Wearable ECG Monitoring System for Multiple-Patient Remote Monitoring. IEEE Sensors J 2016 Jul;16(13):5452-5462. [doi: 10.1109/jsen.2016.2564995]
22. Sparks R, Celler B, Okugami C, Jayasena R, Varnfield M. Telehealth monitoring of patients in the community. J Intell Syst 2016;25(1):37. [doi: 10.1515/jisys-2014-0123]
23. Melillo P, Orrico A, Scala P, Crispino F, Pecchia L. Cloud-Based Smart Health Monitoring System for Automatic Cardiovascular and Fall Risk Assessment in Hypertensive Patients. J Med Syst 2015 Oct;39(10):109. [doi: 10.1007/s10916-015-0294-3] [Medline: 26276015]
24. Doty T, Kellihan B, Jung T, Zao J, Litvan I. The Wearable Multimodal Monitoring System: A Platform to Study Falls and Near-Falls in the Real-World. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population. Los Angeles, CA, USA: Springer; Aug 2, 2015:412-422.
25. Chan W, Zhao Y, Tsui K. Implementation of electronic health monitoring systems at the community level in Hong Kong. In: Lecture Notes in Computer Science. 2017 Oct 28 Presented at: International Conference on Smart Health; 2017; Hong Kong p. 94-103. [doi: 10.1007/978-3-319-67964-8_9]
26. Ghayvat H, Liu J, Mukhopadhyay SC, Gui X. Wellness Sensor Networks: A Proposal and Implementation for Smart Home for Assisted Living. IEEE Sensors J 2015 Dec;15(12):7341-7348. [doi: 10.1109/jsen.2015.2475626]
27. Arshad A, Khan S, Alam A, Tasnim R, Boby R. Health and wellness monitoring of elderly people using intelligent sensing technique. 2016 Presented at: International Conference on Computer and Communication Engineering (ICCCE); 2016; Kuala Lumpur, Malaysia p. 26-27. [doi: 10.1109/iccce.2016.58]
28. Yu L, Chan WM, Zhao Y, Tsui K. Personalized Health Monitoring System of Elderly Wellness at the Community Level in Hong Kong. IEEE Access 2018;6:35558-35567. [doi: 10.1109/access.2018.2848936]
29. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012 Jun 21;344:e3874 [FREE Full text] [doi: 10.1136/bmj.e3874] [Medline: 22723612]
30. Redmond SJ, Xie Y, Chang D, Basilakis J, Lovell NH. Electrocardiogram signal quality measures for unsupervised telehealth environments. Physiol Meas 2012 Sep;33(9):1517-1533. [doi: 10.1088/0967-3334/33/9/1517] [Medline: 22903004]
31. Peyton L, Mouttham A, Ali K, Baarah A, Mouftah H. Real-time analytics and quality of care. In: Sturmberg JP, Martin CM, editors. Handbook of Systems and Complexity in Health. New York, NY: Springer New York; 2013:495-519.

32. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. J Big Data 2014;1(1):2. [doi: 10.1186/2196-1115-1-2]
33. Zhang Y, Fong S, Fiaidhi J, Mohammed S. Real-time clinical decision support system with data stream mining. J Biomed Biotechnol 2012;2012:580186. [doi: 10.1155/2012/580186] [Medline: 22851884]
34. Thommandram A, Pugh J, Eklund J, McGregor C, James A. Classifying neonatal spells using real-time temporal analysis of physiological data streams: Algorithm development. 2013 Presented at: IEEE Point-of-Care Healthcare Technologies (PHT); 2013; Bangalore, India p. 16-18. [doi: 10.1109/pht.2013.6461329]
35. Meharia P, Agrawal D. A hybrid key management scheme for healthcare sensor networks. 2016 Presented at: IEEE International Conference on Communications (ICC); 2016; Kuala Lumpur, Malaysia p. 22-27. [doi: 10.1109/icc.2016.7511484]
36. Rebolledo-Nandi Z, Chavez-Olivera A, Cuevas-Valencia R, Alarcon-Paredes A, Alonso G. Design of a versatile low cost mobile health care monitoring system using an android application. 2015 Presented at: Pan American Health Care Exchanges (PAHCE); March 23-28, 2015; Santiago, Vina del Mar, Chile. [doi: 10.1109/pahce.2015.7173334]
37. Howcroft J, Kofman J, Lemaire ED. Review of fall risk assessment in geriatric populations using inertial sensors. J Neuroeng Rehabil 2013 Aug 08;10(1):91 [FREE Full text] [doi: 10.1186/1743-0003-10-91] [Medline: 23927446]
38. Shany T, Redmond SJ, Marschollek M, Lovell NH. Assessing fall risk using wearable sensors: a practical discussion. A review of the practicalities and challenges associated with the use of wearable sensors for quantification of fall risk in older people. Z Gerontol Geriatr 2012 Dec;45(8):694-706. [doi: 10.1007/s00391-012-0407-2] [Medline: 23184295]
39. Shany T, Redmond SJ, Narayanan MR, Lovell NH. Sensors-Based Wearable Systems for Monitoring of Human Movement and Falls. IEEE Sensors J 2012 Mar;12(3):658-670. [doi: 10.1109/JSEN.2011.2146246]
40. Weiss A, Herman T, Plotnik M, Brozgol M, Maidan I, Giladi N, et al. Med Eng Phys 2010 Mar;32(2):119-125. [doi: 10.1016/j.medengphy.2009.10.015] [Medline: 19942472]
41. Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. J Am Geriatr Soc 2013 Feb;61(2):202-208. [doi: 10.1111/jgs.12106] [Medline: 23350947]
42. Fritz S, Lusardi M. White paper: "walking speed: the sixth vital sign". J Geriatr Phys Ther 2009;32(2):46-49. [Medline: 20039582]
43. Studenski S. Bradypedia: is gait speed ready for clinical use? J Nutr Health Aging 2009 Dec;13(10):878-880. [doi: 10.1007/s12603-009-0245-0] [Medline: 19924347]
44. Perry J, Garrett M, Gronley JK, Mulroy SJ. Classification of walking handicap in the stroke population. Stroke 1995 Jun;26(6):982-989. [doi: 10.1161/01.str.26.6.982] [Medline: 7762050]
45. Quach L, Galica A, Jones R, Procter-Gray E, Manor B, Hannan M, et al. The nonlinear relationship between gait speed and falls: the Maintenance of Balance, Independent Living, Intellect, and Zest in the Elderly of Boston Study. J Am Geriatr Soc 2011 Jun;59(6):1069-1073 [FREE Full text] [doi: 10.1111/j.1532-5415.2011.03408.x] [Medline: 21649615]
46. Chou C, Chien C, Hsueh I, Sheu C, Wang C, Hsieh C. Developing a short form of the Berg Balance Scale for people with stroke. Phys Ther 2006 Feb;86(2):195-204. [Medline: 16445333]
47. Kerr GK, Worringham CJ, Cole MH, Lacherez PF, Wood JM, Silburn PA. Predictors of future falls in Parkinson disease. Neurology 2010 Jul 13;75(2):116-124. [doi: 10.1212/WNL.0b013e3181e7b688] [Medline: 20574039]
48. Guimaraes RM, Isaacs B. Characteristics of the gait in old people who fall. Int Rehabil Med 1980;2(4):177-180. [doi: 10.3109/09638288009163984] [Medline: 7239777]
49. Maki BE. Gait changes in older adults: predictors of falls or indicators of fear. J Am Geriatr Soc 1997 Mar;45(3):313-320. [doi: 10.1111/j.1532-5415.1997.tb00946.x] [Medline: 9063277]
50. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
51. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. Sleep Med Rev 2016 Feb;25:52-73. [doi: 10.1016/j.smrv.2015.01.009] [Medline: 26163057]
52. Pilz LK, Keller LK, Lenssen D, Roenneberg T. Time to rethink sleep quality: PSQI scores reflect sleep quality on workdays. Sleep 2018 May 01;41(5):zsy029. [doi: 10.1093/sleep/zsy029] [Medline: 29420828]
53. Hong Kong Christian Service. Elderly Core Business. 2018. URL: http://www.hkcs.org/en/services/ecb [accessed 2020-09-22]
54. Taylor S, Todd PA. Understanding Information Technology Usage: A Test of Competing Models. Inf Syst Res 1995 Jun;6(2):144-176. [doi: 10.1287/isre.6.2.144]
55. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
56. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quart 2003;27(3):425-478. [doi: 10.2307/30036540]

57. Yu L, Zhao Y, Wang H, Sun TL, Murphy TE, Tsui KL. Assessing elderly's functional balance and mobility via analyzing data from waist-mounted tri-axial wearable accelerometers in time up and go tests. BMC Med Inform Decis Mak 2020:forthcoming.
58. Naddeo S, Verde L, Forastiere M, Pietro G, Sannino G. A real-time m-health monitoring system: An integrated solution combining the use of several wearable sensors and mobile devices. 2017 Presented at: The 10th International Joint Conference on Biomedical Engineering Systems and Technologies; February 21–23, 2017; Porto, Portugal p. 545-552. [doi: 10.5220/0006296105450552]
59. Fan X, Zhao Y, Wang H, Tsui KL. Forecasting one-day-forward wellness conditions for community-dwelling elderly with single lead short electrocardiogram signals. BMC Med Inform Decis Mak 2019 Dec 30;19(1):285 [FREE Full text] [doi: 10.1186/s12911-019-1012-8] [Medline: 31888608]
60. Chow A. Heterogeneous urban traffic data and their integration through kernel-based interpolation. J of Facilities Management 2016 May 03;14(2):165-178. [doi: 10.1108/jfm-08-2015-0025]
61. Camps-Valls G, Gomez-Chova L, Munoz-Mari J, Rojo-Alvarez J, Martinez-Ramon M. Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection. IEEE Trans Geosci Remote Sensing 2008 Jun;46(6):1822-1835. [doi: 10.1109/tgrs.2008.916201]
62. Olkin I. Meta-analysis: methods for combining independent studies. Stat Sci 1992;7(2):226. [doi: 10.17877/DE290R-12874]
63. Choi K, Chung Y. A Data Fusion Algorithm for Estimating Link Travel Time. J Intel Transport Syst 2010 Jun 18;7(3-4):235-260. [doi: 10.1080/714040818]
64. Qiu P, Xiang D. Surveillance of cardiovascular diseases using a multivariate dynamic screening system. Stat Med 2015 Jun 30;34(14):2204-2221. [doi: 10.1002/sim.6477] [Medline: 25757653]
65. Woodall WH, Fogel SL, Steiner SH. The Monitoring and Improvement of Surgical-Outcome Quality. J Qual Technol 2017 Nov 21;47(4):383-399. [doi: 10.1080/00224065.2015.11918141]
66. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD. Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model. Inf Syst Front 2017 Jun 8;21(3):719-734. [doi: 10.1007/s10796-017-9774-y]
67. Lunney A, Cunningham N, Eastin M. Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. Comput Human Behav 2016 Dec;65:114-120. [doi: 10.1016/j.chb.2016.08.007]
68. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157-178. [doi: 10.2307/41410412]
69. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
70. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
71. Mahmood T, Wittenberg P, Zwetsloot IM, Wang H, Tsui KL. Monitoring data quality for telehealth systems in the presence of missing data. Int J Med Inform 2019 Jun;126:156-163. [doi: 10.1016/j.ijmedinf.2019.03.011] [Medline: 31029257]

Abbreviations

3M-TUG: 3-meter timed up and go
10M-SW: 10-meter straight walking
DBP: diastolic blood pressure
SBP: systolic blood pressure
SpO_2: blood oxygen level

Edited by G Eysenbach; submitted 09.04.20; peer-reviewed by K Liu, X Li; comments to author 11.05.20; revised version received 21.05.20; accepted 25.06.20; published 30.09.20.

Please cite as:
Wang H, Zhao Y, Yu L, Liu J, Zwetsloot IM, Cabrera J, Tsui KL
A Personalized Health Monitoring System for Community-Dwelling Elderly People in Hong Kong: Design, Implementation, and Evaluation Study
J Med Internet Res 2020;22(9):e19223
URL: http://www.jmir.org/2020/9/e19223/
doi:10.2196/19223
PMID:32996887

Original Paper

Primary Care Pre-Visit Electronic Patient Questionnaire for Asthma: Uptake Analysis and Predictor Modeling

Andrew Kouri[1], MD, FRCPC; Janet Yamada[2], PhD, NP; Joanna E M Sale[3], MSc, PhD; Sharon E Straus[4], HBSc, MSc, MD, FRCPC; Samir Gupta[1], MD, MSc, FRCPC

[1]Division of Respirology, Department of Medicine, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

[2]Daphne Cockwell School of Nursing, Ryerson University, Toronto, ON, Canada

[3]Li Ka Sing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

[4]Department of Medicine, Division of Geriatrics, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

Corresponding Author:
Samir Gupta, MD, MSc, FRCPC
Division of Respirology, Department of Medicine
St. Michael's Hospital, Unity Health Toronto
Bond Wing, Suite 6042, 30 Bond St.
Toronto, ON, M5B 1W8
Canada
Phone: 1 416 864 6026
Email: samir.gupta@unityhealth.to

Abstract

Background: mHealth tablet-based interventions are increasingly being studied and deployed in various health care settings, yet little knowledge exists regarding patient uptake and acceptance or how patient demographics influence these important implementation metrics.

Objective: To determine which factors influence the uptake and successful completion of an mHealth tablet questionnaire by analyzing its implementation in a primary care setting.

Methods: We prospectively studied a patient-facing electronic touch-tablet asthma questionnaire deployed as part of the Electronic Asthma Management System. We describe tablet uptake and completion rates and corresponding predictor models for these behaviors.

Results: The tablet was offered to and accepted by patients in 891/1715 (52.0%) visits. Patients refused the tablet in 33.0% (439/1330) visits in which it was successfully offered. Patients aged older than 65 years of age (odds ratio [OR] 2.30, 95% CI 1.33-3.95) and with concurrent chronic obstructive pulmonary disease (OR 2.22, 95% CI 1.05-4.67) were more likely to refuse the tablet, and those on an asthma medication (OR 0.55, 95% CI 0.30-0.99) were less likely to refuse it. Once accepted, the questionnaire was completed in 784/891 (88.0%) instances, with those on an asthma medication (OR 0.53, 95% CI 0.32-0.88) being less likely to leave it incomplete.

Conclusions: Older age predicted initial tablet refusal but not tablet questionnaire completion, suggesting that perceptions of mHealth among older adults may negatively impact uptake, independent of usability. The influence of being on an asthma medication suggests that disease severity may also mediate mHealth acceptance. Although use of mHealth questionnaires is growing rapidly across health care settings and diseases, few studies describe their real-world acceptance and its predictors. Our results should be complemented by qualitative methods to identify barriers and enablers to uptake and may inform technological and implementation strategies to drive successful usage.

(J Med Internet Res 2020;22(9):e19358) doi:10.2196/19358

KEYWORDS

electronic questionnaire; tablet; mHealth uptake; asthma; modeling

Introduction

The use of information technology in health care has proliferated rapidly over the past 20 years. For example, patient-facing handheld touchscreen computers (tablets)—a form of mobile health (mHealth)—are increasingly being used and studied across a range of clinical settings [1-6]. Tablets are particularly well suited for collecting information directly from patients and have several potential advantages over traditional paper-based methods, including increased efficiency, facilitated data entry, greater data security, and ease of data transport [7,8]. Tablet-based patient questionnaires are also perceived favorably by patients, can promote greater patient-centered care, and are cost efficient [7,9]. Given these advantages, many clinicians and health care organizations are now integrating tablets into care. In Canada, 3500 physicians are already using electronic medical record–integrated tablet-based mHealth questionnaires, and the Canadian Medical Association has partnered with technology providers to further promote the use of tablets [10].

Patient uptake is critical to realizing the health system benefits of tablet-based questionnaires. However, despite broad interest and increasing use of tablets, there is limited research evaluating real-world mHealth tablet uptake and acceptability. Certain groups, such as older patients and those with lower educational attainment, may have disproportional barriers to accepting and using tablets [6,11] and a higher risk of experiencing technology-induced medical errors [12]. Accordingly, further research is required to understand the differences in uptake across demographics and the reasons for these differences, with a view toward improving implementation processes and technologies to enable broad acceptance.

We recently completed a prospective study of a chronic disease management tool called the Electronic Asthma Management System (eAMS) [13], which consisted of a patient-facing electronic touch-tablet asthma questionnaire for use in waiting rooms, that directly informed a provider-facing computerized clinical decision support system. Herein, we report on the implementation of this tablet-based questionnaire, including patient uptake, completion rates, and predictors of these metrics, in a real-world primary care setting.

Methods

Questionnaire Development

The eAMS questionnaire was created as a web-based app for Apple iPad and requires approximately 10 minutes to complete [14]. Questions were designed to (1) document current asthma control criteria, (2) identify current asthma medications used and adherence to these medications, and (3) collect information required to create a personalized self-management asthma action plan. Questionnaire data were transferred to a computerized clinical decision support system, which provided primary care clinicians with patients' asthma control status, evidence-based recommendations for medication adjustments, and a prepopulated asthma action plan, integrated into the electronic medical record in real time [15]. A questionnaire prototype was developed by asthma and knowledge translation experts in our research group by applying evidence-based electronic questionnaire design principles [16]. After pilot testing in a convenience sample of clinicians and researchers, the prototype's content and usability were iteratively refined using a rapid-cycle design process with serial focus groups (in total, 20 patients with asthma), followed by summative qualitative and quantitative analyses [14,15]. The questionnaire was designed and tested for use across a range of ages (with 30% of included participants aged older than 60 years of age), disease severities, and technological experience [14,15]. The system had a final System Usability Scale score of 84.2. The reported mean System Usability Scale score for web-based systems is 68, with a score of 84.2 corresponding to a subjective usability rating of excellent and a 95th percentile rank. The high System Usability Scale score was supported by favorable Likert-scale question responses and qualitative findings [14].

Study Design and Population

The eAMS patient questionnaire and clinical decision support system were tested in a 1-year prospective interrupted time-series design study across 3 primary care sites in Ontario, Canada [13]. Detailed questionnaire uptake and completion data were captured at 2 of these clinics. We included patients aged older than 16 years, with a physician diagnosis of asthma (these patients were identified according to an electronic medical record search algorithm validated in this population and with the same electronic medical records used in study sites [17]), and with an asthma medication prescription in the prior year (but not necessarily using asthma medication at the time of the study). We excluded patients who were pregnant (as the questionnaire was designed in part to inform creation of a physician-delivered asthma action plan, and there is limited evidence for asthma action plan use during pregnancy [13]), had cognitive or language difficulties, had a life expectancy of <1 year, or had been on a medication for chronic obstructive pulmonary disease in the prior year [17]. The study was approved by St. Michael's Hospital and Hamilton Integrated research ethics boards.

Questionnaire Implementation

Each day, an on-site research assistant at each site used a database query to identify all eligible patients who were booked to see a physician or nurse practitioner (excluding visits exclusively for administration of injections such as the flu vaccine). The research assistant was tasked with identifying patients for questionnaire completion in the clinic waiting room and inviting them to complete the questionnaire on 1 of 2 available tablet devices, prior to their appointment. Patients who had completed the questionnaire within the prior 28 days were not asked to repeat it.

Data Collection

Research assistants recorded tablet uptake and completion information for each eligible patient in a standardized Excel (Microsoft Inc) spreadsheet. This information included whether or not the tablet was given to the patient and the reason it was not given (if applicable) and whether the questionnaire was fully completed and reasons for noncompletion (if applicable, though patients were not required to provide a specific reason if they refused to accept the tablet or complete the questionnaire). We

also collected demographic and clinical data including age, sex, smoking status, asthma medication prescriptions, history of emergency room visits or hospitalizations for asthma (since 2003), the presence or absence of prior physician documentation of a diagnosis of asthma or chronic obstructive pulmonary disease, and whether the patient's asthma was currently under control through an electronic chart audit.

Outcomes

The primary outcomes were tablet questionnaire uptake—the proportion of visits in which the tablet questionnaire was offered to and accepted by patients—and tablet questionnaire completion—the proportion of visits in which the questionnaire was completed by patients when uptake was successful. Secondary outcomes were the reasons for failed uptake and the reasons for failed completion. Among cases where the tablet questionnaire was offered to patients, we also sought to determine if tablet refusal and questionnaire noncompletion were influenced by the following patient-level variables (established a priori, based on clinical relevance): age, sex, smoking status, asthma medication prescription, physician-documented diagnosis of asthma or chronic obstructive pulmonary disease, history of emergency room visit or hospitalization related to asthma, and current asthma control status. Finally, we evaluated the proportion of patients who accepted the tablet questionnaire at least once who were willing to accept it on a subsequent occasion, as well as predictors of accepting it on a subsequent occasion. We compared key outcomes between clinical sites.

Data Analysis

We present summary statistics for tablet uptake and completion rates (by patient visit), and reasons for uptake and completion failures. We used chi-square and Fisher exact tests to compare results between sites. We then constructed a logistic regression model to evaluate the influence of patient characteristics on the odds of tablet refusal and questionnaire noncompletion. In order to account for repeated measures at the patient level and clustering at the clinic level, generalized estimating equations were used [18]. Though the patient visits were not predetermined by the study investigators in this real-world study and follow-up was irregular, generalized estimating equations remain appropriate for analysis as the outcomes of interest were not clinical in nature, thus outcome-dependent follow-up was unlikely [19,20]. In order to ensure that this assumption was valid, another logistic regression model was constructed as a sensitivity analysis, modeling tablet refusal on only first patient visits, removing any potential residual confounding effects of repeated visits. Model performance was assessed by examining the concordance (C) statistic. At the patient level, we present summary statistics for the number of times each patient accepted the tablet questionnaire. We constructed a logistic regression model identifying predictors of refusing the tablet on the second occasion after acceptance on the first occasion. Additional available predictors for this patient-level analysis were time between completion of the first questionnaire and next time being offered it, and whether or not the questionnaire was fully completed on the first occasion it was accepted. All statistical analyses were performed using SAS (university edition; SAS Institute). A P value<.05 was statistically significant.

Results

Primary Outcome: Tablet Questionnaire Uptake

There were 612 eligible patients who made 1715 eligible visits during the study period (visits per patient: median 2, IQR 1-4) (Table 1). The tablet questionnaire was successfully offered and accepted in 891/1715 visits, resulting in an overall uptake of 52.0%. Failed uptake was due to the tablet not being offered in 385/1715 (22.5%) visits and not being accepted in 439/1330 (33.0%) visits in which it was offered. The tablet questionnaire was completed in 784/891 (88.0%) visits in which uptake (tablet offered and accepted) was successful (Table 1).

Table 1. Tablet uptake and completion.

	All, n (%)	Clinic 1, n (%)	Clinic 2, n (%)
Patients	612	437	175
Eligible visits	1715	1348	367
Tablet offered to patient	1715 (100)	1348 (100)	367 (100)
Yes	1330 (77.6)	1084 (80.4)	246 (67.0)
No	385 (22.5)	264 (19.6)	121 (33.0)
Tablet accepted by patient after being offered	1330 (100)	1084 (100)	246 (100)
Yes	891 (67.0)	699 (64.5)	192 (78.0)
No	439 (33.0)	385 (35.5)	54 (22.0)
Questionnaire completed after tablet offered and accepted by patient	891 (100)	699 (100)	192 (100)
Yes	784 (88.0)	618 (88.4)	166 (86.5)
No	107 (12.0)	81 (11.6)	26 (13.5)

Secondary Outcomes: Reasons for Tablet Uptake and Completion Failures

Both technology-related and logistical issues prevented the tablet from being offered to patients. Most patients who refused the tablet did not provide a reason. Once accepted, patients failed to complete the questionnaire for a variety of reasons, including aborting the questionnaire due to difficulties in completing it, technology-related failures, or clinicians calling patients in from the waiting room (Table 2).

Table 2. Reasons for failed tablet uptake and completion.

Reasons	All visits, n (%)	Clinic 1 visits, n (%)	Clinic 2 visits, n (%)	*P* value
Reasons tablet not offered to patients	385 (100)	264 (100)	121 (100)	<.001[a]
Patient missed by research assistant	307 (79.7)	215 (81.4)	92 (76.0)	
Technology-related issue[b]	48 (12.5)	38 (14.4)	10 (8.3)	
Both tablets already in use	12 (3.1)	7 (2.7)	5 (4.1)	
Not documented	18 (4.7)	4 (1.5)	14 (11.6)	
Reasons tablet not accepted by patients when offered	439 (100)	385 (100)	54 (100)	.70[c]
Patient refusal (denies asthma diagnosis)	54 (12.3)	46 (12.0)	8 (14.8)	
Patient refusal (indicates too difficult to use)	24 (5.5)	22 (5.7)	2 (3.7)	
Patient refusal (no reason provided)	351 (80.0)	307 (79.7)	44 (81.5)	
Caregiver refusal	10 (2.3)	10 (2.6)	0 (0.0)	
Reasons questionnaire not completed when tablet accepted	107 (100)	81 (100)	26 (100)	<.001[c]
Patient aborted (no reason provided)	46 (43.0)	45 (55.5)	1 (3.9)	
Patient aborted (indicates too difficult to complete)	17 (15.9)	8 (9.9)	9 (34.6)	
Patient aborted (started, then denied asthma diagnosis)	11 (10.3)	6 (7.4)	5 (19.2)	
Technology-related issue[b]	5 (4.7)	2 (2.5)	3 (11.5)	
Clinician aborted (called patient in prior to finishing)	28 (26.1)	20 (24.7)	8 (30.8)	

[a]Chi-square test was used.

[b]Technology-related issues included poor Wi-Fi or network connectivity and tablet software or hardware malfunctions.

[c]Fisher exact test was used.

Predictors of Tablet Questionnaire Refusal

Among the 1330 visits in which the tablet was successfully offered, we were able to match 1026 visits (77.1%) to patient characteristics for modeling. Among these 1026 visits, the tablet was accepted on 679 (66.2%) occasions and refused on 347 (33.8%) occasions. In the multivariable model, being 65 years and older was associated with the greatest odds of tablet questionnaire refusal (odds ratio [OR] 2.30, 95% CI 1.34-3.95). Having a previously documented diagnosis of chronic obstructive pulmonary disease was also associated with increased odds of refusal (OR 2.22, 95% CI 1.05-4.67). Conversely, a current prescription for an asthma medication was associated with lower odds of refusal (OR 0.55, 95% CI 0.30-0.99). The model demonstrated good predictive ability, with a C statistic of 0.67. All predictors are shown in Table 3. Patient asthma control level was missing in 40% of visits (a known care gap [21]) and was therefore excluded from the model.

In the multivariable model constructed with only the data from 451 first patient visits, the tablet was accepted on 407 (90.2%) occasions and refused on 44 (9.8%) occasions. In order to limit the potential for overspecification, only the covariates meeting or approaching statistical significance in the original model or those that were thought to be clinically important were included in the model, which included age ≥65 years, sex, current asthma medication prescription, physician-documented diagnosis of chronic obstructive pulmonary disease, and clinic site. In this sensitivity analysis, being 65 years of age or older was again associated with the greatest odds of tablet questionnaire refusal (OR 2.79, 95% CI 1.39-5.59), controlling for other confounders. Prescription of an asthma medication and physician diagnosis of chronic obstructive pulmonary disease also remained statistically significant predictors. Full results from the sensitivity analysis can be found in Multimedia Appendix 1.

Table 3. Predictors of tablet refusal.

Patient-level descriptors	Visits (n=1026), n (%)	Tablet refused (proportion of visits), n (%)	Odds ratio (95% CI)
Age			
<65 years old	750 (73.1)	208 (27.7)	reference
≥65 years old	276 (26.9)	139 (50.4)	2.30 (1.33-3.95)
Sex			
Male	263 (25.6)	81 (30.8)	reference
Female	763 (74.4)	266 (34.9)	1.18 (0.68-2.06)
Current asthma medication prescription			
No asthma medications	218 (21.3)	90 (41.3)	reference
Any asthma medication	808 (78.7)	257 (31.8)	0.55 (0.30-0.99)
Smoking status			
Lifelong nonsmoker	489 (47.7)	161 (32.9)	reference
Current or former smoker	432 (42.1)	136 (31.5)	0.84 (0.48-1.46)
Smoking history not documented	105 (10.2)	50 (47.6)	1.74 (0.79-3.83)
Physician-documented diagnosis of COPD[a]			
Absent	887 (86.5)	274 (30.9)	reference
Present	139 (13.5)	73 (52.5)	2.22 (1.05-4.67)
Physician-documented diagnosis of asthma			
Absent	310 (30.2)	106 (34.2)	reference
Present	716 (69.8)	241 (33.7)	1.13 (0.68-1.89)
Prior ER[b] visit or hospitalization for asthma			
No	951 (92.7)	330 (34.7)	reference
Yes	75 (7.3)	17 (22.7)	0.59 (0.15-2.32)
Clinic site			
Site 2	169 (16.5)	35 (20.7)	reference
Site 1	857 (83.5)	312 (36.4)	1.63 (0.79-3.35)

[a]COPD: chronic obstructive pulmonary disease.

[b]ER: emergency room.

Predictors of Failure to Complete Questionnaire After Accepting Tablet

Among the 891 visits in which patients accepted the tablet, we were able to match 706 visits (79.2%) to patient characteristics for modeling. After removing 27 visits where the questionnaire was not completed for reasons outside of the patient's control (such as the patient being called in prior to finishing or technology-related issues), 679 visits remained, among which the questionnaire was completed on 605 (89.1%) occasions and not completed on 74 (10.9%) occasions. In the multivariable model considering these visits, use of asthma medications was the only statistically significant predictor and was associated with lower odds of failing to complete the questionnaire (OR 0.53, 95% CI 0.32-0.88). The logistic model demonstrated adequate predictive ability with a C statistic of 0.64. All predictors are shown in Table 4. Asthma control level was missing in 40% of visits and was therefore excluded from the model.

Table 4. Predictors of failure to complete the questionnaire after tablet acceptance.

Patient-level descriptors	Visits (n=1026), n (%)	Tablet refused (proportion of visits), n (%)	Odds ratio (95% CI)
Age			
<65 years old	542 (79.8)	53 (9.8)	reference
≥65 years old	137 (20.2)	21 (15.3)	1.48 (0.88-2.49)
Sex			
Male	182 (26.8)	24 (13.2)	reference
Female	497 (73.2)	50 (10.1)	0.77 (0.45-1.31)
Current asthma medication prescription			
No asthma medications	128 (18.8)	22 (17.2)	reference
Any asthma medication	551 (81.2)	52 (9.4)	0.53 (0.32-0.88)
Smoking status			
Lifelong nonsmoker	328 (48.3)	35 (10.7)	reference
Current or former smoker	296 (43.6)	36 (12.2)	1.00 (0.60-1.67)
Smoking history not documented	55 (8.1)	2 (3.6)	0.48 (0.15-1.62)
Physician-documented diagnosis of COPD[a]			
Absent	613 (90.3)	62 (10.1)	reference
Present	66 (9.7)	12 (18.2)	1.90 (0.96-3.75)
Physician-documented diagnosis of asthma			
Absent	204 (30.0)	30 (14.7)	reference
Present	475 (70.0)	44 (9.3)	0.79 (0.47-1.34)
Prior ER[b] visit or hospitalization for asthma			
No	621 (91.5)	70 (11.3)	reference
Yes	58 (8.5)	4 (6.9)	0.66 (0.24-1.81)
Clinic site			
Site 2	134 (19.7)	15 (11.2)	reference
Site 1	545 (80.3)	59 (10.8)	0.88 (0.47-1.67)

[a]COPD: chronic obstructive pulmonary disease.

[b]ER: emergency room.

Willingness to Accept the Tablet on More Than One Visit

Of the 612 eligible patients in the study, 507 (82.8%) accepted the tablet at least once, 208 (34.0%) accepted the tablet at least twice, 84 (13.7%) accepted the tablet at least three times, and 45 (7.4%) accepted it four or more times.

Among the 507 patients who accepted the tablet at least once, 248 (48.9%) were offered the tablet a second time, among whom 208 (83.9%) accepted it.

In a multivariable model including age older than 65 years, sex, study site, time between accepting the first questionnaire and the next time it was offered, and whether or not the patient completed the questionnaire on the occasion in which it was first accepted, the only statistically significant predictor of refusing to accept the questionnaire a second time after accepting it the first time was noncompletion during the first attempt. Not having completed the questionnaire on the first instance had an odds ratio of 5.73 (95% CI 2.31-14.21) for refusing the questionnaire on the second eligible opportunity (for full details of model, see Multimedia Appendix 1).

Discussion

Principal Results

Our analysis revealed that a simple tablet-based questionnaire for use in waiting rooms was successfully offered and accepted by patients in only 52.0% (891/1715) and completed in only 45.7% (784/1715) of eligible visits in primary care over a period of 1 year. Although use of tablet-based questionnaires is growing rapidly in health care, few studies have explored their use in primary care settings, and knowledge about predictors of their real-world uptake is still limited. Although other investigators have reported on patient surveys of tablet acceptability [6,11,22-29], to our knowledge, ours is the first study to examine

tablet questionnaire uptake and completion rates in a primary care setting *and* to quantitatively identify predictors of uptake and successful completion.

The first barrier encountered to tablet questionnaire uptake was failure of the study site research assistants to offer the tablet to eligible patients. This accounted for 17.9% of instances of failed uptake (307/1715). The main reason for this was research assistants missing an eligible patient in the busy clinic environments. The number of missed patients differed between the 2 clinic sites, suggesting that clinic-specific factors such as patient flow, clinic staff engagement, and physical layout may impact successful tablet implementation. This finding aligns with the results of previous systematic reviews [30] of eHealth implementation in general, which have highlighted the importance of good fit between eHealth systems and clinic workflows. One strategy to mitigate this workflow barrier could be to tie tablet distribution to the patient registration process upon arrival, such that the same personnel member who registers the patient (eg, a clinic receptionist) distributes the tablet. However, identifying appropriate patients, distributing, and managing tablets, including collecting and disinfecting between patients [14], may not be feasible for a clinic receptionist. Overall, our findings support the notion that strategies to successfully deliver tablets need tailoring to each clinic environment and that prelaunch workflow analysis [31] and pilot testing may be beneficial.

Of note, technology issues such as Wi-Fi or network failures and hardware or software issues prevented tablet uptake in only 48/1715 (2.8%) cases and completion in only 5/891 (1.0%) cases (Table 2). Though technology failures are recognized barriers to eHealth implementation success [30], they were only minor contributors to failed uptake in our study. This trend is likely to continue as smartphone use becomes more ubiquitous [32], and electronic questionnaires can instead be offered on patients' personal devices.

The second major barrier to uptake was patient refusal to accept the tablet questionnaire, which occurred in 33.0% (439/1330) of visits (Table 1). Refusal levels did not differ between clinics, suggesting that this barrier is mediated mostly by patient-level factors (Table 2). In our multivariable model, patient age older than 65 years of age was the strongest predictor of tablet refusal (Table 3). This finding could be related to known decreased acceptance and comfort-levels with mobile technology among older adults [33,34] and may also reflect a decreased willingness among older adults to adopt newer paradigms of care in which patients are expected to play a more active role [35-37]. Previous researchers have shown somewhat conflicting results with regard to the influence of older age on tablet usage. Some authors have shown that increasing age is associated with decreased preference for tablet questionnaires over paper version [27], increased time needed to complete tablet questionnaires [38], and increased difficulty using tablets [6,11,22,28,29]. However, others have reported no significant influence of increasing age on tablet questionnaire usability, feasibility, and preference over paper versions [4,5,25,39]. Interestingly, being older than 65 years of age was not associated with failure to complete the tablet questionnaire in patients who accepted it. This finding may reflect the fact that our questionnaire development process included older patients (30% were 60 years of age or older) [14], and thus if the initial acceptance barrier was overcome, the electronic questionnaire had good usability across age ranges. These findings underscore the importance of including older individuals in the development process of mHealth tools but also that optimal tool usability across age ranges alone is not sufficient to drive uptake, and other implementation strategies to engage older individuals need to be considered as well.

Patients with a physician diagnosis of chronic obstructive pulmonary disease were also more likely to refuse the tablet questionnaire (Table 3). There is often a degree of overlap between asthma and chronic obstructive pulmonary disease in patients with airways disease, and some patients who experience this overlap may identify more as having chronic obstructive pulmonary disease than asthma [40,41]. It isn't clear currently if and how patients with asthma–chronic obstructive pulmonary disease overlap engage differently with mHealth interventions targeted at either condition, as this group is poorly studied, but our findings suggest that there may be barriers to mHealth acceptance targeted specifically at one of their codiagnoses.

The only statistically significant predictor in both models was having a prescription for an asthma medication, which was associated with lower odds of both tablet questionnaire refusal and noncompletion (refusal: OR 0.55, 95% CI 0.30-0.99; noncompletion: OR 0.53, 95% CI 0.32-0.88). The fact that patients using asthma medications were less likely to refuse and more likely to complete the tablet questionnaire may reflect a greater interest in and willingness to invest in disease management among patients with more active disease (ie therapy-meriting disease). It is also of note that a fair proportion of both tablet refusals (54/439, 12.3%) and failed questionnaire completions (11/107, 10.3%) were due to patients denying that they had asthma (Table 2). Although use of a probabilistic electronic medical record search algorithm may have incorrectly identified some patients as having asthma [17], it is also likely that some patients with asthma who are not actively taking asthma medications do not self-identify as currently having asthma. Although health status–based differences in mHealth uptake have not previously been reported, research into how an individual's health status influences desire for autonomy and preference for actively participating in care provides conflicting evidence across different patient populations and care settings [37,42,43]. In asthma in particular, there is recent evidence to suggest that there is an overall high level of desire among patients for autonomy and collaborative decision making regardless of disease severity [44,45]. However, how this relates to mHealth and tablet use is not clear and warrants further study.

The barriers and predictors discussed above illuminate important practical factors requiring consideration when implementing a tablet-based questionnaire. Many are likely relevant to any mHealth questionnaire, including smartphone questionnaires. Ideally, strategies to optimize usage of such tools will involve formal measurement of barriers and enablers and a theory-based approach to overcoming and leveraging these, respectively. To this end, we used complementary quantitative and qualitative research methods to specifically identify barriers and enablers to mHealth questionnaire uptake, through a Theoretical Domains

Framework analysis [46]. We will use these findings to match specific behavior-change strategies to the identified barriers and enablers, in order to inform system changes and implementation strategies that will drive usage.

Encouragingly, we found that when the tablet was successfully offered to and accepted by patients, overall, questionnaire completion occurred in 88% (784/891) of cases. This finding corresponds with our previous study's findings [14,15] of high questionnaire usability and further validates the importance of patient engagement and iterative questionnaire design with a focus on both content and usability. The unexpected finding that older patient age did not negatively influence tablet questionnaire completion underscores this importance further and supports our decision to deliberately include patients from a wide age range in our tablet questionnaire development process. This finding provides an important addition to the literature, as it implies that provided initial acceptance barriers can be overcome, optimizing mobile interface usability across patient demographics can lead to successful engagement in patient populations that are traditionally thought to be more difficult to reach (for example, older adults). This finding has implications for the design of future implementation strategies for mobile technology in older adults.

Finally, we assessed patients' willingness to complete tablet questionnaires on more than one occasion. The vast majority (208/248, 83.9%) of patients who accepted the tablet the first time accepted it again on the second opportunity. This suggests that once initial buy-in is achieved with patients, the majority see sufficient value in the process to continue to participate. The only significant predictor of refusing the tablet on the second occasion after having accepted it on the first was not having completed the tablet fully on the first attempt (OR 5.73, 95% CI 2.31-14.21). Programs considering routine use of tablet questionnaires should ensure that patients have adequate time and support to complete the questionnaire on the first attempt, as our findings suggest that a failed first attempt significantly deters future participation.

Limitations

Our study has some limitations. Although asthma patients were identified by a validated electronic medical record–based search algorithm [17] and were further required to have been prescribed an asthma medication in the prior year, some patients without asthma were likely erroneously identified due to imperfect algorithm specificity. Patients denied an asthma diagnosis in 65 visits, but this represented only 3.8% of the total 1715 visits and is unlikely to have influenced our main conclusions. We were unable to include about 20% of visits in our predictive models due to an inability to match those visits to patient characteristics. This resulted from the fact that patient characteristics were collected independently, and at a separate point in the study, whereby certain patients who were offered the questionnaire did not have their clinical characteristics recorded. This process can be assumed to have occurred at random; there was no difference in the tablet questionnaire acceptance level between patients who had full data and the whole cohort. We were also unable to obtain real-time qualitative patient feedback on tablet use during the study, as we designed our study to not interfere with clinic workflow as much as possible. However, we have completed an accompanying qualitative analysis of barriers and enablers to mobile health questionnaire acceptance and uptake in patients with asthma using this tablet questionnaire. It should also be noted that tablet questionnaires can be delivered via various operating systems and tablet sizes. Although we would not expect metrics related to offering the patients the tablet, or willingness of patients to accept the tablet to be affected by such differences, our findings regarding tablet completion rates may not be generalizable to other hardware and operating system types. Finally, as the denominator for our primary analysis was patient visits, there is the possibility that repeated measures from patients being offered the tablet on multiple occasions might have biased our analysis. To account for this, we used generalized estimating equations, which enable a population average model that adjusts for the correlated nature of repeated measurements. We also replicated our main results in a sensitivity analysis which included only a single visit for each patient.

Conclusion

In summary, we found that an electronic tablet-based asthma questionnaire implemented in a primary care waiting room setting had an overall uptake of only 52.0% (891/1715), even with involvement of dedicated research personnel. Patients refused to accept or to complete the tablet for various reasons. In particular, older age was an important predictor of tablet refusal, despite older patients having been included during tool development. Given that older adults, particularly those with chronic diseases that require close monitoring, likely stand to benefit the most from improvements in quality of care that may be realized through such mHealth questionnaires, our findings suggest that future research is required to specifically identify and address age-related barriers to successful uptake of such technology. The practical barriers and statistical predictors identified in our study may be relevant for electronic questionnaire usage across settings, diseases, and patient demographics. These factors, particularly if complemented by qualitative methods to identify barriers and enablers to uptake, can be used by clinics and programs employing such questionnaires, to target improvements to the intervention and its implementation strategy in order to drive usage.

Acknowledgments

This work was supported by the Keenan Research Summer Student Program, the Canadian Respiratory Research Network and the Canadian Institutes of Health Research. AK is supported by the Thorpe Family Fellowship in Respiratory Medicine at St. Michael's Hospital. JY is supported by the Lung Association Provincial Grant-in-Aid/National Grant Review competition. SG

is supported by the Michael Locke Term Chair in Knowledge Translation and Rare Lung Disease Research. Funders had no role in the design of the study; in the collection, analysis, or interpretation of data; or in manuscript preparation.

Conflicts of Interest

SG owns the intellectual property associated with eAMS and the eAMS questionnaire and would have an ownership interest in any commercial enterprise derived from this system. Other authors report no potential conflicts of interest.

Multimedia Appendix 1
Supplementary tables for regression analyses of tablet refusal at first patient visit only, and refusal of the tablet after having accepted it on first occasion.
[DOC File , 20 KB - jmir_v22i9e19358_app1.doc]

References

1. Freynhagen R, Baron R, Tölle T, Stemmler E, Gockel U, Stevens M, et al. Screening of neuropathic pain components in patients with chronic back pain associated with nerve root compression: a prospective observational pilot study (MIPORT). Curr Med Res Opin 2006 Mar;22(3):529-537. [doi: 10.1185/030079906X89874] [Medline: 16574036]
2. O'Leary KJ, Lohman ME, Culver E, Killarney A, Randy SG, Liebovitz DM. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2016 Jan;23(1):159-165. [doi: 10.1093/jamia/ocv058] [Medline: 26078412]
3. Richter JG, Becker A, Koch T, Nixdorf M, Willers R, Monser R, et al. Self-assessments of patients via Tablet PC in routine patient care: comparison with standardised paper questionnaires. Ann Rheum Dis 2008 Dec;67(12):1739-1741. [doi: 10.1136/ard.2008.090209] [Medline: 18647853]
4. Salaffi F, Gasparini S, Ciapetti A, Gutierrez M, Grassi W. Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. Rheumatology (Oxford) 2013 Nov;52(11):2062-2070 [FREE Full text] [doi: 10.1093/rheumatology/ket276] [Medline: 23955646]
5. Kim H, Park HC, Yoon SM, Kim TH, Kim J, Kang MK, et al. Evaluation of quality of life using a tablet PC-based survey in cancer patients treated with radiotherapy: a multi-institutional prospective randomized crossover comparison of paper and tablet PC-based questionnaires (KROG 12-01). Support Care Cancer 2016 Oct;24(10):4399-4406. [doi: 10.1007/s00520-016-3280-5] [Medline: 27220523]
6. Suzuki E, Mackenzie L, Sanson-Fisher R, Carey M, D'Este C, Asada H, et al. Acceptability of a Touch Screen Tablet Psychosocial Survey Administered to Radiation Therapy Patients in Japan. Int J Behav Med 2016 Aug;23(4):485-491. [doi: 10.1007/s12529-015-9502-2] [Medline: 26249725]
7. Abernethy AP, Herndon JE, Wheeler JL, Patwardhan M, Shaw H, Lyerly HK, et al. Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. Health Serv Res 2008 Dec;43(6):1975-1991 [FREE Full text] [doi: 10.1111/j.1475-6773.2008.00887.x] [Medline: 18761678]
8. Newell SM, Logan HL, Guo Y, Marks JG, Shepperd JA. Evaluating tablet computers as a survey tool in rural communities. J Rural Health 2015;31(1):108-117 [FREE Full text] [doi: 10.1111/jrh.12095] [Medline: 25243953]
9. Fritz F, Balhorn S, Riek M, Breil B, Dugas M. Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. Int J Med Inform 2012 May;81(5):303-313. [doi: 10.1016/j.ijmedinf.2011.12.008] [Medline: 22236957]
10. Joule Partners with CognisantMD. CognisantMD. URL: https://www.cognisantmd.com/joule-partners-cognisantmd/ [accessed 2019-07-04]
11. Zarghom S, Di Fonzo D, Leung F. Does Socioeconomic Status Affect Patients' Ease of Use of a Touch-Screen (iPad) Patient Survey? Interact J Med Res 2013 Jan 11;2(1):e1. [doi: 10.2196/ijmr.2314]
12. Kastner M, Lottridge D, Marquez C, Newton D, Straus SE. Usability evaluation of a clinical decision support tool for osteoporosis disease management. Implementation Sci 2010 Dec 10;5(1):96. [doi: 10.1186/1748-5908-5-96]
13. Gupta S, Price C, Agarwal G, Chan D, Goel S, Boulet L, et al. The Electronic Asthma Management System (eAMS) improves primary care asthma management. Eur Respir J 2019 Apr;53(4):1802241. [doi: 10.1183/13993003.02241-2018] [Medline: 30765503]
14. Lam Shin Cheung V, Kastner M, Sale JE, Straus S, Kaplan A, Boulet L, et al. Development process and patient usability preferences for a touch screen tablet–based questionnaire. Health Informatics J 2019 Jan 23:146045821882474. [doi: 10.1177/1460458218824749]
15. Gupta S, Lam Shin Cheung V, Kastner M, Straus S, Kaplan A, Boulet L, et al. Patient preferences for a touch screen tablet-based asthma questionnaire. J Asthma 2019 Jul;56(7):771-781. [doi: 10.1080/02770903.2018.1490750] [Medline: 29972087]
16. Palmblad M, Tiplady B. Electronic diaries and questionnaires: designing user interfaces that are easy for all patients to use. Qual Life Res 2004 Sep;13(7):1199-1207. [doi: 10.1023/B:QURE.0000037501.92374.e1] [Medline: 15473498]

17. Xi N, Wallace R, Agarwal G, Chan D, Gershon A, Gupta S. Identifying patients with asthma in primary care electronic medical record systems Chart analysis-based electronic algorithm validation study. Can Fam Physician 2015 Oct;61(10):e474-e483 [FREE Full text] [Medline: 26759847]
18. Hanley J, Negassa A, Edwardes MD, Forrester J. Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol 2003 Feb 15;157(4):364-375. [doi: 10.1093/aje/kwf215] [Medline: 12578807]
19. Pullenayegum EM, Lim LS. Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. Stat Methods Med Res 2016 Jul 11;25(6):2992-3014. [doi: 10.1177/0962280214536537]
20. Lin H, Scharfstein DO, Rosenheck RA. Analysis of longitudinal data with irregular, outcome-dependent follow-up. J Royal Statistical Soc B 2004 Aug;66(3):791-813. [doi: 10.1111/j.1467-9868.2004.b5543.x]
21. Price C, Agarwal G, Chan D, Goel S, Kaplan AG, Boulet L, et al. Large care gaps in primary care management of asthma: a longitudinal practice audit. BMJ Open 2019 Jan 29;9(1):e022506 [FREE Full text] [doi: 10.1136/bmjopen-2018-022506] [Medline: 30696669]
22. Miller Jr DP, Weaver KE, Case LD, Babcock D, Lawler D, Denizard-Thompson N, et al. Usability of a Novel Mobile Health iPad App by Vulnerable Populations. JMIR Mhealth Uhealth 2017 Apr 11;5(4):e43. [doi: 10.2196/mhealth.7268]
23. Hobden B, Bryant J, Carey M, Sanson-Fisher R, Oldmeadow C. Computer tablet or telephone? A randomised controlled trial exploring two methods of collecting data from drug and alcohol outpatients. Addictive Behaviors 2017 Aug;71:111-117. [doi: 10.1016/j.addbeh.2017.03.009]
24. Brinker TJ, Brieske CM, Esser S, Klode J, Mons U, Batra A, et al. A Face-Aging App for Smoking Cessation in a Waiting Room Setting: Pilot Study in an HIV Outpatient Clinic. J Med Internet Res 2018 Aug 15;20(8):e10976. [doi: 10.2196/10976]
25. Moura LM, Schwamm E, Moura Junior V, Seitz MP, Hsu J, Cole AJ, et al. Feasibility of the collection of patient-reported outcomes in an ambulatory neurology clinic. Neurology 2016 Nov 04;87(23):2435-2442. [doi: 10.1212/wnl.0000000000003409]
26. Carey M, Noble N, Mansfield E, Waller A, Henskens F, Sanson-Fisher R. The Role of eHealth in Optimizing Preventive Care in the Primary Care Setting. J Med Internet Res 2015 May;17(5):e126 [FREE Full text] [doi: 10.2196/jmir.3817] [Medline: 26001983]
27. Barentsz MW, Wessels H, van Diest PJ, Pijnappel RM, Haaring C, van der Pol CC, et al. Tablet, Web-Based, or Paper Questionnaires for Measuring Anxiety in Patients Suspected of Breast Cancer: Patients' Preferences and Quality of Collected Data. J Med Internet Res 2014 Oct 31;16(10):e239. [doi: 10.2196/jmir.3578]
28. Hess R, Santucci A, McTigue K, Fischer G, Kapoor W. Patient difficulty using tablet computers to screen in primary care. J Gen Intern Med 2008 Apr;23(4):476-480 [FREE Full text] [doi: 10.1007/s11606-007-0500-1] [Medline: 18373148]
29. Aiello EJ, Taplin S, Reid R, Hobbs M, Seger D, Kamel H, et al. In a randomized controlled trial, patients preferred electronic data collection of breast cancer risk-factor information in a mammography setting. J Clin Epidemiol 2006 Jan;59(1):77-81. [doi: 10.1016/j.jclinepi.2005.07.007] [Medline: 16360564]
30. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
31. Sheehan B, Bakken S. Approaches to workflow analysis in healthcare settings. NI 2012 (2012) 2012;2012:371 [FREE Full text] [Medline: 24199123]
32. The Internet and Digital Technology Internet. Statistics Canada. 2019 Jul 04. URL: https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2017032-eng.htm [accessed 2019-07-04]
33. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
34. Harrington CN, Ruzic L, Sanford JA. Universally Accessible mHealth Apps for Older Adults: Towards Increasing Adoption and Sustained Engagement. In: Antona M, Stephanidis C, editors. Universal Access in Human–Computer Interaction. Human and Technological Environments. UAHCI 2017. Lecture Notes in Computer Science, vol 10279. Cham: Springer; 2017.
35. Bastiaens H, Van Royen P, Pavlic DR, Raposo V, Baker R. Older people's preferences for involvement in their own care: A qualitative study in primary health care in 11 European countries. Patient Education and Counseling 2007 Sep;68(1):33-42. [doi: 10.1016/j.pec.2007.03.025]
36. Belcher VN, Fried TR, Agostini JV, Tinetti ME. Views of older adults on patient participation in medication-related decision making. J Gen Intern Med 2006 Apr;21(4):298-303. [doi: 10.1111/j.1525-1497.2006.00329.x]
37. Levinson W, Kao A, Kuby A, Thisted RA. Not all patients want to participate in decision making. A national study of public preferences. J Gen Intern Med 2005 Jun;20(6):531-535 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.04101.x] [Medline: 15987329]
38. Pollom EL, Wang E, Bui TT, Ognibene G, von Eyben R, Divi V, et al. A prospective study of electronic quality of life assessment using tablet devices during and after treatment of head and neck cancers. Oral Oncology 2015 Dec;51(12):1132-1137. [doi: 10.1016/j.oraloncology.2015.10.003]

39. McCleary NJ, Wigler D, Berry D, Sato K, Abrams T, Chan J, et al. Feasibility of computer-based self-administered cancer-specific geriatric assessment in older patients with gastrointestinal malignancy. Oncologist 2013;18(1):64-72 [FREE Full text] [doi: 10.1634/theoncologist.2012-0241] [Medline: 23287880]
40. Gibson PG, Simpson JL. The overlap syndrome of asthma and COPD: what are its features and how important is it? Thorax 2009 Aug;64(8):728-735. [doi: 10.1136/thx.2008.108027] [Medline: 19638566]
41. Zeki A, Schivo M, Chan A, Albertson T, Louie S. The Asthma-COPD Overlap Syndrome: A Common Clinical Problem in the Elderly. J Allergy (Cairo) 2011;2011:861926 [FREE Full text] [doi: 10.1155/2011/861926] [Medline: 22121384]
42. Cullati S, Courvoisier DS, Charvet-Bérard AI, Perneger TV. Desire for autonomy in health care decisions: A general population survey. Patient Education and Counseling 2011 Apr;83(1):134-138. [doi: 10.1016/j.pec.2010.04.025]
43. Chung GS, Lawrence RE, Curlin FA, Arora V, Meltzer DO. Predictors of hospitalised patients' preferences for physician-directed medical decision-making. J Med Ethics 2012 Feb;38(2):77-82. [doi: 10.1136/jme.2010.040618] [Medline: 21697296]
44. Caress A, Beaver K, Luker K, Campbell M, Woodcock A. Involvement in treatment decisions: what do adults with asthma want and what do they get? Results of a cross sectional survey. Thorax 2005 Mar;60(3):199-205 [FREE Full text] [doi: 10.1136/thx.2004.029041] [Medline: 15741435]
45. Adams RJ, Smith BJ, Ruffin RE. Patient preferences for autonomy in decision making in asthma management. Thorax 2001 Feb;56(2):126-132 [FREE Full text] [Medline: 11209101]
46. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implementation Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9]

Abbreviations

eAMS: Electronic Asthma Management System
mHealth: mobile health
OR: odds ratio

Edited by G Eysenbach; submitted 21.04.20; peer-reviewed by D Dalan, D Ryan; comments to author 29.06.20; revised version received 22.07.20; accepted 26.07.20; published 18.09.20.

Please cite as:
Kouri A, Yamada J, Sale JEM, Straus SE, Gupta S
Primary Care Pre-Visit Electronic Patient Questionnaire for Asthma: Uptake Analysis and Predictor Modeling
J Med Internet Res 2020;22(9):e19358
URL: http://www.jmir.org/2020/9/e19358/
doi:10.2196/19358
PMID:32945779

Original Paper

Sociodemographic Representativeness in a Nationwide Web-Based Survey of the View of Men on Involvement in Health Care Decision-Making: Cross-Sectional Questionnaire Study

Søren F Birkeland[1], LLM, MD, PhD; Anders K Haakonsson[1], MSc, PhD; Susanne S Pedersen[2,3], MSc, PhD; Nina Rottmann[2,4], MSc, PhD; Michael J Barry[5], MD; Sören Möller[1], MSc, PhD

[1]Open Patient Data Explorative Network (OPEN), Odense University Hospital and Department of Clinical Medicine, University of Southern Denmark, Odense, Denmark

[2]Department of Psychology, University of Southern Denmark, Odense, Denmark

[3]Department of Cardiology, Odense University Hospital, Odense, Denmark

[4]REHPA, The Danish Knowledge Centre for Rehabilitation and Palliative Care, Odense University Hospital and Department of Clinical Medicine, University of Southern Denmark, Nyborg, Denmark

[5]MGH Division of General Internal Medicine, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Søren F Birkeland, LLM, MD, PhD
Open Patient Data Explorative Network (OPEN)
Odense University Hospital and Department of Clinical Medicine
University of Southern Denmark
J. B. Winsløws Vej 9 a, 3. floor
Odense, DK-5000
Denmark
Phone: 45 29166730
Email: sbirkeland@health.sdu.dk

Abstract

Background: Being able to generalize research findings to a broader population outside of the study sample is an important goal in surveys on the internet. We conducted a nationwide, cross-sectional, web-based survey with vignettes illustrating different levels of patient involvement to investigate men's preferences regarding participation in health care decision-making. Following randomization into vignette variants, we distributed the survey among men aged 45 to 70 years through the state-authorized digital mailbox provided by the Danish authorities for secure communication with citizens.

Objective: This study aimed to investigate the sociodemographic representativeness of our sample of men obtained in a nationwide web-based survey using the digital mailbox.

Methods: Response rate estimates were established, and comparisons were made between responders and nonresponders in terms of age profiles (eg, average age) and municipality-level information on sociodemographic characteristics.

Results: Among 22,288 men invited during two waves, a total of 6756 (30.31%) participants responded to the survey. In adjusted analyses, responders' characteristics mostly resembled those of nonresponders. Response rates, however, were significantly higher in older men (odds ratio [OR] 2.83 for responses among those aged 65-70 years compared with those aged 45-49 years, 95% CI 2.58-3.11; $P<.001$) and in rural areas (OR 1.10 compared with urban areas, 95% CI 1.03-1.18; $P=.005$). Furthermore, response rates appeared lower in areas with a higher tax base (OR 0.89 in the highest tertile, 95% CI 0.81-0.98; $P=.02$).

Conclusions: Overall, the general population of men aged 45 to 70 years was represented very well by the responders to our web-based survey. However, the imbalances identified highlight the importance of supplementing survey findings with studies of the representativeness of other characteristics of the sample like trait and preference features, so that proper statistical corrections can be made in upcoming analyses of survey responses whenever needed.

(J Med Internet Res 2020;22(9):e19517) doi:10.2196/19517

KEYWORDS

research methodology; electronic data capture; internet-based survey; representativeness; generalizability; user involvement; patient satisfaction; bioethics; medical law; cancer

Introduction

Research has suggested that communication breaches are underlying issues in many complaints about health care delivery [1-3]. Beckman et al found that patients' feelings of being deserted and poor information delivery were central themes in malpractice suits, and later research pointed in the same direction [2,3]. Insufficient patient involvement in decision-making has therefore been proposed as an important underlying issue when people file a malpractice complaint [4-6]. People's lack of "ownership" of decisions about the care they have received may very well be an underlying cause in many complaints, and perhaps this is particularly the case if treatment leads to an undesirable outcome. Nonetheless, even if greater patient involvement seems to be an obvious starting point to increase patient satisfaction with health care and prevent complaints, there is still scant research to support this notion.

We therefore conducted a cross-sectional questionnaire study to increase our understanding of patient preferences for involvement in health care decision-making. The study was designed as a nationwide internet survey in the general population benefiting from the opportunities in Denmark for survey distribution through a web portal used for communication between authorities and citizens. As is the case with other survey approaches, however, using the internet for data collection may raise concerns about nonresponse bias and sample representativeness [7]. A representative survey sample can be defined as "one that has strong external validity in relationship to the target population the sample is meant to represent" [8]. This implies that results from the survey analyses can be generalized with confidence to the population of interest. Correspondingly, nonresponse and poor coverage of the sample may bias survey findings if the responding sample differs from the characteristics of the target population in nonnegligible ways [8]. For example, responders may more often be better off economically than the target population, thereby lowering the sample's representativeness [8]. Similarly, response rates (RRs) may vary with age and location (eg, areas with various ethnic populations) [9]. In this paper, we report on the representativeness of our web-based survey in terms of *sociodemographic characteristics* through comparisons of our sample with nonresponders and national statistics data.

Methods

Setting and Participants

We used a web platform (REDCap [10]) for the survey and identified the sample with civil registration numbers through use of the Danish Health Data Authority (DHDA). Danish citizens are all registered in the civil registration system with unique personal identification numbers. We recruited participants using personal invitations delivered to the participants through coupling of civil registration numbers to the "digital mailbox" provided by the Danish authorities for safe communication with citizens. Adult Danish citizens are registered to use the digital mailbox as default, although a small proportion of citizens have actively deregistered (9.9% in 2017) [9]. Digital mailbox communication is encrypted, and thus, its security is higher than that of usual email and mails sent by regular postal service [9].

Variables and Measurement

The survey illustrated various levels of patient involvement in health care decision-making through use of multiple case vignette versions. We used decision-making with regard to having a prostate-specific antigen (PSA) test for prostate cancer (PCa) screening as a model situation and measured responders' imagined satisfaction with health care and readiness to initiate malpractice litigation about the health care received. Responders were randomized into one of 30 different scenarios with an identical core structure. There were differences regarding the degree of patient involvement (five levels), the decision to have a PSA test, and outcomes (three possibilities; details have been provided previously [11]). Measures comprised standardized validated instruments (eg, personality) and purpose-designed questions (eg, sociodemographic characteristics). Regarding participant age, we chose the age range of 45 to 70 years with reference to international guidelines about PCa screening [12,13]. During survey development, we collaborated with male adult health user representatives to optimize the survey's content and acceptability.

We examined the overall representativeness of the sample obtained through the procedures described above. We estimated RRs and compared responders and nonresponders with respect to their age profile. Furthermore, we made comparisons regarding municipality-level sociodemographic characteristics derived from the 98 municipality codes available from the Danish municipality statistics database [14]. Data were used as standard measures for the state and municipalities in Denmark, as well as for research purposes, and included statistics information of municipality-level population density, tax per citizen figures, proportion of citizens with higher education, and proportion of citizens of non-Western origin [15,16]. Reporting in this article follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational studies [17].

Study Size, Quantitative Variables, and Statistical Methods

For the project, we drew a random sample of 12,000 male Danish health care users in the age range of 45 to 70 years, according to the following sample size estimation. When the primary outcome measure (readiness to complain) was targeted, 100 participants per subgroup were needed to obtain a 0.90 power to detect a 0.45 SD ("medium") effect between groups with a 5% risk of type 1 error and a bidirectional two-sample homoscedastic *t* test [18,19]. To compensate for nonresponse and subgroup skewness, we intended to include an additional 300 participants per group, thereby totaling 12,000 participant

invitations (400 per group and 30 groups) [20]. In order to anticipate the possibility of even lower participation, we obtained DHDA permission for sending up to 36,000 invitations in total (through up to three consecutive waves). Ultimately, two waves of surveys were required to achieve the necessary sample size.

We present RRs as counts and proportions, with stratification into 5-year age groups, survey wave groups (the person was part of which of the two survey waves), and municipality types (four categories) [21]. We compared the proportions between the groups by logistic regression (both unadjusted and mutually adjusted) and compared the participants with the general population by logistic regression with weights according to population size, using data from Statistics Denmark [22]. Comprehensive digital mailbox systems like the Danish system are not in place in all countries. Therefore, to get an impression of the general representativeness of our sample, we compared responders to all nonresponders in "A comparisons" without consideration for their opportunity to respond (having the digital mailbox). In "B comparisons," we made comparisons between responders and nonresponders among those having the digital mailbox only in order to accentuate any active decision on whether to participate. Conducting both comparisons in parallel is necessary to allow for taking into account digital mailbox noncoverage and to rule out that we only obtained responses from a highly selected group of the population through our web-survey solution. Odds ratios (ORs) denote the probability of response (ie, OR >1 means more likely to respond). Analyses were carried out using Stata version 15.0 (Stata Corp LP).

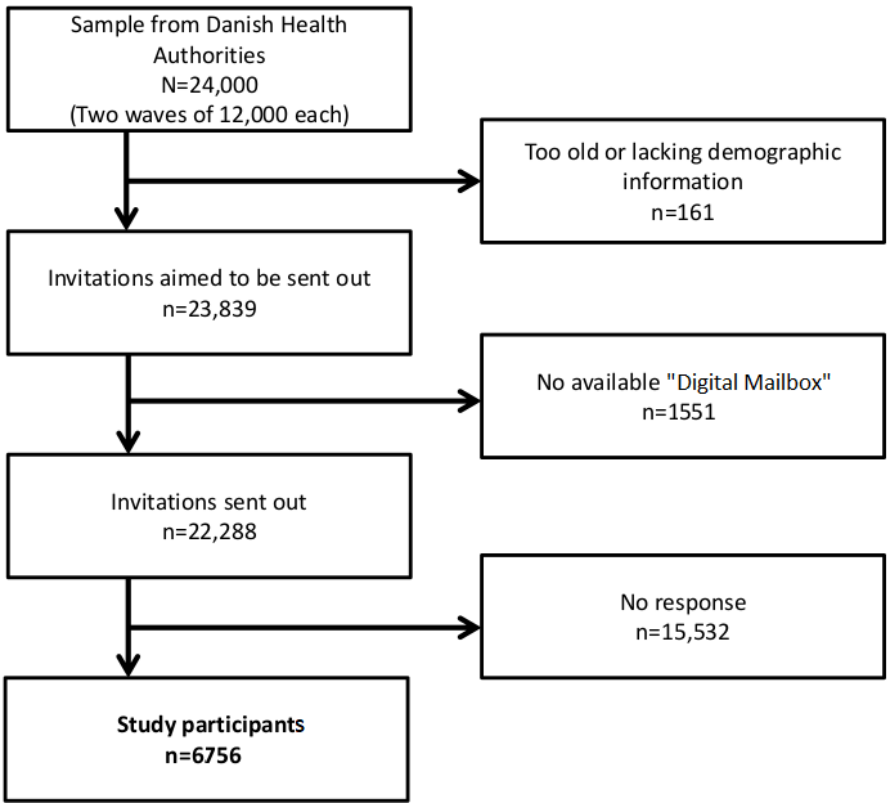
Ethics Approval and Consent to Participate

When conducting web-based surveys, ethical issues may arise with regard to procuring valid informed consent and respecting responders' privacy [7]. The Regional Scientific Ethics Committee for Southern Denmark evaluated the project and concluded that the project could be implemented without their permission (case number 20182000-99). However, after ensuring that data management was in compliance with EU regulation 2016/679 and Directive 95/46/EC, General Data Protection Regulation, we obtained Data Protection Agency authorization through the regional municipality and DHDA permission to conduct the survey (number FSEID-00003692). Invitation letters to potential participants explained the purpose of the study and were distributed with a link to the questionnaire. Considering a person's privilege to decline participation in our study, in addition to the introductory invitation letter, we considered that it was appropriate to send only one reminder after 14 days to augment the RR.

Results

Participants were invited on January 24, 2019, and one reminder was sent after 14 days. To obtain a satisfactory sample size (n=100) in all (n=30) questionnaire variants, two consecutive waves of invitations were necessary. The second wave of invitations was sent out on March 07, 2019, with a reminder after 14 days. We thereby obtained a sample of 6756 participants. The flow chart is presented in Figure 1, and sample characteristics are presented in Table 1.

Figure 1. Flowchart showing the inclusion of study participants.



Among 161 excluded individuals, 117 were excluded because they either were too old or had unknown age according to register information from DHDA, 6 were excluded as they explicitly asked to be deleted from the research project, and 38

were excluded because they were too old when they responded to the survey. The overall RR was 30.31% (6756/22,288) among digital mailbox invitations and 28.34% (6756/23,839) in the general population. It appears from our study that the digital mailbox covered 93.49% (22,288/23,839) of our target population.

Table 1. Sample characteristics.

Characteristic	Responders[a] (N=6756), median (IQR) or n	Nonresponders (N=15,532), median (IQR) or n	No digital mailbox (N=1551), median (IQR) or n
Age (years)	59 (53-65)	55 (50-62)	60 (54-66)
45-49 (N=4721)	935	3633	153
50-54 (N=5256)	1225	3751	280
55-59 (N=4804)	1368	3147	289
60-64 (N=4276)	1406	2515	355
65-70 (N=4782)	1822	2486	474
Municipality type			
Urban municipality (N=10,564)	2828	7049	687
Rural municipality (N=6962)	2042	4443	477
Urban-rural municipality (N=3947)	1183	2525	239
Outskirts municipality (N=2366)	703	1515	148
Invitation wave			
Wave 1 (N=11,869)	3395	7716	758
Wave 2 (N=11,970)	3361	7816	793

[a]Responder numbers and response rates regarding the 30 different questionnaire variants of the survey are shown in Multimedia Appendix 1.

In Tables 2 and 3, responders and nonresponders are compared. It appeared that older men and men from rural areas had higher RRs.

Table 2. Comparison of responders and nonresponders according to age and dwelling.

Characteristic	Comparison A[a] (N=23,839), OR (95% CI)[b]	Comparison A, *P* value	Comparison B[c] (N=22,288), OR (95% CI)[b]	Comparison B, *P* value
Age (years)		<.001		<.001
45-49	1 (reference)	N/A[d]	1 (reference)	N/A
50-54	1.12 (1.18-1.35)	<.001	1.27 (1.15-1.40)	<.001
55-59	1.61 (1.47-1.77)	<.001	1.69 (1.53-1.86)	<.001
60-64	1.98 (1.80-2.18)	<.001	2.17 (1.97-2.39)	<.001
65-70	2.49 (2.27-2.73)	<.001	2.85 (2.59-3.13)	<.001
Municipality type		<.001		<.001
Urban municipality	1 (reference)	N/A	1 (reference)	N/A
Rural municipality	1.14 (1.06-1.21)	<.001	1.15 (1.07-1.23)	<.001
Urban-rural municipality	1.17 (1.08-1.27)	<.001	1.17 (1.08-1.27)	<.001
Outskirts municipality	1.16 (1.05-1.28)	.004	1.16 (1.05-1.28)	.004
Invitation wave		N/A		N/A
Wave 1	1 (reference)	N/A	1 (reference)	N/A
Wave 2	1.03 (0.97-1.09)	.37	1.02 (0.97-1.08)	.43

[a]Comparison A compares responders to all nonresponders with or without the digital mailbox.

[b]Odds ratios (ORs) denote the probability of response (ie, OR >1 means more likely to respond).

[c]Comparison B compares responders to nonresponders with the digital mailbox.

[d]N/A: not applicable.

Table 3. Comparison of responders and nonresponders according to age and dwelling adjusted for age, municipality, and wave.

Characteristic	Comparison A[a] (N=23,839), OR (95% CI)[b]	Comparison A, *P* value	Comparison B[c] (N=22,288), OR (95% CI)[b]	Comparison B, *P* value
Age (years)		N/A[d]		N/A
45-49	1 (reference)	N/A	1 (reference)	N/A
50-54	1.23 (1.12-1.35)	<.001	1.27 (1.15-1.40)	<.001
55-59	1.61 (1.46-1.77)	<.001	1.69 (1.53-1.86)	<.001
60-64	1.97 (1.79-2.17)	<.001	2.16 (1.96-2.38)	<.001
65-70	2.48 (2.26-2.72)	<.001	2.83 (2.58-3.11)	<.001
Municipality type		N/A		N/A
Urban municipality	1 (reference)	N/A	1 (reference)	N/A
Rural municipality	1.10 (1.03-1.18)	.006	1.10 (1.03-1.18)	.005
Urban-rural municipality	1.16 (1.07-1.26)	<.001	1.15 (1.06-1.25)	.001
Outskirts municipality	1.10 (0.99-1.21)	.07	1.08 (0.98-1.20)	.12
Invitation wave		N/A		N/A
Wave 1	1 (reference)	N/A	1 (reference)	N/A
Wave 2	1.02 (0.96-1.08)	.55	1.02 (0.96-1.08)	.59

[a]Comparison A compares responders to all nonresponders with or without the digital mailbox.

[b]Odds ratios (ORs) denote the probability of response (ie, OR >1 means more likely to respond).

[c]Comparison B compares responders to nonresponders with the digital mailbox.

[d]N/A: not applicable.

Table 4 illustrates the statistical differences between responders and nonresponders across questionnaire variants. One would expect a tendency toward lower RRs with increasing length of the questionnaire, that is, with higher group number from 1 through 10 and from main variant A through C. However, no clear association could be demonstrated.

As a proxy for the amount of resources required to complete the survey, we measured the time used by responders across variants (Multimedia Appendix 2). It is worth noting that no relevant difference could be established regarding response time across questionnaire variants. Correspondingly, when looking at RRs, no clear relevant association appeared between questionnaire variant and RR. Although “Group 10” had a notably lower RR than did the other groups, “Group 9” that had a similar length demonstrated no decrease in the RR.

Table 5 compares responders with the entire Danish population of men aged 45 to 70 years. It appears that compared with the general population, younger men and urban men are underrepresented in the sample.

Table 4. Comparison of responders and nonresponders regarding questionnaire variants.

Variants	Comparison B[a] (N=22,288), OR (95% CI)[b]	Comparison B, *P* value	Adjusted comparison B[a] (N=22,288), OR (95% CI)[b,c]	Adjusted comparison B, *P* value
Randomization and vignette type		.004		N/A[d]
Number and vignette characteristics		.02		N/A
No patient participation[e]				
1	1 (reference)	N/A	1 (reference)	N/A
2	1.10 (0.97-1.25)	.13	1.10 (0.97-1.25)	.15
Various means of informing patient and patient participation in decision making[f]				
3	1.07 (0.94-1.21)	.32	1.06 (0.93-1.21)	.36
4	1.03 (0.91-1.17)	.62	1.04 (0.91-1.19)	.55
5	1.02 (0.90-1.16)	.77	1.01 (0.89-1.15)	.85
6	1.00 (0.88-1.13)	.97	0.99 (0.87-1.13)	.87
7	0.97 (0.85-1.11)	.66	0.96 (0.84-1.10)	.56
8	0.99 (0.87-1.12)	.85	0.99 (0.87-1.12)	.84
Patient involvement through shared decision making and use of a decision aid[g]				
9	1.03 (0.90-1.17)	.70	1.02 (0.90-1.16)	.75
10	0.85 (0.75-0.97)	.02	0.85 (0.74-0.96)	.01
Alternative and course		.006		N/A
A: No cancer detected	1 (reference)	N/A	1 (reference)	N/A
B: Treatable PCa[h]	1.01 (0.94-1.08)	.79	1.01 (0.94-1.09)	.72
C: Fatal PCa	0.91 (0.85-0.98)	.008	0.91 (0.84-0.97)	.007

[a]Comparison B compares responders to nonresponders among all individuals with the digital mailbox.

[b]Odds ratios (ORs) denote the probability of response (ie, OR >1 means more likely to respond).

[c]Adjusted for age and geography and mutually adjusted.

[d]N/A: not applicable.

[e]For example, in one version of the vignette, the fictional doctor performs the prostate-specific antigen (PSA) test without any test information and the patient described in the vignette is later successfully treated for prostate cancer (PCa) (alternative B).

[f]For example, in one version, the patient chooses to have a PSA test following brief information about the test (showing no PCa; alternative A), and in other versions, the patient chooses not to have a test after being slightly nudged in favor or in disfavor of the PSA test and is later diagnosed with nontreatable PCa (alternative C).

[g]For example, in one version, the patient is subject to a shared decision-making aid and dialogue with the doctor, chooses not to take the PSA test, and is later diagnosed with treatable PCa (alternative B).

[h]PCa: prostate cancer.

Table 5. Comparison of responders with the entire Danish population of men.

Characteristic	Sample (all responders) (N=6756), n (%)	Denmark (N=951,247)[a], n (%)	*P* value	*P* value adjusted[b]
Age (years)			<.001	<.001
45-49	935 (13.84%)	195,838 (20.59%)		
50-54	1225 (18.13%)	212,293 (22.32%)		
55-59	1368 (20.25%)	188,671 (19.83%)		
60-64	1,406 (20.81%)	169,672 (17.84%)		
65-70	1822 (26.97%)	187,926 (19.76%)		
Municipality type			<.001	.006
Urban municipality	2828 (41.86%)	425,073 (44.69%)		
Rural municipality	2042 (30.22%)	279,102 (29.34%)		
Urban-rural municipality	1183 (17.51%)	159,240 (16.74%)		
Outskirts municipality	703 (10.41%)	90,985 (9.57%)		

[a]Danish men aged 45 to 70 years as of January 1, 2019.

[b]Adjusted for age and municipality type.

Tables 6 and 7 compare municipality-level characteristics in responders and nonresponders. In unadjusted models, higher population density, higher tax base, higher proportion of educated citizens, and higher proportion of citizens from non-Western countries were associated with lower RRs. Mutual adjustments were made; however, the associations weakened. Most consistently, RRs were found to be lower in population dense and higher tax base areas. Counts and frequencies corresponding to the data in Tables 6 and 7 are presented in Multimedia Appendix 3.

Table 6. Comparison of municipality-level characteristics in responders and nonresponders.

Characteristic	Comparison A[a] (N=23,839), OR[b] (95% CI)	Comparison A, *P* value	Comparison B[c] (N=22,288), OR (95% CI)	Comparison B, *P* value
Population density (citizens/km^2)				
First tertile (least population density)	1 (reference)	N/A[d]	1 (reference)	N/A
Second tertile (middle population density)	1.01 (0.95-1.08)	.75	1.01 (0.94-1.08)	.81
Third tertile (most population density)	0.84 (0.79-0.90)	<.001	0.85 (0.79-0.91)	<.001
Tax per citizen				
First tertile (lowest tax base)	1 (reference)	N/A	1 (reference)	N/A
Second tertile (middle tax base)	1.00 (0.94-1.07)	.94	0.97 (0.93-1.07)	.91
Third tertile (highest tax base)	0.86 (0.80-0.92)	<.001	0.85 (0.79-0.91)	<.001
Proportion of citizens aged 25-64 years with higher education				
First tertile (fewest with high education)	1 (reference)	N/A	1 (reference)	N/A
Second tertile (middle proportion with high education)	1.02 (0.96-1.09)	.52	1.02 (0.95-1.09)	.64
Third tertile (most with high education)	0.89 (0.83-0.95)	.001	0.89 (0.83-0.95)	.001
Number of citizens from non-Western countries per 10,000 people				
First tertile (fewest non-Western)	1 (reference)	N/A	1 (reference)	N/A
Second tertile (middle number non-Western)	1.05 (0.98-1.12)	.15	1.07 (1.00-1.15)	.048
Third tertile (most non-Western)	0.89 (0.83-0.95)	.001	0.91 (0.84-0.97)	.007

[a]Comparison A compares responders to all nonresponders with or without the digital mailbox.

[b]OR: odds ratio.

[c]Comparison B compares responders to nonresponders with the digital mailbox.

[d]N/A: not applicable.

Table 7. Comparison of municipality-level characteristics adjusted for population density, tax per citizen, proportion with higher education, and number of citizens from non-Western countries.

Characteristic	Comparison A[a] (N=23,839), OR[b] (95% CI)	Comparison A, *P* value	Comparison B[c] (N=22,288), OR (95% CI)	Comparison B, *P* value
Population density (citizens/km^2)				
First tertile (least population density)	1 (reference)	N/A[d]	1 (reference)	N/A
Second tertile (middle population density)	0.98 (0.90-1.08)	.71	0.98 (0.89-1.07)	.62
Third tertile (most population density)	0.85 (0.73-1.00)	.049	0.84 (0.71-0.99)	.03
Tax per citizen				
First tertile (lowest tax base)	1 (reference)	N/A	1 (reference)	N/A
Second tertile (middle tax base)	0.98 (0.91-1.06)	.58	0.97 (0.90-1.05)	.45
Third tertile (highest tax base)	0.91 (0.83-1.00)	.054	0.89 (0.81-0.98)	.02
Proportion of citizens aged 25-64 years with higher education				
First tertile (fewest with high education)	1 (reference)	N/A	1 (reference)	N/A
Second tertile (middle proportion with high education)	1.07 (0.98-1.18)	.14	1.07 (0.98-1.18)	.15
Third tertile (most with high education)	1.03 (0.92-1.16)	.59	1.03 (0.92-1.16)	.57
Number of citizens from non-Western countries per 10,000 people				
First tertile (fewest non-Western)	1 (reference)	N/A	1 (reference)	N/A
Second tertile (middle number non-Western)	1.05 (0.98-1.13)	.18	1.08 (1.00-1.16)	.052
Third tertile (most non-Western)	1.04 (0.92-1.19)	.50	1.08 (0.95-1.23)	.22

[a]Comparison A compares responders to all nonresponders with or without the digital mailbox.

[b]OR: odds ratio.

[c]Comparison B compares responders to nonresponders with the digital mailbox.

[d]N/A: not applicable.

Discussion

Principal Findings

Using web-based surveys to collect data for public health research provides an opportunity to get easy access to potential responders while reducing efforts and research costs [7]. As is the case with other types of surveys, however, responders sometimes may not accurately represent the group of interest. In this study, we report on the representativeness of a large sample of adult men recruited through the use of a national web-based communication channel (the Danish digital mailbox). Our findings do not indicate that we received responses from only a highly selected group of the population. Nevertheless, we found that older men and men living in rural areas were more likely to respond than younger men, while RRs were lower in high economic resource areas. We discuss the findings in detail with reference to the existing literature below.

Comments Regarding the Use of Digital Mailbox Solutions for Research

There are multiple benefits of using web surveys for health research [9,23]. For example, participants may enter their responses confidentially and directly into the electronic database, allowing for more complete responses to sensitive questions and making subsequent data management much easier [23,24]. Furthermore, web surveys benefit from the ability to automatically branch into different scenarios, skipping irrelevant questions, etc [24]. As a result, savings are potentially larger compared with traditional mail or telephone surveys [25]. Uneven distribution in the use of information technologies among groups with different backgrounds may, however, challenge the use of web-based surveys for research [7]. In this regard, distributing surveys through a public platform, which is essentially mandatory for citizens to use, may be an attractive solution. Danish public authorities like hospitals and municipalities use the digital mailbox to inform citizens about tax issues, medical examinations, etc, adding to its legitimacy as an important platform for information exchange [9,26]. This may itself promote survey participation and, in part, prevent nonresponse, if not misused. It is noteworthy that 93.49% (22,288/23,839) of our target population was registered with the digital mailbox, a finding that corresponds very well with available public reports on digital mailbox coverage (92.9% among all Danish men as of the first quarter of 2019 [27]). As discussed below, this relatively high coverage is not necessarily reflected in a high RR.

The digital mailbox has been previously used for national-level survey research. Researchers recently reported on their

experiences with using the digital mailbox for distribution of "The Danish Health and Morbidity Surveys" to describe the status and trends in health and morbidity in the adult Danish population aged 16 years or older [9]. The authors found that in 2017, 90.1% of their sample was registered to use the digital mailbox; however, there were variations across age groups from 98.7% among individuals aged 24 years to 68.7% among individuals aged 65 years. Among men aged 45 years, they found RRs between 57.1% and 66.3%. Their survey may generally appeal to a broader audience thereby contributing to a higher RR. The web-based survey methodology constitutes a particularly attractive opportunity for sending reminders with very little effort, and "The Danish Health and Morbidity Surveys" indeed benefitted from this ability by using a total of four reminders plus the introduction letter to increase the RR. Using multiple reminders, however, could potentially raise ethical concerns with regard to respecting citizens' rights not to join a research project.

Discussion Regarding the Response Rate and Representativeness

In "Encyclopedia of Survey Research Methods," Davern concluded that the standards for true representativeness in surveys are rarely met; however, "the biases produced by failures often are not severe enough to threaten the ultimate value of the survey findings" [8]. Likewise, there is no clear answer to the question about when an RR is acceptable [25]. Based on a review of 30 published studies, the RR itself was concluded to not be a good indicator of the magnitude of nonresponse bias [25,28]. Although very high RRs tended to reduce the chance of nonresponse bias, when bias did occur, the degree of bias was not necessarily low [28]. Furthermore, considerable variation often appeared in the degree of bias among variables within a study.

The RR generally can be affected by many factors [29,30]. For example, a smaller proportion of individuals responding probably would be expected in a more time-consuming survey. Previous research on the relationship between survey length and the RR has predominantly focused on other survey modalities (postal service and phone interviews) [31]. Our findings suggest that the relationship between the amount of survey material and the RR may not be that clear for web-based surveys. In any case, our RR (30%) was lower than that usually seen in Danish population-based surveys [32,33]. In addition to the impact of only sending one reminder, the smaller RR may mirror the fact that the questionnaire was not available as a hard copy. Additionally, it may reflect that the RR is sometimes lower among men than among women [7,33]. Furthermore, it should be kept in mind that there was no direct benefit for survey responders apart from the sense of contributing to scientific knowledge about a patient's wish for involvement in health care decisions. Finally, to solve the problem with missing values, we constructed the web survey in such a way that participation without fully completing the questionnaire was not possible. This may have contributed to the smaller RR.

Our finding that RRs are higher in older men conflicts with the finding in a previous Danish survey [9] but seems to agree with the finding in another Danish investigation [33]. Keeping in mind that PCa risk rises with age, it is understandable that older men will be more concerned about this issue [12]. The higher RR may, however, also reflect that even if younger citizens have better access to the internet, older and retired people have more time available for research participation [7]. In accordance with previous findings, the RR seemed smaller in high income populations [33]. Associations overall were rather similar between comparisons of responders to all nonresponders and comparisons of responders to nonresponders with the digital mailbox, suggesting that nonresponders without the digital mailbox were fairly similar to nonresponders with the digital mailbox in terms of the variables under study.

Interpretation of the Study Findings

Our findings make it necessary to consider the possible consequences for further analyses of our survey data and the question of whether the composition of our sample tends to overrepresent particular viewpoints on health care decision-making. Regarding the association between age and preferences for patient involvement, some previous studies pointed toward older people having preferences for less participation in decision-making [34-37]. On the other hand, in the review by Hubbard et al that included 11 studies on the association between age and role preferences in decisions about cancer treatment, no general conclusion could be drawn [38]. Five studies reported that younger people were more likely to prefer a collaborative or active role in decision-making, but regarding PCa in particular, no relevant association with responder age was found in any one of three studies.

We found the RR to be much higher in rural municipalities. Furthermore, the RR was lower in areas with a higher tax base. Although the RR initially seemed to be lower in higher education areas, this association faded in adjusted analyses. Additionally, the RR seemed to be lower in areas with a higher proportion of citizens from non-Western countries, which might partly reflect the fact that the questionnaire was only available in Danish, although this association also disappeared in adjusted analyses controlling for the association with municipality-level population density and average tax per citizen.

A Cochrane review on patient decision aids suggested that the desire for involvement in health care decision-making, rather than being a stable trait, should be considered an adaptable way of thinking ("state") [39]. People simply have to be informed about why they should participate in a certain decision and understand the importance of their own preferences for outcomes of options (which outcomes matter most to them) before they are asked about their desire to participate. In spite of this, some studies have investigated associations between sociodemographic factors and preferences for involvement in decisions. Degner and Sloan found a nonrelevant trend for individuals from more rural areas to give away control to their physicians to a greater extent [36]. With regard to the impact of education, in the study by Rovner et al, college-educated men tended to want to make their own decisions about health care in benign prostatic hyperplasia, whereas noncollege-educated men tended to desire a shared approach [40]. Additionally, those preferring that decisions should be made mainly by the patient had a higher income.

Correspondingly, in a large survey, Levinson et al found that more educated and healthier people were more likely to prefer an active role in decision-making [41]. On the other hand, African-American and Hispanic responders were more likely to prefer a passive role in decision-making. Generally, preferences for involvement increased with age up to 45 years and then declined. In a later study, Peek et al confirmed no racial differences in preferences for participation in decision-making [42].

While previous studies suggest that preferences for participation in health care decision-making may vary with the aforementioned factors, the latter factors may play no clear-cut role when it comes to decisions about cancer treatment. Hence, it was generally concluded from the review by Hubbard et al of 31 papers in total investigating patient preferences for participation in decision-making that the evidence on associations with age as well as gender, education level, marital status, socioeconomic status, and health status was inconclusive [38]. However, sample sizes were small in the majority of the studies. Patient preferences for involvement in decision-making are hypothesized to vary with age, socioeconomic status, etc [37], and therefore, statistical weighting may be warranted to control for skewness.

Limitations

Some previous studies have suggested gender differences regarding preferences for participation in decision-making [34-36,41,43]. This hardly can be controlled for in a survey including only male responders, but points to possible limitations regarding the generalizability of survey RR findings to women. In this regard, however, the role of gender has been questioned, at least when it comes to cancer issues [38]. Furthermore, it should be noted that only municipality-level register data on responders' sociodemographic characteristics were used. For example, the income of individual responders was not available. Likewise, more complex issues with regard to, for example, psychological disposition was not taken into consideration. In this regard, preferences for involvement in decision-making actually have been hypothesized to vary among personality types, while individuals' likelihood of survey participation simultaneously may vary with personality style [44-46].

Conclusion

The generalizability of survey findings to target populations is an important goal in research using web survey methodology. We wanted to establish the representativeness of responders in a large national survey investigating the desire of men to participate in decision-making about undergoing a PSA test, using the national digital mailbox platform. Our comparisons of responders to nonresponders point toward a reasonable representativeness of the sample. With regard to the result that responders did not fully represent men aged 45 to 70 years, previous research findings suggest that the variation found in our sample may not necessarily deteriorate forthcoming analyses on preferences for involvement in decision-making. In any case, identification of imbalances allows for statistical corrections to be made during the analysis of survey responses.

Acknowledgments

We would like to thank the survey responders for their participation and contributions to the study. The project was funded by grants of €40,000 from the Danish Health Insurance Foundation and €5700 from the Lilly & Herbert Hansen's Foundation. The funding bodies had no influence on the design of the study. Moreover, the funding bodies had no influence on collection, analysis, and interpretation of data or writing of the manuscript.

Authors' Contributions

SB collected, analyzed, and interpreted the data used in this study and was a major contributor in writing the manuscript. AH assisted in collecting the data. SM assisted in analyzing the data. All authors assisted in interpreting the data. All authors read, commented on, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Respondent numbers and response rates regarding the different questionnaire scenario variants of the survey.
[DOCX File , 26 KB - jmir_v22i9e19517_app1.docx]

Multimedia Appendix 2
Time spent answering the questionnaire.
[DOCX File , 26 KB - jmir_v22i9e19517_app2.docx]

Multimedia Appendix 3
Counts and proportions for comparisons of municipality-level characteristics in respondents and nonrespondents.
[DOCX File , 25 KB - jmir_v22i9e19517_app3.docx]

References

1. Bismark M, Dauer E, Paterson R, Studdert D. Accountability sought by patients following adverse events from medical care: the New Zealand experience. CMAJ 2006 Oct 10;175(8):889-894 [FREE Full text] [doi: 10.1503/cmaj.060429] [Medline: 17030939]
2. Beckman HB, Markakis KM, Suchman AL, Frankel RM. The doctor-patient relationship and malpractice. Lessons from plaintiff depositions. Arch Intern Med 1994 Jun 27;154(12):1365-1370. [Medline: 8002688]
3. Birkeland S, Depont Christensen R, Damsbo N, Kragstrup J. Characteristics of complaints resulting in disciplinary actions against Danish GPs. Scand J Prim Health Care 2013 Sep;31(3):153-157 [FREE Full text] [doi: 10.3109/02813432.2013.823768] [Medline: 23906082]
4. Huntington B, Kuhn N. Communication gaffes: a root cause of malpractice claims. Proc (Bayl Univ Med Cent) 2003 Apr;16(2):157-61; discussion 161 [FREE Full text] [doi: 10.1080/08998280.2003.11927898] [Medline: 16278732]
5. King JS, Moulton BW. Rethinking informed consent: the case for shared medical decision-making. Am J Law Med 2006;32(4):429-501. [doi: 10.1177/009885880603200401] [Medline: 17240730]
6. Birkeland S. Informed Consent Obtainment, Malpractice Litigation, and the Potential Role of Shared Decision-making Approaches. Eur. J. Health Law 2017 May 31;24(3):264-284 [FREE Full text] [doi: 10.1163/15718093-12341410]
7. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003 Jan;57(1):68-73 [FREE Full text] [doi: 10.1136/jech.57.1.68] [Medline: 12490652]
8. Davern M. Representative Sample. In: Lavrakas P, editor. Encyclopedia of Survey Research Methods. Thousand Oaks, California: SAGE Publications, Inc; 2008.
9. Jensen HA, Ekholm O, Davidsen M, Christensen AI. The Danish health and morbidity surveys: study design and participant characteristics. BMC Med Res Methodol 2019 May 03;19(1):91 [FREE Full text] [doi: 10.1186/s12874-019-0733-9] [Medline: 31053088]
10. REDCap. URL: https://www.project-redcap.org/ [accessed 2020-08-11]
11. Birkeland S, Pedersen SS, Haakonsson AK, Barry MJ, Rottmann N. Men's view on participation in decisions about prostate-specific antigen (PSA) screening: patient and public involvement in development of a survey. BMC Med Inform Decis Mak 2020 Apr 06;20(1):65 [FREE Full text] [doi: 10.1186/s12911-020-1077-4] [Medline: 32252729]
12. US Preventive Services Task Force, Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, et al. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2018 May 08;319(18):1901-1913. [doi: 10.1001/jama.2018.3710] [Medline: 29801017]
13. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol 2017 Apr;71(4):618-629. [doi: 10.1016/j.eururo.2016.08.003] [Medline: 27568654]
14. Social- og Indenrigsministeriet. URL: http://www.noegletal.dk/ [accessed 2020-07-30]
15. Olsen KR, Sørensen TH, Gyrd-Hansen D. [Association between productivity, list size, patient and practice characteristics in general practice]. Ugeskr Laeger 2010 Apr 19;172(16):1192-1196. [Medline: 20423660]
16. Birkeland S, Christensen RD, Damsbo N, Kragstrup J. Patient complaint cases in primary health care: what are the characteristics of general practitioners involved? Biomed Res Int 2013;2013:807204 [FREE Full text] [doi: 10.1155/2013/807204] [Medline: 24027764]
17. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
18. DeWitt T, Brady MK. Rethinking Service Recovery Strategies. Journal of Service Research 2016 Jun 29;6(2):193-207. [doi: 10.1177/1094670503257048]
19. Altman D. Practical Statistics for Medical Research. London: Chapman and Hall; 1991.
20. Barry MJ, Wexler RM, Brackett CD, Sepucha KR, Simmons LH, Gerstein BS, et al. Responses to a Decision Aid on Prostate Cancer Screening in Primary Care Practices. Am J Prev Med 2015 Oct;49(4):520-525. [doi: 10.1016/j.amepre.2015.03.002] [Medline: 25960395]
21. Kristensen I, Kjeldsen C, Dalgaard T. Landdistriktskommuner - indikatorer for landdistrikt. PURE.au.dk. Tjele: Danmarks Jordbrugsforskning; 2007 Feb 28. URL: https://pure.au.dk/ws/files/4594133/RapLanddistrikt2007_20070201.pdf [accessed 2020-08-11]
22. Danmarks Statistik. URL: https://www.dst.dk/da [accessed 2020-07-30]
23. Couper M. Web surveys: a review of issues and approaches. Public Opin Q 2000;64(4):464-494. [Medline: 11171027]
24. Ryan JM, Corry JR, Attewell R, Smithson MJ. A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. Qual Life Res 2002 Feb;11(1):19-26. [doi: 10.1023/a:1014415709997] [Medline: 12003052]
25. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. BMC Med Res Methodol 2012 Aug 31;12:132 [FREE Full text] [doi: 10.1186/1471-2288-12-132] [Medline: 22938205]

26. Porter SR, Whitcomb ME. Mixed-Mode Contacts In Web Surveys: Paper is Not Necessarily Better. Public Opinion Quarterly 2007 Sep 18;71(4):635-648. [doi: 10.1093/poq/nfm038]
27. Digitaliseringsstyrelsen. URL: https://digst.dk/it-loesninger/digital-post/om-loesningen/tal-og-statistik-om-digital-post/ [accessed 2020-07-30]
28. Groves R. Nonresponse Rates and Nonresponse Bias in Household Surveys. The Public Opinion Quarterly 2006;70(5):646-675 [FREE Full text] [doi: 10.1093/poq/nfl033]
29. Marcus B, Bosnjak M, Lindner S, Pilischenko S, Schütz A. Compensating for Low Topic Interest and Long Surveys. Social Science Computer Review 2016 Aug 18;25(3):372-383 [FREE Full text] [doi: 10.1177/0894439307297606]
30. Turk T, Elhady MT, Rashed S, Abdelkhalek M, Nasef SA, Khallaf AM, et al. Quality of reporting web-based and non-web-based survey studies: What authors, reviewers and consumers should consider. PLoS One 2018;13(6):e0194239 [FREE Full text] [doi: 10.1371/journal.pone.0194239] [Medline: 29912881]
31. Burkhart Q, Orr N, Brown JA, Hays RD, Cleary PD, Beckett MK, et al. Associations of Mail Survey Length and Layout With Response Rates. Med Care Res Rev 2019 Nov 20:1077558719888407. [doi: 10.1177/1077558719888407] [Medline: 31747849]
32. Rasmussen S, Søndergaard J, Larsen PV, Balasubramaniam K, Elnegaard S, Svendsen RP, et al. The Danish Symptom Cohort: Questionnaire and Feasibility in the Nationwide Study on Symptom Experience and Healthcare-Seeking among 100 000 Individuals. Int J Family Med 2014;2014:187280 [FREE Full text] [doi: 10.1155/2014/187280] [Medline: 25147736]
33. Elnegaard S, Andersen RS, Pedersen AF, Larsen PV, Søndergaard J, Rasmussen S, et al. Self-reported symptoms and healthcare seeking in the general population--exploring "The Symptom Iceberg". BMC Public Health 2015 Jul 21;15:685 [FREE Full text] [doi: 10.1186/s12889-015-2034-5] [Medline: 26195232]
34. Marahrens L, Kern R, Ziemssen T, Fritsche A, Martus P, Ziemssen F, et al. Patients' preferences for involvement in the decision-making process for treating diabetic retinopathy. BMC Ophthalmol 2017 Aug 09;17(1):139 [FREE Full text] [doi: 10.1186/s12886-017-0526-z] [Medline: 28793881]
35. Stiggelbout AM, Kiebert GM. A role for the sick role. Patient preferences regarding information and participation in clinical decision-making. CMAJ 1997 Aug 15;157(4):383-389 [FREE Full text] [Medline: 9275945]
36. Degner LF, Sloan JA. Decision making during serious illness: what role do patients really want to play? J Clin Epidemiol 1992 Sep;45(9):941-950. [doi: 10.1016/0895-4356(92)90110-9] [Medline: 1432023]
37. Robinson A, Thomson R. Variability in patient preferences for participating in medical decision making: implication for the use of decision support tools. Qual Health Care 2001 Sep;10 Suppl 1:i34-i38 [FREE Full text] [doi: 10.1136/qhc.0100034] [Medline: 11533436]
38. Hubbard G, Kidd L, Donaghy E. Preferences for involvement in treatment decision making of patients with cancer: a review of the literature. Eur J Oncol Nurs 2008 Sep;12(4):299-318. [doi: 10.1016/j.ejon.2008.03.004] [Medline: 18486552]
39. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
40. Rovner DR, Wills CE, Bonham V, Williams G, Lillie J, Kelly-Blake K, et al. Decision aids for benign prostatic hyperplasia: applicability across race and education. Med Decis Making 2004;24(4):359-366. [doi: 10.1177/0272989X04267010] [Medline: 15271274]
41. Levinson W, Kao A, Kuby A, Thisted RA. Not all patients want to participate in decision making. A national study of public preferences. J Gen Intern Med 2005 Jun;20(6):531-535 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.04101.x] [Medline: 15987329]
42. Peek ME, Tang H, Cargill A, Chin MH. Are there racial differences in patients' shared decision-making preferences and behaviors among patients with diabetes? Med Decis Making 2011;31(3):422-431 [FREE Full text] [doi: 10.1177/0272989X10384739] [Medline: 21127318]
43. Flynn KE, Smith MA, Vanness D. A typology of preferences for participation in healthcare decision making. Soc Sci Med 2006 Sep;63(5):1158-1169 [FREE Full text] [doi: 10.1016/j.socscimed.2006.03.030] [Medline: 16697096]
44. Reamer E, Yang F, Holmes-Rovner M, Liu J, Xu J. Influence of Men's Personality and Social Support on Treatment Decision-Making for Localized Prostate Cancer. Biomed Res Int 2017;2017:1467056 [FREE Full text] [doi: 10.1155/2017/1467056] [Medline: 28785574]
45. Orom H, Penner LA, West BT, Downs TM, Rayford W, Underwood W. Personality predicts prostate cancer treatment decision-making difficulty and satisfaction. Psychooncology 2009 Mar;18(3):290-299 [FREE Full text] [doi: 10.1002/pon.1385] [Medline: 18821530]
46. Nestler S, Thielsch M, Vasilev E, Back M. Will They Stay or Will They Go? Personality Predictors of Dropout in an Online Study. International Journal of Internet Science 2015:37-48.

Abbreviations

DHDA: Danish Health Data Authority
OR: odds ratio

PCa: prostate cancer
PSA: prostate-specific antigen
RR: response rate
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by G Eysenbach; submitted 21.04.20; peer-reviewed by X Liu, I Buljan, M Mahmic Kaknjo; comments to author 12.06.20; accepted 25.06.20; published 02.09.20.

Please cite as:
Birkeland SF, Haakonsson AK, Pedersen SS, Rottmann N, Barry MJ, Möller S
Sociodemographic Representativeness in a Nationwide Web-Based Survey of the View of Men on Involvement in Health Care Decision-Making: Cross-Sectional Questionnaire Study
J Med Internet Res 2020;22(9):e19517
URL: https://www.jmir.org/2020/9/e19517
doi:10.2196/19517
PMID:32663149

Original Paper

Online Health Resource Use by Individuals With Inflammatory Bowel Disease: Analysis Using the National Health Interview Survey

Rong Yin[1*], MS; David M Neyens[1*], MPH, PhD

Department of Industrial Engineering, Clemson University, Clemson, SC, United States

[*]all authors contributed equally

Corresponding Author:
David M Neyens, MPH, PhD
Department of Industrial Engineering
Clemson University
100 Freeman Hall
Clemson, SC, 29634
United States
Phone: 1 8646564719
Email: dneyens@clemson.edu

Abstract

Background: The internet has enabled convenient and efficient health information searching which is valuable for individuals with chronic conditions requiring some level of self-management. However, there is little research evaluating what factors may impact the use of the internet for health-related tasks for specific clinical populations, such as individuals with inflammatory bowel diseases.

Objective: Our goal was to investigate the factors that influence internet use in acquiring health information by individuals with inflammatory bowel diseases. Specifically, we identified factors associated with internet searching behavior and using the internet for completing health-related tasks.

Methods: We used 2016 National Health Interview Survey weighted data to develop logistic regression models to predict the likelihood that individuals with inflammatory bowel diseases would use the internet for 2 types of tasks: seeking health information through online searches and using the internet to perform health-related tasks including scheduling appointments and emailing care providers.

Results: 2016 National Health Interview Survey weighted data include more than 3 million weighted adult respondents with inflammatory bowel diseases (approximately 1.29% of adults in the weighted data set). Our results suggest that approximately 66.3% of those with inflammatory bowel diseases reported using the internet at least once a day, and approximately 14.7% reported being dissatisfied with their current health care. About 62.3% of those with inflammatory bowel diseases reported that they had looked up health information online, 16.3% of those with inflammatory bowel diseases reported that they had scheduled an appointment with a health care provider online, and 21.6% reported having used a computer to communicate with a health provider by email. We found that women who were self-regulating their care were more likely to look up health information online than others. Both middle-aged and older adults with inflammatory bowel diseases who were unsatisfied with their current health care were less likely to look up health information online. Frequent internet users who were worried about medical costs were more likely to look up health information online. Similarly, the results from our statistical models suggest that individuals with inflammatory bowel diseases who were frequent internet users were more likely to use the internet for specific health-related tasks. Additionally, women with inflammatory bowel diseases who reported being married were less likely to use the internet for specific health-related tasks.

Conclusions: For those with inflammatory bowel diseases, there are additional socioeconomic and behavioral factors that impact the use of the internet for health information and health-related tasks. Future research should evaluate how these factors moderate the use of the internet and identify how online resources can support clinical populations in ways that improve access to information, support health self-management, and subsequently improve health outcomes.

(J Med Internet Res 2020;22(9):e15352) doi:10.2196/15352

KEYWORDS

internet; searching behavior; access to information; inflammatory bowel disease; logistic regression model; National Health Interview Survey

Introduction

Background

The internet is seen as a reliable alternative source of health information [1,2], and people seek health information online to gain additional information about health conditions or procedures [3], as well as to discuss their specific condition and health status through online discussion groups [4]. The internet may provide a convenient method for patients to obtain health information regardless of geographical restrictions [5-7] or access to care providers. Past research [8-12] has found that using the internet to search health information leads to better health outcomes, and the internet is believed to be a good source of health information to support developing health knowledge, ongoing long-term self-management of care, and monitoring the condition of patients. Research [13] has found that most people use the internet to acquire specific information regarding their own health status or that of their family or friends.

Individuals with chronic diseases are a unique user population in terms of their potential use of online health information in self-management of their health. The prevalence of chronic diseases is high in the United States; Ward et al [14] reported that nearly 50% of adults have one chronic disease, and 25% have multiple conditions. Past research suggests that searching health information online may be a common behavior for people with chronic health conditions [15] and that online information seekers' health literacy and engagement may correlate with their ability to manage their chronic health conditions [16]. It has been shown that individuals with chronic diseases are more willing to search health information on the internet than those without such conditions [17]. In addition, patients who have chronic diseases but who do not have health insurance are more willing to search for health information on the internet than individuals with insurance [17], supporting results from other studies [3,18] that suggest that the involvement and motivation of users impact their engagement in online health information searching, with highly motivated users, such as those with chronic diseases, applying more effort in the information searching task. Additionally, there are multiple factors, including a person's gender, age, and socioeconomic status that influence an individuals' online information searching behavior and internet usage [1,15,19-23].

To ensure the effectiveness of the internet related to health information, the US Department of Health and Human Services [24] has provided design guidelines to improve the user experience of individuals with various levels of health literacy, paying special attention to people with limited abilities. Not only are those with low health literacy less likely to use the internet for information searching and emailing [25], they are also more likely to forget information and experience working memory overload when interacting with websites [24] compared to internet users with higher health literacy. These users have been found to spend 9 times longer conducting information searching tasks than higher literacy users, and they tend to read word by word rather than glancing at the entire page for the more relevant information [26]. In addition, there are other barriers for all online health information seekers including limited accessibility to the content published in research journals, the complexity of the clinical language used, and the inability to evaluate the reliability of health information websites [16]. Lee et al [16] argue that these barriers could be reduced by increasing the involvement of health professions in guiding the health information seeking process and improving general health literacy.

Crohn Disease and Ulcerative Colitis

Crohn disease and ulcerative colitis are collectively referred to as inflammatory bowel disease (IBD) [27,28], a chronic condition that affects the intestines, colon, and bowel [29]. It is a complex, incurable disease [30] that can result in long-term disability or mortality [31], and its highest incidence occurs in younger adults [31,32]. A recent study [33] suggested that the incidence of IBD has seen a dramatic increase to over 0.3% in North America and many European countries, and the incidence of IBD is expected to continuously increase [31].

Generally, the majority of studies related to IBD focus on its pathology and medical treatment. Although some studies have focused on the diagnosis of IBD [29], predictors of its disabling consequences [34], its pathogenesis [27,28,35,36], and the dietary habits of those with IBD [37,38], few have examined which factors may influence individuals with IBD to search the internet for health care–related information. Yet, the management of IBD depends on self-management of the disease and a level of health literacy. It has been found that many health websites did not provide appropriate coverage of prognoses, side effects, and additional health risks associated with IBD but did cover symptoms, complications, and treatment options [39]. Additionally, it was reported that information related to self-management of IBD was not widely included in health websites [39], and thus the use of online search behavior associated with IBD is an important area of research.

Research Objectives

The overall objective of this study was to investigate the factors that influence the use of the internet to acquire health information for individuals with IBD. We examined 2 types of internet-related activities: searching the internet for health information and using the internet for health-related tasks such as scheduling appointments with health care providers and communicating with a health care provider by email. We evaluated a number of potential factors that might impact how an individual with IBD uses the internet for health information. Previous research has shown that a number of factors impact internet usage for health information in general populations including: gender [5,6,40,41], age [40,41], level of education [42], health literacy [25], health insurance coverage [17], and level of income [19,23].

Methods

Data Source: National Health Interview Survey

The National Health Interview Survey (NHIS), which is conducted by the National Center for Health Statistics, covers broad health topics [43]. The data that are collected are weighted to represent the general population of the United States. The topics and the questions in the survey have evolved over time, and thus, the type of data collected each year varies. The 2016 NHIS [44] included questions asking respondents to self-identify as having IBD (Crohn disease and ulcerative colitis). For our study, several variables in the original data were recoded and combined to form categories to support the analysis and interpretation of the results of our statistical models. The original variable names in the NHIS data files are included in parentheses to facilitate an understanding of how we coded and used the data.

Dependent Variables

This study focused on the behaviors and experiences, during the year preceding the interview, of adult individuals who reported having IBD (ULCCOLEV). The dependent variables in this study were related to internet usage: (1) individuals searching for health information on the internet (HIT1A), (2) individuals using the internet to schedule appointments with health care providers (HIT3A), and (3) individuals using the internet to communicate with health care providers via email (HIT4A). All dependent variables were recoded as binary variables (1, they reported that they had done the activity in the previous 12 months; 0; they had not).

Independent Variables

Demographic variables such as sex (SEX) and age (AGE_P) were used in the analysis. The age variable was recoded into 3 groups: younger adults (18-35 years old), middle-age adults (36-55 years old), and older adults (older than 55). We recoded marriage status (R_MARITL) as a binary variable (1, married; 0, not married) where not married included never married, divorced, widowed, separated, as well as preferred not to answer and nonresponses. Parental status (PAR_STAT) of participants was recoded as being a parent of a child or not a parent of a child. Work status (DOINGLWA) of participants was recoded as employed or not employed.

It is possible that individuals with multiple chronic conditions may use the internet differently than those with a single chronic condition because of the complexity of managing multiple conditions. It is possible that they may receive conflicting medical advice for diverse chronic conditions [45,46]. Therefore, 7 other chronic conditions were also included in our analysis as binary variables: hypertension (HYPEV), high cholesterol (CHLEV), coronary heart disease (CHDEV), asthma (AASMEV), cancer (CANEV), diabetes (DIVEV1) and chronic/long-term liver conditions (LIVEV).

Other variables that may impact an individual's online information searching behaviors were also included in the analysis such as socioeconomic considerations, the level of satisfaction with health care services, and internet usage frequency. Whether the respondent reported having trouble finding a care provider in the previous 12 months (APRVTRYR) was recoded as reported trouble in finding a care provider and reported no trouble in finding a care provider. The respondents who reported being worried about paying medical bills (AWORPAY) were recoded as worried and not worried, with the former category including those who were very worried and those who were somewhat worried. A new variable was created to indicate whether participants were self-regulating care in a number of possible ways. This self-regulating care included whether the respondents reported doing at least one of the following actions: skipping medication doses (ARX12_1), taking less medicine (ARX12_2), delaying filling a prescription (ARX12_3), asking a doctor for less expensive medication (ARX12_4), and using alternative therapies (ARX12_6). A binary variable was created to identify whether the participants reported having seen or talked to a general practitioner in the prior year (AHCSYR9). A variable was also created to determine whether the participants tried to purchase health insurance directly in the prior 3 years by combining the 2 relevant variables of "Tried to purchase health insurance directly" (AINDINS2) and "Purchased health insurance directly" (AINDPRCH). The satisfaction of participants in their health care (ASISATHC) was recoded as satisfied and not satisfied, with the satisfied category including those who reported being very or somewhat satisfied with their health care services. A variable was created identifying frequent internet users based on the respondent's frequency of internet usage (AWEBOFNO and AWEBOFTP). Frequent internet users were identified as such if the internet was used at least once a day (ie, at least 7 times per week) and were classified as not frequent internet users otherwise.

Statistical Analysis

The data were analyzed using R (version 3.5.0). Specifically, the svyglm function (Survey package; version 3.34) [47] was used for logistic regression, and stepwise deletion was used to remove insignificant parameters from the model in order to identify the best model for each dependent variable. As the weighted sample size was large, α=.01 was used to assess significance.

Results

Descriptive Statistics

After applying the data weights, the sample size of individuals who reported having IBD was 3,155,477 (approximately 1.29% of all the adults in the weighted data set); approximately 64.4% (2,032,022) of the respondents were female, the average age of the respondents was 52.8 (SE 0.87) years, and approximately 49.9% of the respondents (1,575,168) reported being married. Approximately 80.7% (2,544,995/3,155,477) of the respondents reported having seen or talked to a general practitioner in the previous year, with very few (273,977/3,155,477, 8.7%) reporting having trouble finding a provider in the previous 12 months, although 14.7% (464,376/3,155,477) reported being dissatisfied with their health care. Approximately 42.6% (1,344,253/3,155,477) and 41.2% (1,288,836/3,155,477) of the respondents also reported having hypertension or high cholesterol, respectively, which were the 2 highest prevalences

of comorbidities examined for individuals who had IBD. More than half of the respondents (1,965,639/3,155,477, 62.3%) reported looking up health information online, and approximately 66.3% (2,090,505/3,155,477) reported being frequent internet users, using it at least daily. In terms of the health-related tasks, 16.3% (515,253/3,155,477) of those with IBD reported scheduling an appointment with a health care provide online, and 21.6% (680,872/3,155,477) reported having used computer to communicate with a health provider by email. The complete demographic information of the respondents is in Table 1.

Table 1. The characteristics of the sample of survey respondents who reported having IBD.

Variable	Weighted, n (%)
Age	
Younger adults (18-35 years old)	454,950 (14.4)
Middle age adults (36-55 years old)	1,159,430 (36.7)
Older adults (>55 years old)	1,541,097 (48.8)
Sex	
Male	1,123,455 (35.6)
Female	2,032,022 (64.4)
Married	1,575,168 (49.9)
Employed	1,548,101 (49.1)
Has at least one child	670,310 (21.2)
Looked up health information online	1,965,639 (62.3)
Used computers to schedule an appointment with a health care provider	515,253 (16.3)
Used computer to communicate with a health care provider by email	680,872 (21.6)
Reported having hypertension	1,344,253 (42.6)
Reported having high cholesterol	1,298,836 (41.2)
Reported having coronary heart disease	320,715 (10.2)
Reported having asthma	636,538 (20.2)
Reported having cancer	491,356 (15.6)
Reported having diabetes	564,795 (17.9)
Reported having chronic/long-term liver conditions	127,679 (4.0)
Reported having trouble in finding a provider in the previous 12 months	273,977 (8.7)
Reported being worried about paying medical bills	1,732,203 (54.9)
Reported multiple types of self-regulating care	1,192,446 (37.9)
Reported having seen or talked to a general doctor in the previous year	2,544,995 (80.7)
Reported trying to purchase health insurance directly in the previous 3 years	426,541 (13.5)
Reported being unsatisfied with their health care	464,376 (14.7)
Used the internet frequently (at least daily usage)	2,090,505 (66.3)
Reported being worried about medical costs	1,618,723 (51.3)

Looking Up Health Information on the Internet

A binary logit model was created to evaluate how individuals with IBD use the internet for information seeking (Table 2). Among the individuals with IBD, those who also had asthma were more likely to look up health information online compared to others (adjusted odds ratio [OR] 2.97, 99% CI 1.17 to 7.54). Although several different types of chronic conditions were initially included in the model, only the variable indicating asthma was a significant predictor impacting the likelihood of those with IBD looking up health information online.

Both middle-aged and older women were less likely to look up health information online compared to others (adjusted OR 0.07, 99% CI 0.004 to 0.96 and adjusted OR 0.02, 99% CI 0.001 to 0.29, respectively). Women with IBD who reported self-regulating care were more likely to look up health information online than others (adjusted OR 9.87, 99% CI 1.49 to 65.37). Both middle-aged (36-55 years old) and older (over 55 years old) adults who were married were more likely to look up health information online (adjusted OR 22.20, 99% CI 1.46 to 336.97 and adjusted OR 23.81, 99% CI 1.75 to 327.01, respectively). Both middle-aged and older adults who were

unsatisfied with their current health care were less likely to look up health information online (adjusted OR 0.03, 99% CI 0.002 to 0.58 and 0.03, 99% CI 0.001 to 0.71, respectively). Individuals who were employed and were unsatisfied with their current health care were less likely to look up health information online (adjusted OR 0.07, 99% CI 0.007 to 0.62). Additionally, frequent internet users who were worried about the medical costs of an illness/accident were more likely to look up health information online (adjusted OR 12.18, 99% CI 2.08 to 72.24).

Table 2. Binary logit model for the likelihood of looking up health information on internet.

Parameter	Estimate	99% CI	SE	*t* value	*P* value	Adjusted OR[a]	99% CI
Intercept	–2.95	(–4.91, –0.99)	0.76	–3.87	<.001	0.05	(0.007, 0.37)
Female	3.08	(0.75, 5.42)	0.91	3.40	.001	21.76	(2.12, 225.88)
Middle-aged adults	0.98	(–1.11, 3.08)	0.81	1.21	.228	—[b]	
Older adults	1.43	(–0.59, 3.44)	0.78	1.83	.068	—	
Married	–2.72	(–5.03, –0.42)	0.90	–3.04	.002	0.07	(0.007, 0.66)
Employed	0.95	(–0.06, 1.95)	0.39	2.42	.016	—	
Had asthma	1.09	(0.16, 2.02)	0.36	3.02	.003	2.97	(1.17, 7.54)
Self-regulating care	–1.30	(–2.72, 0.13)	0.55	–2.34	.019	—	
Unsatisfied with health care	4.15	(1.08, 7.22)	1.19	3.49	.001	63.52	(2.94, 1366.49)
Worried about medical costs of illness/accident	–1.30	(-2.57, -0.02)	0.50	–2.62	.009	0.27	(0.08, 0.98)
Frequent internet users	2.60	(1.47, 3.73)	0.44	5.92	<.001	13.42	(4.35, 41.68)
Female × middle-aged adults	–2.72	(–5.40, –0.04)	1.04	–2.62	.009	0.07	(0.004, 0.96)
Female × older adults	–3.91	(–6.59, –1.23)	1.04	–3.76	<.001	0.02	(0.001, 0.29)
Female × self-regulating care	2.29	(0.40, 4.18)	0.73	3.12	.002	9.87	(1.49, 65.37)
Middle-aged adults × married	3.10	(0.38, 5.82)	1.06	2.93	.004	22.20	(1.46, 336.97)
Older adults × married	3.17	(0.56, 5.79)	1.01	3.13	.002	23.81	(1.75, 327.01
Middle-aged adults × unsatisfied with health care	–3.51	(–6.47, –0.55)	1.15	–3.06	.002	0.03	(0.002, 0.58)
Older adults × unsatisfied with health care	–3.48	(–6.61, –0.34)	1.22	–2.86	.004	0.03	(0.001, 0.71)
Employed × unsatisfied with health care	–2.72	(–4.97, –0.48)	0.87	–3.12	.002	0.07	(0.007, 0.62)
Worried about medical costs of illness/accident × frequent internet users	2.50	(0.73, 4.28)	0.69	3.64	<.001	12.18	(2.08, 72.24)

[a]OR: odds ratio.

[b]No statistically significant differences were found at α=.01.

Using Computers to Schedule an Appointment With a Health Care Provider

A binary logistic regression model was created to predict the likelihood that an individual with IBD used a computer to schedule an appointment with their care provider (see Table 3). Those who reported self-regulating their care were more likely to use the internet to schedule an appointment with a provider than those who did not self-regulate (adjusted OR 2.61, 99% CI 1.05 to 6.49). Those who were frequent internet users were more likely to use the internet to schedule an appointment with a provider than nonusers or infrequent users (adjusted OR 15.18, 99% CI 3.56 to 64.72). Women who reported being married were less likely to use the internet to schedule an appointment with a provider (adjusted OR 0.07, 99% CI 0.007 to 0.75).

Table 3. Binary logit model for the likelihood of using the internet to schedule an appointment with a health care provider.

Parameter	Estimate	99% CI	SE	t value	*P* value	Adjusted OR[a]	99% CI
Intercept	–5.82	(–8.23, –3.42)	0.93	–6.24	<.001	0.003	(<0.001,0.03)
Female	1.84	(–0.12, 3.79)	0.76	2.42	.016	__[b]	—
Married	2.10	(0.09, 4.11)	0.78	2.69	.007	8.17	(1.09,60.95)
Self-regulating care	0.96	(0.05, 1.87)	0.35	2.72	.007	2.61	(1.05,6.49)
Frequent internet users	2.72	(1.27, 4.17)	0.56	4.82	<.001	15.18	(3.56,64.72)
Female × married	–2.60	(–4.92, –0.29)	0.90	–2.90	.004	0.07	(0.007,0.75)

[a]OR: odds ratio.

[b]No statistically significant differences was found at α=.01.

Using Email to Communicate With a Health Care Provider

A binary logistic regression model was created to predict the likelihood that an individual with IBD used email to communicate with their care provider (see Table 4). Those who were frequent internet users were more likely to report using email to communicate with a provider (adjusted OR 8.41, 99% CI 3.22 to 21.76). Women who reported being married were less likely to report using email to communicate with a care provider than others (adjusted OR 0.15, 99% CI 0.02 to 0.93).

Table 4. Binary logit model for the likelihood of emailing a health care provider.

Parameter	Estimate	99% CI	SE	t value	*P* value	Adjusted OR[a]	99% CI
Intercept	–4.02	(–5.60, –2.43)	0.61	-6.54	<.001	0.02	(0.003,0.09)
Female	1.36	(–0.10, 2.83)	0.57	2.41	.017	__[b]	—
Married	1.42	(–0.07, 2.91)	0.58	2.45	.014	—	—
Frequent internet users	2.13	(1.17, 3.08)	0.37	5.75	<.001	8.41	(3.22,21.76)
Female × married	–1.88	(–3.69, –0.07)	0.70	–2.67	.008	0.15	(0.02,0.93)

[a]OR: odds ratio.

[b]No statistically significant difference was found at α=.01.

Discussion

Principal Findings

Our study examined the use of the internet by individuals with IBD to seek health information and to perform health-related activities. The population of interest was examined because these chronic conditions are often self-managed [48], and for those with IBD, understanding their own chronic conditions, experiences, and psychosocial factors can be a critical aspect of their treatment process [49]. Therefore, information acquisition and use are vital for those with chronic conditions to be able to self-regulate their health conditions [50].

In general, previous studies [5,40-42,51] suggest that the gender and age of individuals impact their internet usage for health information. In our model, women who self-regulated their care were more likely to look up health information online. Whereas, women in the middle-age and older age groups were both less likely to look up health information online. It has been suggested that younger individuals are more likely to use the internet than older individuals [49], and the same may be true for using the internet for health information seeking. Future research should continue to examine how the gender and age interaction influence searching for health information on the internet. The main effect of age was not significant in our study which is inconsistent with the findings of previous studies [42,50]. This may due to the fact that we defined age as a 3-level categorical variable (younger adults, middle-age adults, and older adults) and not as a continuous variable. Future studies could examine the impact of age as a continuous variable on the internet usage by individuals with specific chronic conditions including those with IBD.

As the literature suggests, individuals in poor health tend to use the internet more frequently than healthy individuals to look up health information [5,52,53]. Previous research [15,17] has suggested that individuals with multiple chronic health conditions are more likely to use the internet to acquire information with the expectation that it will help improve their condition. Our results suggest that individuals who reported having asthma in addition to IBD were more likely to use the internet for health care information searching. No other comorbidities were significant predictors in our models. Future research should more comprehensively examine comorbidity categories and types to identify if the results for IBD mirror those from previous studies [15,17].

Those who reported self-regulating their care were more likely to use the internet to schedule appointments with health care providers. Additionally, women who self-regulated their care were more likely to look up health information on the internet. This may relate to the fact that those who self-regulated care

may utilize these online resources as part of their self-regulating behaviors, for example, searching for suggestions to support self-regulating their care through self-medicating [7]. There are a number of potential reasons that an individual self-regulates care, such as trying to avoid medication side effects or trying to switch to alternative medication or treatment plans [54]. This type of behavior is critically important for individuals with IBD as self-management is a major aspect of the treatment plans [55]. Future work should further evaluate the underlying mechanisms that lead to individuals choosing to self-regulate their care and how the design of health information and internet-supported health tasks support those types of behaviors. Additionally, being dissatisfied with health care has been shown to influence the likelihood of using the internet for health information seeking [56,57]. Our study suggests those who were unsatisfied with their current health care and who were employed were less likely to look up health information online, the same was true for middle aged and older adults who were unsatisfied with their current health care. This may also relate to different information needs when trying to find a reasonable alternative treatment plan or trying to switch health providers [54].

Identifying factors that might impact the use of the internet for health-related tasks and health information searching can identify demographic and specific issues that might lead to targeted interventions and an examination of how online information is designed for and presented to these populations. According to Kittler et al [58], in 2004, 38% of physicians exchanged emails with their patients regularly, and Hobbs et al [59] found that approximately 37% of patients would have agreed to pay out of pocket to be able to communicate with their physicians by email. The estimates of email communication rates with health care providers are likely much higher today than in 2004. In fact, in 2015, a study of patient email communication with health providers suggested that the email use rate ranged from 18.7% to 50.7% among in 14 European countries and that men were found to be more likely to email health providers than women [41]. In our study, we found that 21.6% had emailed a health provider and that those who were frequent internet users were more likely to use email to communicate with their doctors, whereas married women with IBD were less likely to use email in this way. Future research should evaluate if there are other factors that impact the use of these services.

As expected, frequent internet users were shown to be more likely to use the internet to seek health information, schedule an appointment, and email health providers. In this study, we categorized frequent internet users as individuals who used the internet at least daily, yet many people currently use the internet on a more constant basis, and this variable may not capture differences between daily users and more constant users of the internet. Future research should more specifically examine the impact of internet usage frequency on how individuals with IBD use the internet for health care related activities. It would also be interesting to examine the frequency of internet use as a continuous variable and how that would impact the estimates of using the internet for health care tasks for those with IBD.

Limitations

There are several limitations of this study that should be addressed in future research. The focus of the NHIS survey was not specifically related to the use of the internet for health care–related tasks, nor was it specifically focused on individuals with IBD. Future work could specifically focus on this clinical population and on specific internet-related tasks. Additionally, with the frequent changes to health IT and in the adoption of health technology, it is possible that this survey did not capture some of the specific uses of technology for health-related purposes or possible technologies (eg, smartphones and health-related apps). There may also be other factors that influence the use of the internet for health-related activities that were not captured by the survey, and thus, were not included in this analysis. For example, some insurance companies require that their customers refill their medications online, a situation not captured by the survey. Nor were socioeconomic variables related to internet access included. Additionally, there are other factors that may impact the use of the internet in conducting health-related tasks (eg, mental health comorbidities, cognitive abilities, health literacy skills [60], complexity of the information search tasks, and credibility of target website [61]) that should be evaluated in future studies. The specific underlying mechanisms for self-regulating care, the way self-regulating care can be defined and implemented, and other related behaviors should be evaluated in future research.

In addition, to facilitate this analysis, most of the survey responses were categorized into binary variables that combined some answers with nonanswers and “I don’t know” responses. For example, internet use was transformed into a binary variable of frequent internet use versus infrequent internet use. These dichotomized variables may impact the findings associated with specific variables. Thus, future research could also examine the variables on a broader continuum in order to identify any additional nuances in the data. Additionally, future research should use different methods to identify why some relationships between variables were significant and also to identify the underlying causes so that future information strategies account for these differences and leverage what we know about the individuals with IBD and their internet health-related behaviors.

Conclusions

As the use of health information technology increases and evolves, it is critical to understand what specific clinical groups are using these resources, how they are doing so, and how those resources can best support health care self-management and disease prevention. This study examined using the internet for health information seeking tasks by individuals with IBD. As expected, frequent internet users were more likely to use the internet for health-related tasks. Our study demonstrates there are a number of factors and complex subgroups that impact the likelihood of individuals with IBD using the internet for information seeking. Future research should further investigate how these factors and groups (eg, women trying to self-regulate care) use the internet for health information and how the use of the internet shapes self-management of their health. Future research should also attempt to identify information design strategies and specific health-related task strategies for this

population. In addition, human factors studies should be conducted to identify if and how online resources can support these populations in ways that improve access to information and health outcomes.

Acknowledgments

The authors would like to thank Xiaoxia Li for her assistance with the preliminary data analysis for this manuscript. The authors would also like to thank the EASt lab for their feedback on a previous version of this paper.

Authors' Contributions

Both authors contributed equally to the manuscript.

Conflicts of Interest

None declared.

References

1. Mayer DK, Terrin NC, Kreps GL, Menon U, McCance K, Parsons SK, et al. Cancer survivors information seeking behaviors: a comparison of survivors who do and do not seek information about cancer. Patient Educ Couns 2007 Mar;65(3):342-350. [doi: 10.1016/j.pec.2006.08.015] [Medline: 17029864]
2. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de RSE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
3. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
4. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
5. Brochu F, Robins S, Miner SA, Grunberg PH, Chan P, Lo K, et al. Searching the internet for infertility information: a survey of patient needs and preferences. J Med Internet Res 2019 Dec 12;21(12):e15132 [FREE Full text] [doi: 10.2196/15132] [Medline: 31829963]
6. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128:151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]
7. Bessell TL, Anderson JN, Silagy CA, Sansom LN, Hiller JE. Surfing, self-medicating and safety: buying non-prescription and complementary medicines via the internet. Qual Saf Health Care 2003 Apr;12(2):88-92 [FREE Full text] [Medline: 12679503]
8. Claridy MD, Hudson MM, Caplan L, Mitby PA, Leisenring W, Smith SA, et al. Patterns of Internet-based health information seeking in adult survivors of childhood cancer. Pediatr Blood Cancer 2018 Dec;65(5):e26954 [FREE Full text] [doi: 10.1002/pbc.26954] [Medline: 29350454]
9. Cassell MM, Jackson C, Cheuvront B. Health communication on the Internet: an effective channel for health behavior change? J Health Commun 1998;3(1):71-79. [Medline: 10947375]
10. Kreps GL. The Relevance of Health Literacy to mHealth. Stud Health Technol Inform 2017;240:347-355. [Medline: 28972527]
11. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
12. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
13. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]
14. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. Prev Chronic Dis 2014 Apr 17;11:E62 [FREE Full text] [doi: 10.5888/pcd11.130389] [Medline: 24742395]
15. Weaver JB, Mays D, Weaver SS, Hopkins GL, Eroglu D, Bernhardt JM. Health information-seeking behaviors, health indicators, and health risks. Am J Public Health 2010 Aug;100(8):1520-1525. [doi: 10.2105/AJPH.2009.180521] [Medline: 20558794]
16. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
17. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
18. Park S, Go E. Health information seeking on the Internet: The role of involvement in searching for and assessing online health information. Health Mark Q 2016;33(4):327-341. [doi: 10.1080/07359683.2016.1238653] [Medline: 27841742]

19. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
20. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
21. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
22. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
23. Renahy E, Parizot I, Chauvin P. Health information seeking on the Internet: a double divide? Results from a representative survey in the Paris metropolitan area, France, 2005-2006. BMC Public Health 2008;8:69 [FREE Full text] [doi: 10.1186/1471-2458-8-69] [Medline: 18291024]
24. Hou SI. Health Literacy Online: A Guide to Writing and Designing Easy-to-Use Health Web Sites. Health Promot Pract 2012 Sep;13(5):577-580. [doi: 10.1177/1524839912446480] [Medline: 22763891]
25. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]
26. Kodagoda N, Wong B. Effects of low & high literacy on user performance in information search and retrieval. 2008 Presented at: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers:Culture, Creativity, Interaction; 2008 Sep 1; Swinton, UK, p. 173-181. [doi: 10.14236/ewic/HCI2008.17]
27. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat Clin Pract Gastroenterol Hepatol 2006 Jul;3(7):390-407. [doi: 10.1038/ncpgasthep0528] [Medline: 16819502]
28. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001 May 31;411(6837):599-603. [doi: 10.1038/35079107] [Medline: 11385576]
29. Best WR, Becktel JM, Singleton JW, Kern F. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976 Mar;70(3):439-444. [Medline: 1248701]
30. Boyapati R, Satsangi J, Ho G. Pathogenesis of Crohn's disease. F1000Prime Rep 2015;7:44 [FREE Full text] [doi: 10.12703/P7-44] [Medline: 26097717]
31. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol 2015 Dec;12(12):720-727. [doi: 10.1038/nrgastro.2015.150] [Medline: 26323879]
32. Loftus EV. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology 2004 May;126(6):1504-1517. [Medline: 15168363]
33. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2018 Dec 23;390(10114):2769-2778. [doi: 10.1016/S0140-6736(17)32448-0] [Medline: 29050646]
34. Colombel J. Measuring Disability in IBD: The IBD Disability Index. Gastroenterol Hepatol (N Y) 2013 May;9(5):300-302 [FREE Full text] [Medline: 23943665]
35. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001 May 31;411(6837):603-606. [doi: 10.1038/35079114] [Medline: 11385577]
36. Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 2004 May;36(5):476-480. [doi: 10.1038/ng1345] [Medline: 15107852]
37. Vagianos K, Clara I, Carr R, Graff LA, Walker JR, Targownik LE, et al. What Are Adults With Inflammatory Bowel Disease (IBD) Eating? A Closer Look at the Dietary Habits of a Population-Based Canadian IBD Cohort. JPEN J Parenter Enteral Nutr 2016 Mar;40(3):405-411. [doi: 10.1177/0148607114549254] [Medline: 25189173]
38. de Vries JHM, Dijkhuizen M, Tap P, Witteman BJM. Patient's Dietary Beliefs and Behaviours in Inflammatory Bowel Disease. Dig Dis 2019;37(2):131-139 [FREE Full text] [doi: 10.1159/000494022] [Medline: 30391940]
39. Promislow S, Walker JR, Taheri M, Bernstein CN. How well does the Internet answer patients' questions about inflammatory bowel disease? Can J Gastroenterol 2010 Nov;24(11):671-677 [FREE Full text] [Medline: 21157582]
40. Baumann E, Czerwinski F, Reifegerste D. Gender-Specific Determinants and Patterns of Online Health Information Seeking: Results From a Representative German Health Survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
41. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
42. Ryan C, Lewis J. Computer and Internet use in the United States. Am Community Surv Reports, ACS-37, US Census Bur Washington, DC. URL: https://www.census.gov/content/dam/Census/library/publications/2017/acs/acs-37.pdf [accessed 2019-02-28] [WebCite Cache ID 76WDpaBnl]

43. National Health Interview Survey 2011. CDC/National Center for Health Statistics. URL: https://www.cdc.gov/nchs/nhis/index.htm [accessed 2019-02-28] [WebCite Cache ID 76WFvKgnU]
44. CDC/National CFHS. National Health Interview Survey. 2016. URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2016/samadult_layout.pdf [accessed 2019-03-07] [WebCite Cache ID 76hUxRzB9]
45. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch Intern Med 2002 Nov 11;162(20):2269-2276. [Medline: 12418941]
46. Benjamin RM. Multiple chronic conditions: a public health challenge. Public Health Rep 2010;125(5):626-627 [FREE Full text] [Medline: 20873276]
47. Lumley T. Analysis of complex survey Samples. J Stat Soft 2004;9(8):1-19. [doi: 10.18637/jss.v009.i08]
48. Kennedy AP, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, et al. A randomised controlled trial to assess the effectiveness and cost of a patient orientated self management approach to chronic inflammatory bowel disease. Gut 2004 Nov;53(11):1639-1645 [FREE Full text] [doi: 10.1136/gut.2003.034256] [Medline: 15479685]
49. Casati J, Toner BB, de Rooy EC, Drossman DA, Maunder RG. Concerns of patients with inflammatory bowel disease: a review of emerging themes. Dig Dis Sci 2000 Jan;45(1):26-31. [doi: 10.1023/a:1005492806777] [Medline: 10695609]
50. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999 Jan;37(1):5-14. [Medline: 10413387]
51. Morrell RW, Mayhorn CB, Bennett J. A survey of world wide web use in middle-aged and older adults. Hum Factors 2000;42(2):175-182. [doi: 10.1518/001872000779656444] [Medline: 11022878]
52. Houston TK, Allison JJ. Users of Internet health information: differences by health status. J Med Internet Res 2002 Dec;4(2):E7 [FREE Full text] [doi: 10.2196/jmir.4.2.e7] [Medline: 12554554]
53. Ahadzadeh AS, Pahlevan SS, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
54. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
55. Plevinsky JM, Greenley RN, Fishman LN. Self-management in patients with inflammatory bowel disease: strategies, outcomes, and integration into clinical care. Clin Exp Gastroenterol 2016;9:259-267 [FREE Full text] [doi: 10.2147/CEG.S106302] [Medline: 27601930]
56. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15 Suppl 3:279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]
57. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
58. Kittler AF, Carlson GL, Harris C, Lippincott M, Pizziferri L, Volk LA, et al. Primary care physician attitudes towards using a secure web-based portal designed to facilitate electronic communication with patients. Inform Prim Care 2004;12(3):129-138. [Medline: 15606985]
59. Hobbs J, Wald J, Jagannath YS, Kittler A, Pizziferri L, Volk LA, et al. Opportunities to enhance patient and physician e-mail contact. Int J Med Inform 2003 Apr;70(1):1-9. [Medline: 12706177]
60. Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. Psychol Aging 2013 Dec;28(4):1124-1139. [doi: 10.1037/a0033911] [Medline: 24364414]
61. Hong T. Contributing factors to the use of health-related websites. J Health Commun 2006 Mar;11(2):149-165. [doi: 10.1080/10810730500526679] [Medline: 16537285]

Abbreviations

IBD: inflammatory bowel disease
OR: odds ratio
NHIS: National Health Interview Survey

Edited by G Eysenbach; submitted 03.07.19; peer-reviewed by J del Hoyo Francisco, T Ewais; comments to author 21.10.19; revised version received 28.01.20; accepted 26.07.20; published 24.09.20.

Please cite as:
Yin R, Neyens DM
Online Health Resource Use by Individuals With Inflammatory Bowel Disease: Analysis Using the National Health Interview Survey
J Med Internet Res 2020;22(9):e15352
URL: http://www.jmir.org/2020/9/e15352/
doi:10.2196/15352
PMID:32969831

Original Paper

Consumer-Grade Wearable Device for Predicting Frailty in Canadian Home Care Service Clients: Prospective Observational Proof-of-Concept Study

Ben Kim[1], PhD; Sandra M McKay[2,3], MBA, PhD; Joon Lee[4,5,6], PhD

[1]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

[2]VHA Home Healthcare, Toronto, ON, Canada

[3]School of Physical Therapy, University of Toronto, Toronto, ON, Canada

[4]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[5]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[6]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Joon Lee, PhD
Data Intelligence for Health Lab
Cumming School of Medicine
University of Calgary
3280 Hospital Drive NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 2968
Email: joonwu.lee@ucalgary.ca

Abstract

Background: Frailty has detrimental health impacts on older home care clients and is associated with increased hospitalization and long-term care admission. The prevalence of frailty among home care clients is poorly understood and ranges from 4.0% to 59.1%. Although frailty screening tools exist, their inconsistent use in practice calls for more innovative and easier-to-use tools. Owing to increases in the capacity of wearable devices, as well as in technology literacy and adoption in Canadian older adults, wearable devices are emerging as a viable tool to assess frailty in this population.

Objective: The objective of this study was to prove that using a wearable device for assessing frailty in older home care clients could be possible.

Methods: From June 2018 to September 2019, we recruited home care clients aged 55 years and older to be monitored over a minimum of 8 days using a wearable device. Detailed sociodemographic information and patient assessments including degree of comorbidity and activities of daily living were collected. Frailty was measured using the Fried Frailty Index. Data collected from the wearable device were used to derive variables including daily step count, total sleep time, deep sleep time, light sleep time, awake time, sleep quality, heart rate, and heart rate standard deviation. Using both wearable and conventional assessment data, multiple logistic regression models were fitted via a sequential stepwise feature selection to predict frailty.

Results: A total of 37 older home care clients completed the study. The mean age was 82.27 (SD 10.84) years, and 76% (28/37) were female; 13 participants were frail, significantly older (P<.01), utilized more home care service (P=.01), walked less (P=.04), slept longer (P=.01), and had longer deep sleep time (P<.01). Total sleep time (r=0.41, P=.01) and deep sleep time (r=0.53, P<.01) were moderately correlated with frailty. The logistic regression model fitted with deep sleep time, step count, age, and education level yielded the best predictive performance with an area under the receiver operating characteristics curve value of 0.90 (Hosmer-Lemeshow P=.88).

Conclusions: We proved that a wearable device could be used to assess frailty for older home care clients. Wearable data complemented the existing assessments and enhanced predictive power. Wearable technology can be used to identify vulnerable older adults who may benefit from additional home care services.

(J Med Internet Res 2020;22(9):e19732) doi:10.2196/19732

KEYWORDS

frailty; mobile health; wearables; physical activity; home care; prediction; predictive modeling, older adults; activities of daily living, sleep

Introduction

Frailty has detrimental health impacts among community-dwelling older adults. Frailty is associated with higher mortality [1-3], functional impairment [4,5], hospitalization [2,3], long-term care facility admission [3], and disability in activities of daily living [4]. Frailty also increases the demand on formal and informal caregivers, including home and community care services and family members [6]. A recent study [7] identified that caregiver burden can be predicted based on the physical frailty level of geriatric patients. Due to its significant impact on health outcomes and its burden on health care systems, improved screening and monitoring of frailty for community-dwelling older adults is deemed vital [8].

The prevalence of frailty among community-dwelling older adults is poorly understood. A systematic review [9] reported that its prevalence ranges between 4.0% and 59.1%; varying operational definitions and the heterogeneity of tools used in the studies resulted in a wide range of estimates. However, the prevalence range narrows to 4.0% to 17.0% when only the prevalence of physical phenotype frailty is aggregated by excluding social or cognitive deficits [9].

Both home and community health care are challenged with increased demand, primarily due to the aging population and emphasis on aging-in-place [10]. The demand for home and community health care service is expected to continue to rise in an effort to keep patients in their own community to reduce health care costs [11]. Screening and monitoring frailty in this population can benefit the home and community health care sector in multiple ways. Effective frailty intervention programs involve lifestyle changes including improving nutrition, increasing physical activity, and modifying the home environment [12]. Home and community health care clinicians are uniquely situated to deliver and monitor such interventions in a longitudinal manner, which can contribute to successful lifestyle changes. Screening for frailty at the community level can also help the home and community health care sector to identify vulnerable groups and allocate resources more efficiently [13].

Tools to screen community-dwelling older adults for frailty exist, but they have been used inconsistently and are often impractical or have been invalidated [14]. Wearable devices have been suggested as a potential tool to monitor frailty, and a few research studies [15-18] have explored this possibility. These studies explored the feasibility of using research-grade wearable devices, such as ActiGraph or independently developed wearable devices. These studies provide evidence for the internal construct validity of research-grade wearable devices to screen for frailty [19], as well as for a strong association between varying sleep quality parameters and frailty [20-22]. Consumer-grade wearable devices are a promising tool to monitor frailty as they have become smaller, cheaper, and ever more accessible in recent years [23], with older adults being the fastest growing group of wearable device users [24]. Research studies [25-27] have demonstrated the reliability of these devices for measuring step count, sleep quality, and heart rate compared to gold standard measures that are used in laboratory and clinical settings. Further validation studies demonstrated a high agreement between consumer-grade and medical-grade devices among specific populations, including patients with chronic obstructive pulmonary disease [28], pediatric patients [29], patients in intensive care units [30], and patients in cardiac rehabilitation [31].

Recognizing the need for an innovative solution to measure frailty in community-dwelling older adults, we set out to investigate the possibility of using consumer-grade wearable devices. We examined the data generated from a wearable device worn by home care clients to identify associations with frailty. We also aimed to identify key wearable device measures that can predict the status of frailty. Study procedure, tools, and statistical analyses are described. The results of the study are then presented, followed by a discussion where new findings are interpreted and compared to existing knowledge. The implications for frailty research studies, for wearable device research studies, and in home and community health care sectors, as well as the limitations of the study are presented.

Methods

Study Design

A prospective observational study was conducted to meet the study objectives. Participants were asked to wear a wearable device for a minimum of 8 days. At the end of the study, participants were assessed for frailty, activities of daily living, and level of comorbidity.

Recruitment

Home care clients in the Greater Toronto Area were recruited through VHA Home Healthcare from August 2018 to September 2019. VHA Home Healthcare is a home care agency that serves over 3000 clients throughout the Greater Toronto Area and other metropolitan areas in Ontario, Canada. Patients 55 years or older who had been receiving personal support service for more than 3 months were eligible for the study. Patients who were diagnosed with primary neuromuscular pathology, dependent on wheelchair, in an end-of-life program, or had cognitive impairments that could interfere with the use of wearable devices were excluded. Eligible home care patients were identified using VHA's electronic medical record system.

Wearable Device

The Xiaomi Mi Band Pulse 1S (Mi Band, hereafter) is a commercially available wearable device that is worn on the wrist. It uses a triaxial accelerometer to capture motions to approximate step count and sleep events. It is equipped with an optical heart rate sensor (photoplethysmography) to measure minute-by-minute heart rate. While the Mi Band can be worn on either the wrist or neck (as a pendant), its placement was

limited to the wrist for the study. The reliability and internal consistency of Mi Band's performance for measuring step count when walking and jogging has been validated [32,33]. Wrist-worn wearable devices displayed systematically lower heart rate during exercise, but the Mi Band demonstrated the highest accuracy [32].

We collected daily step count, light sleep time, deep sleep time, total sleep time, awake time, sleep quality, mean heart rate, and heart rate standard deviation. Sleep quality was calculated as the percentage of sleep duration over total sleep time; sleep duration was determined by subtracting awake time from total sleep time [34,35]. Heart rate was measured in beats per minute. A pool of 10 devices were used in rotation and sanitized throughout the study. The adherence to wearing the device was defined as 10 hours or more of wear time per day [36].

Frailty Assessment

Frailty was assessed using the Fried Frailty Index, a tool that has been developed for and used widely with community-dwelling older adults [1]. The Fried Frailty Index assesses phenotypic frailty based on 5 criteria: weight loss, exhaustion, slowness, weakness, and low physical activity. The index categorizes frailty into 3 stages based on the number of criteria that are met: nonfrail, prefrail, and frail corresponding to scores of 0, 1-2, and 3-5, respectively [1]. We dichotomized the Fried Frailty Index into a frail group for those with a score of 3 or higher and a nonfrail group for those with a score of 2 or lower [1].

Other Variables

Sociodemographic variables were collected using a short background questionnaire and through review of the patient's medical chart. These sociodemographic variables included age, sex, weight, height, ethnicity, level of education, income, and marital status. The level of comorbidity was assessed using the Charlson Comorbidity Index (CCI) [37]. The level of activities of daily living was assessed with the Katz index of independence [38]. The number of hours of service received per week was collected by reviewing the patient's medical chart.

Statistical Analysis

Descriptive statistics and univariate comparisons of means, medians, and proportions were performed to describe the sociodemographic information and patient assessments according to their frailty status. The level of education was condensed into 2 levels: high school (some or completed) and postsecondary. Household income was categorized into a lower income, those who earned $30,000 (approximately US $22,653) per year or less, and higher income, those who earned $30,000 or higher per year. Ethnicity was categorized into 2 levels: Caucasian and others which included aboriginal identity, Latin American, African American, South Asian, Southeast Asian, East Asian, Filipino, Arab, and West Asian.

Wearable device data were examined for participants adherence level, and days with less than 10 hours of wear time were excluded. Heart rate measurements of zero were generated when the device failed to have good skin contact. Such measurements were treated as missing and were removed from the analyses.

The Shapiro-Wilk test was performed to check for normality. To check for significant differences between patients who were frail and patients who were nonfrail, when the assumption of normal distribution was met, a two-tailed independent *t* test was used, while the Mann-Whitney *U* test was performed otherwise. The chi-square test was performed for categorical variables. The posthoc chi-square test was performed when significance was observed.

Pearson and Spearman correlation statistics were used to examine the relationship between frailty, sociodemographic information, patient assessments, and the data collected from the wearable devices.

Multiple logistic regression models were generated to predict frailty status. A sequential stepwise feature selection method was used to select the variables to be fitted into the models. The feature selection was used on the pool of sociodemographic and patient assessment variables to determine the features to be included in model 1. Model 2 was built by applying feature selection to the variables derived from the wearable device data. Model 3 used all available variables in a feature selection algorithm; the selected variables were used to build the logistic regression model. The Hosmer-Lemeshow test was performed to test the goodness-of-fit for each model. The predictive performance of each model was evaluated and compared using the area under the receiver operating characteristics curve (AUROC).

Statistical significance was set at α=.05 for all statistical results. The significance level for posthoc tests was corrected using the Bonferroni method. All statistical analyses were performed using R (version 3.6.0) in R studio (version 1.2.1335; R Studio Inc). Stepwise feature selection was performed using the function (stepAIC, version 7.3-51.4) from the MASS library [39].

Ethics, Consent, and Permissions

This study received ethics approval from the Office of Research Ethics Board at the University of Waterloo (ORE22842).

Results

Recruitment

A total of 72 older adults responded to the mailed recruitment brochure. All 72 older adults were contacted, and 45 agreed to participate in the study; 4 participants withdrew before completion of the 8-day study period. Data attrition due to technical issues resulted in data from 4 participants not being included. In total, 37 older home care clients were included in the study.

Participant Characteristics

Participants were 57 to 96 years of age, with a mean age of 82.23 (SD 10.84) years and 76% (28/37) were female (Table 1). The prevalence of frailty among the study population was 35% (13/37). On average, participants were observed for 9.43 (SD 1.99) days. Participants who were frail (mean age: 83.92 years) were significantly older (P<.001) than those who were nonfrail (mean age: 80.61 years). There was a significant difference in the income level between older adults who were

frail and those who were nonfrail (*P*=.03). Posthoc comparisons within each of the 3 income levels showed no statistical significance (low income: *P*=.93; mid to high income: *P*>.999) after correcting the α level with the Bonferroni method. Frail patients received significantly greater hours of home care services per week compared to the hours received by patients who were nonfrail (*P*=.01). The resulting *P* values of the Shapiro-Wilk normality tests are presented in Multimedia Appendix 1. The results of group difference tests are presented in Multimedia Appendix 2.

Table 1. Baseline sociodemographic and patient characteristics stratified by frailty status.

Characteristics	Frail (n=13)	Nonfrail (n=24)	*P* value
Age (years), mean (SD)	83.92 (9.66)	80.61 (13.96)	<.001[a]
Sex, n (%)			>.999[b]
Male	3 (23)	6 (25)	
Female	10 (77)	18 (75)	
BMI (kg/m^2), mean (SD)	26.96 (6.70)	28.54 (5.43)	.44[c]
ADL[d] score, mean (SD)	4.62 (1.45)	5.08 (0.88)	.43[a]
CCI[e] score, mean (SD)	1.92 (1.26)	1.25 (1.11)	.11[a]
Marital status, n (%)			.29[b]
Single	1 (8)	7 (29)	
Divorced or separated	2 (15)	5 (21)	
Widowed	4 (31)	7 (29)	
Currently married	6 (46)	5 (21)	
Education, n (%)			.12[b]
High school or less	8 (62)	7 (29)	
Postsecondary or higher	5 (38)	17 (71)	
Income, n (%)			.03[b]
Prefer not to answer	7 (54)	3 (12)	.06[f]
Low income	4 (31)	13 (54)	.93[f]
Mid to high income	2 (15)	8 (33)	>.999[f]
Ethnicity, n (%)			.71[b]
White	10 (77)	21 (88)	
Other	3 (23)	3 (12)	
Personal support service, hours per week	5.15 (3.51)	2.77 (1.85)	.01[a]

[a]Mann-Whitney *U* test was used.

[b]Chi-square test was used.

[c]An independent *t* test was used.

[d]ADL: activities of daily living; Katz index of independence was used.

[e]CCI: Charlson Comorbidity Index.

[f]Posthoc chi-square test was used.

Frailty and Wearable Device Data

On average, older adults wore the device for 20.03 (1.64) hours per day (Table 2). Home care clients who were frail reported significantly lower daily step counts than their nonfrail counterparts did (mean steps per day: 367.11 vs. 1023.95, respectively; *P*=.04). Total sleep time (*P*=.01) and deep sleep time (*P*<.01) were significantly longer for older adults who were frail, but no difference was found for light sleep time (*P*=.04). No difference was found for heart rate measures. Box plots corresponding to Table 2 are presented in Multimedia Appendix 3.

Table 2. Difference in the data collected from the wearable device between frail and nonfrail participants.

Measures	Frail (n=13), mean (SD)	Nonfrail (n=24), mean (SD)	*P* value
Worn time (hours per day)	20.66 (1.03)	19.69 (1.82)	.16[a]
Daily step count	367.11 (272.63)	1023.95 (863.83)	.04[a]
Sleep measures			
Deep sleep time (minutes)	138.90 (64.00)	75.65 (39.12)	<.001[a]
Light sleep time (minutes)	350.88 (130.56)	312.78 (82.32)	.35[b]
Total sleep time (minutes)	489.78 (139.54)	388.44 (93.28)	.01[a]
Awake time (minutes)	36.03 (24.27)	65.05 (57.97)	.17[a]
Sleep quality (%)	92.48 (5.62)	78.95 (26.53)	.08[a]
Heart rate measures			
Heart rate (bpm)	82.77 (10.25)	77.43 (8.66)	.13[b]
Heart rate SD (bpm)	22.12 (7.61)	18.78 (4.54)	.17[b]

[a]Mann-Whitney *U* test was used.

[b]An independent *t* test was used.

Factors Correlated With Frailty

The correlation between wearable data and frailty is summarized in Table 3. Daily step count was negatively correlated with frailty level (r=–0.52, P<.001). All 5 sleep measures were moderately correlated with frailty. Education level was moderately correlated with frailty status (r=–0.40, P=.02). No relationship was found between heart rate measures and frailty status.

Table 3. Correlations between wearable device data, patient characteristics, and frailty.

	Frailty	
	Correlation coefficient	*P* value
Daily step count	–0.52	.001
Sleep measures		
Total sleep time	0.52	.001
Deep sleep time	0.47	.003
Light sleep time	0.35	.03
Sleep quality	0.56	<.001
Awake time	–0.54	<.001
Heart rate measures		
Mean heart rate	0.11	.54
Heart rate SD	–0.25	.16
Sociodemographic		
Age	0.29	.08
Sex	0.074	.66
BMI	–0.068	.69
Income level	–0.066	.74
Education level	–0.40	.02
Patient assessments		
ADL[a] score	–0.18	.27
CCI[b] score	0.16	.33
Personal support hours	0.23	.17

[a]ADL: activities of daily living; Katz index of independence was used.

[b]CCI: Charlson Comorbidity Index.

Frailty Prediction

Model Description

A total of 3 multiple variable logistic regression models were fitted to predict frailty with the sociodemographic variables, patient assessments, and wearable data (Table 4). Income was excluded from the feature selection method since a high number of participants declined to answer. Model 1 formulation began by fitting the sociodemographic variables and patient assessments. The feature selection method resulted in a model that contains CCI and education level. Model 2 used variables derived from the wearable device data only. The resulting model was fitted with step count, deep sleep time, awake time, and heart rate standard deviation. Model 3 used all available variables and was fitted with deep sleep time, step count, age, and education level.

Table 4. Three frailty prediction models and the variables selected by the stepwise feature selection method.

Models	Variable pool	Selected variables
Model 1	Sociodemographic and patient assessment variables	CCI[a], education level
Model 2	Wearable device–derived variables	Step count, deep sleep time, light sleep time, heart rate standard deviation
Model 3	Sociodemographic, patient assessment, and wearable device–derived variables	Deep sleep time, step count, age, education level

[a]CCI: Charlson Comorbidity Index.

Model Evaluation

Table 5 shows the results of multiple logistic regression analyses and the factors predictive of frailty. Model 1 showed no significant association. For model 2, deep sleep time was a significant predictor of frailty (P<.01). Increasing deep sleep time was significantly associated with increased odds of frailty (adjusted odds ratio [OR] 1.02, 95% CI 1.01-1.05, P<.01). For model 3, deep sleep time (P=.02) and age (P=.03) were significant predictors. Increasing deep sleep time was associated

with an increase in the odds of frailty (adjusted OR 1.03, 95% CI 1.01-1.07, *P*=.02), whereas increasing age was associated with a decrease in the odds of frailty (adjusted OR 0.90, 95% CI 0.80-0.99, *P*=.03).

All 3 models were evaluated for their goodness of fit using the Hosmer-Lemeshow statistic. Overall, no model showed statistical significance on this test, indicating they had acceptable goodness-of-fit, and the predicted frailty matched the observed frailty status (Table 6).

When the predictive performance was evaluated by AUROC, all 3 models showed medium to high values. Model 1 (AUROC 0.77), based on sociodemographic and patient assessment variables, was outperformed by model 2 (AUROC 0.88), which was fitted with wearable device variables. Model 3 (AUROC 0.90) had the best predictive performance (Table 6). The receiver operating characteristic curves are shown in Figure 1 for each model.

Table 5. Multiple logistic regression of factors associated with frailty.

Model and variables	Adjusted OR[a] (95 % CI)	*P* value
Model 1		
CCI[b]	1.78 (0.95, 3.66)	.09
Education level—high school or below	reference	—
Education level—postsecondary education or higher	0.22 (0.04, 0.96)	.05
Model 2		
Step count	1.00 (1.00, 1.00)	.17
Deep sleep time	1.02 (1.01, 1.05)	.02
Awake time	0.97 (0.93, 1.01)	.18
Heart rate standard deviation	1.17 (0.99, 1.46)	.10
Model 3		
Deep sleep time	1.03 (1.01, 1.07)	.04
Step count	1.00 (1.00, 1.00)	.06
Age	0.90 (0.80, 0.99)	.04
Education level—high school or less	reference	—
Education level—postsecondary education or higher	0.11 (0.01, 0.94)	.06

[a]OR: odds ratio.

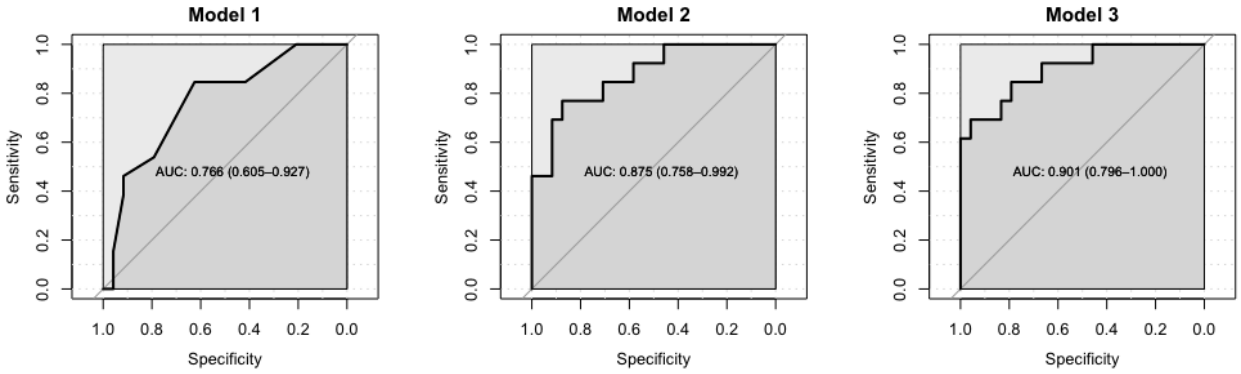
[b]CCI: Charlson Comorbidity Index.

Table 6. Summary of model performance in predicting frailty status.

Models	Accuracy	Sensitivity	Specificity	AUROC[a]	Hosmer-Lemeshow test *P* value
Model 1: Sociodemographic and patient assessment variables	0.76	0.46	0.92	0.77	0.73
Model 2:Wearable device derived variables	0.81	0.69	0.88	0.88	0.95
Model 3: All variables from models 1 and 2	0.81	0.69	0.88	0.90	0.85

[a]AUROC: area under the receiver operating characteristics curve.

Figure 1. The receiver operating characteristics curves (with area under the curve) for all models fitted to predict frailty. AUC: area under the curve.



Discussion

Principal Findings

The growing aging population in Canada and the emphasis on aging-in-place call for innovative ways to improve efficiency in the home and community health care sector. There is an increasing interest in integrating information and communication technology such as consumer-grade wearable devices into health care delivery due to their rising popularity, ease-of-use, and the potential usefulness of continuously collected data [40]. The aim of this study was to investigate the possibility of assessing and predicting frailty using a wearable device.

We observed 37 older home care clients for a minimum of 8 days. The prevalence of frailty in the study sample, 35% (13/37), was similar to that found in other research studies examining home care clients [3,41]. Many research studies [9,42] reported a significantly higher prevalence of frailty in older women compared to prevalence in older men, but this was not observed in our study sample. However, another study [3] that examined the same population did not find any significant difference between the sexes. Overall, the study sample seemed reasonably representative of the home care population. Previous research studies [1,43] reported an association between income and education level and frailty. Our study sample had significantly different income levels between the 2 frailty groups (P=.03). However, the posthoc chi-square analysis results did not reach statistical significance (low income: P=.93; mid to high income: P>.999). Education level was moderately correlated with frailty level (r=–0.40, P=.02). Overall, our study sample displayed the general characteristics of frail populations [1,43].

Our study found a significantly higher utilization of home care service by older adults who were frail compared to utilization by older adults who were nonfrail (mean hours per week: 5.15 vs 2.77; P=.01). Unfortunately, the current system fails to meet all care needs of home care clients as indicated by the increased hours of informal care and caregiver distress for the home care clients with more severe frailty [13]. Resulting adverse health outcomes and increased health care utilization [3] highlight the need for a better allocation of home care service to those who stand to benefit the most.

In our study sample, older adults who were nonfrail walked significantly more than the older adults who were frail. This result is in line with the findings of previous research studies where reduced daily step count and physical activity were observed for frail community-dwelling older adults [44] and ICU patients [45]. In a previous study [46], daily step count was significantly related to frailty. Our study extended this evidence outside the controlled settings and beyond 24-hour monitoring period [47,48], and demonstrated the relationship in an unsupervised setting.

Sleep measures including longer total sleep, deep sleep, and light sleep durations; awake time; and sleep quality were shown to be related to more severe frailty. This is contrary to the common knowledge of deterioration of sleep quality and quantity with aging [49]. However, in epidemiological studies, a longer sleep duration was associated with an increased risk of heart disease and all-cause mortality [50]. The lowest mortality risk was found for those who sleep about 7 hours a night [51], while men who slept more than 8 hours per day had a tripled risk of heart disease [52]. This relationship was shown in our study sample where older adults who were nonfrail and older adults who were frail had significantly different total sleep durations (P=.01). Older adults who were nonfrail had a mean total sleep duration of 6.48 hours (close to 7 hours), while their frail counterparts slept for 8.16 hours. These findings demonstrate the additional information wearable devices provide over conventional sleep quality assessments.

In this study, we built logistic regression models using a sequential stepwise feature selection method. Feature selection in general can help improve predictive performance [53]. It minimizes the number of features needed in a model, which was critical given the small sample size of this study. While manual feature selection based on expert knowledge could have been a feasible alternative, our goal was to maximize frailty prediction performance in our data set by utilizing an empirical feature selection method. The analysis of multiple logistic regression models showed that wearable device data were a superior source of information for predicting frailty than sociodemographic information and patient assessments. However, the highest AUROC of 0.90 was achieved with the model that used wearable device data, sociodemographic, and patient assessment information. Previously, a similar study [16] that used a neck-worn wearable device to obtain step count and physical activity-related variables achieved an AUROC of 0.88 in discriminating the prefrail group from the frail and nonfrail groups. Another study [48] used 2 research-grade wearable devices concurrently and achieved an AUROC of 0.86 in

discriminating 3 frailty states using stride length. Both studies were limited due to their short 48-hour observational period and being conducted in a laboratory setting. Our study demonstrated that unsupervised monitoring of frailty at home using a wearable device is possible. Our results corroborate that wearable technology should complement, rather than replace, the existing practice [54].

Many mobile health and telehealth apps have been successful at delivering health care while improving efficiency [55]. A study [56] that examined telehealth for frail older adults found the most cost-effective telehealth program used automated monitoring of vital signs to reduce health service use and facilitate remote follow-up. Wearable devices are becoming increasingly affordable and are capable of offering a similar use case as telehealth apps with their automated monitoring of physical activities, sleep, and heart rate. The range of information collected from wearable devices are also increasing with the advancement of sensor technology such as electrocardiogram, blood glucose level, oxygen saturation level, and electrodermal activity. When coupled with well-calibrated algorithms that enable early detection of health deteriorations such as frailty, cost savings can be further increased. The added value of wearable devices in assessing frailty for home care clients and community-dwelling older adults should be carefully evaluated for their feasibility in real-life settings. Each home or community health care system is unique, including but not limited to their target population, geographical area, and funding structure. Future research should consider these factors when evaluating the clinical value and cost savings of wearable devices.

Future research should confirm the predictive power of data derived from wearable devices and extend it beyond the home and community care sector. Our results indicated that wearable devices are a valid tool when an adequate analytical process is used. We recommend that future home care research studies leverage the potential of consumer-grade wearable devices to help identify vulnerable and frail groups who may benefit from additional home care services and increased access to health care.

Limitations

Our study has several limitations. First, the small study sample prevented us from stratifying patients into nonfrail, prefrail, and frail groups. A third frailty state could have helped us demonstrate gradient measures of wearable data. The small sample size also limited the number of variables that could be used in developing multiple logistic regression models. The 3 logistic regression models were each fitted with 2 to 4 features. They exceeded the common rule of 1-in-10 and which may have increased the risk of overfitting [57]. The small sample size precluded partitioning our data into training and test sets. As a result, the reported predictive performance overestimated the performance that would be found on a different sample of older adults. A further caution should be taken when interpreting the results of the Hosmer-Lemeshow test due to the small sample size.

Our research used an 8-day observation period. While this was longer than the observation periods of most other studies using wearable devices, an even longer observational period may be required to reveal new patterns that are not observable within 8 days such as weekdays versus weekends and seasonal differences. Lastly, the validation studies that examined the Mi Band [32,33] were conducted in younger participants, limiting their generalizability to older adults of this study.

Conclusions

In this study, we proved the concept of using a wrist-worn consumer-grade wearable device to assess frailty among older home care clients. Data collected from the wearable device, such as total sleep time and deep sleep time, were associated with frailty. The frailty prediction model based on variables selected from wearable devices, sociodemographic variable, and patient assessment variables achieved the highest AUROC of 0.90, compared to the AUROC of the other predictive models that either used only sociodemographic and assessment variables or only wearable device–derived variables.

Acknowledgments

This work was supported by a Core Research Project Grant from the AGE-WELL Network of Centres of Excellence (WP7.3).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The results of Shapiro-Wilk normality tests for all continuous variables.
[DOCX File , 15 KB - jmir_v22i9e19732_app1.docx]

Multimedia Appendix 2
T test statistics and chi-square test statistics for comparisons between the frail and nonfrail participants with respect to baseline sociodemographic and patient characteristics (n=37).
[DOCX File , 17 KB - jmir_v22i9e19732_app2.docx]

Multimedia Appendix 3

Boxplots of the wearable device data comparing frail and nonfrail participants.
[DOCX File , 181 KB - jmir_v22i9e19732_app3.docx]

References

1. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [Medline: 11253156]
2. Shamliyan T, Talley KMC, Ramakrishnan R, Kane RL. Association of frailty with survival: a systematic literature review. Ageing Res Rev 2013 Mar;12(2):719-736. [doi: 10.1016/j.arr.2012.03.001] [Medline: 22426304]
3. Campitelli MA, Bronskill SE, Hogan DB, Diong C, Amuah JE, Gill S, et al. The prevalence and health consequences of frailty in a population-based older home care cohort: a comparison of different measures. BMC Geriatr 2016 Jul 07;16:133 [FREE Full text] [doi: 10.1186/s12877-016-0309-z] [Medline: 27388294]
4. Vermeulen J, Neyens JCL, van RE, Spreeuwenberg MD, de WLP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. BMC Geriatr 2011 Jul;11:33 [FREE Full text] [doi: 10.1186/1471-2318-11-33] [Medline: 21722355]
5. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. J Gerontol A Biol Sci Med Sci 2004 Mar;59(3):255-263. [Medline: 15031310]
6. Sinha M. Portrait of caregivers, 2012. Statistics Canada Catalogue no. 89-652-X. Ottawa: Statistics Canada; 2013 Sep. URL: https://www150.statcan.gc.ca/n1/pub/89-652-x/89-652-x2013001-eng.pdf [accessed 2018-06-13] [WebCite Cache ID 709i6D5LK]
7. Ringer TJ, Hazzan AA, Kennedy CC, Karampatos S, Patterson C, Marr S, et al. Care recipients' physical frailty is independently associated with subjective burden in informal caregivers in the community setting: a cross-sectional study. BMC Geriatr 2016 Nov 17;16(1):186 [FREE Full text] [doi: 10.1186/s12877-016-0355-6] [Medline: 27855633]
8. Muscedere J, Andrew MK, Bagshaw SM, Estabrooks C, Hogan D, Holroyd-Leduc J, et al. Screening for Frailty in Canada's Health Care System: A Time for Action. Can. J. Aging 2016 May 23;35(3):281-297. [doi: 10.1017/s0714980816000301]
9. Collard RM, Boter H, Schoevers RA, Oude VRC. Prevalence of frailty in community-dwelling older persons: a systematic review. J Am Geriatr Soc 2012 Aug;60(8):1487-1492. [doi: 10.1111/j.1532-5415.2012.04054.x] [Medline: 22881367]
10. Home health care and related services, 2017. Statistics Canada Catalogue no. 11-001-X. Ottawa: Statistics Canada; 2019 31. URL: https://www150.statcan.gc.ca/n1/en/daily-quotidien/190731/dq190731c-eng.pdf?st=siWEGqn7 [accessed 2020-05-05]
11. Integration through information communication technology for home care in Canada: Final Report. Canadian Home Care Association. Ottawa: The Canadian Home Care Association; 2008 Mar. URL: https://cdnhomecare.ca/wp-content/uploads/2020/03/Information-Technology-in-Home-Care-March-2008-English.pdf [accessed 2019-04-21]
12. Puts MTE, Toubasi S, Andrew MK, Ashe MC, Ploeg J, Atkinson E, et al. Interventions to prevent or reduce the level of frailty in community-dwelling older adults: a scoping review of the literature and international policies. Age Ageing 2017 May 01;46(3):383-392 [FREE Full text] [doi: 10.1093/ageing/afw247] [Medline: 28064173]
13. Maxwell CJ, Campitelli MA, Diong C, Mondor L, Hogan DB, Amuah JE, et al. Variation in the health outcomes associated with frailty among home care clients: relevance of caregiver distress and client sex. BMC Geriatr 2018 Sep 12;18(1):211 [FREE Full text] [doi: 10.1186/s12877-018-0899-8] [Medline: 30208884]
14. Dwyer JT, Gahche JJ, Weiler M, Arensberg MB. Screening Community-Living Older Adults for Protein Energy Malnutrition and Frailty: Update and Next Steps. J Community Health 2019 Sep 30;45(3):640-660. [doi: 10.1007/s10900-019-00739-1]
15. Rumer KK, Saraswathula A, Melcher ML. Prehabilitation in our most frail surgical patients. Current Opinion in Organ Transplantation 2016;21(2):188-193. [doi: 10.1097/mot.0000000000000295]
16. Razjouyan J, Naik A, Horstman M, Kunik M, Amirmazaheri M, Zhou H, et al. Wearable Sensors and the Assessment of Frailty among Vulnerable Older Adults: An Observational Cohort Study. Sensors 2018 Apr 26;18(5):1336. [doi: 10.3390/s18051336]
17. Mohler MJ, Wendel CS, Taylor-Piliae RE, Toosizadeh N, Najafi B. Motor Performance and Physical Activity as Predictors of Prospective Falls in Community-Dwelling Older Adults by Frailty Level: Application of Wearable Technology. Gerontology 2016 Apr 30;62(6):654-664. [doi: 10.1159/000445889]
18. Parvaneh S, Mohler J, Toosizadeh N, Grewal GS, Najafi B. Postural Transitions during Activities of Daily Living Could Identify Frailty Status: Application of Wearable Technology to Identify Frailty during Unsupervised Condition. Gerontology 2017 Mar 11;63(5):479-487. [doi: 10.1159/000460292]
19. Chen S, Honda T, Chen T, Narazaki K, Haeuchi Y, Supartini A, et al. Screening for frailty phenotype with objectively-measured physical activity in a west Japanese suburban community: evidence from the Sasaguri Genkimon Study. BMC Geriatr 2015 Apr 02;15:36 [FREE Full text] [doi: 10.1186/s12877-015-0037-9] [Medline: 25887474]
20. Nóbrega PVDN, Maciel ACC, de Almeida Holanda CM, Oliveira Guerra R, Araújo JF. Sleep and frailty syndrome in elderly residents of long-stay institutions: a cross-sectional study. Geriatr Gerontol Int 2014 Jul;14(3):605-612. [doi: 10.1111/ggi.12144] [Medline: 24020502]

21. Ensrud KE, Blackwell TL, Redline S, Ancoli-Israel S, Paudel ML, Cawthon PM, Osteoporotic Fractures in Men Study Group. Sleep disturbances and frailty status in older community-dwelling men. J Am Geriatr Soc 2009 Nov;57(11):2085-2093 [FREE Full text] [doi: 10.1111/j.1532-5415.2009.02490.x] [Medline: 19793160]
22. Vaz Fragoso CA, Gahbauer EA, Van Ness PH, Gill TM. Sleep-wake disturbances and frailty in community-living older persons. J Am Geriatr Soc 2009 Nov;57(11):2094-2100 [FREE Full text] [doi: 10.1111/j.1532-5415.2009.02522.x] [Medline: 19793356]
23. Baig MM, Afifi S, GholamHosseini H, Mirza F. A Systematic Review of Wearable Sensors and IoT-Based Monitoring Applications for Older Adults - a Focus on Ageing Population and Independent Living. J Med Syst 2019 Jun 15;43(8):233. [doi: 10.1007/s10916-019-1365-7] [Medline: 31203472]
24. Older Americans Drive Growth of Wearables Internet. eMarketer. 2018 Dec 13. URL: https://www.emarketer.com/content/older-americans-drive-growth-of-wearables [accessed 2019-10-11]
25. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
26. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. J Pers Med 2017 May 24;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
27. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of Wrist-Worn Heart Rate Monitors. JAMA Cardiol 2016 Oct 12:104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
28. Vooijs M, Alpay LL, Snoeck-Stroband JB, Beerthuizen T, Siemonsma PC, Abbink JJ, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. Interact J Med Res 2014;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]
29. Voss C, Gardner RF, Dean P, Harris K. Validity of Commercial Activity Trackers in Children With Congenital Heart Disease. Can J Cardiol 2017 Jun;33(6):799-805. [doi: 10.1016/j.cjca.2016.11.024] [Medline: 28347581]
30. Kroll RR, Boyd JG, Maslove DM. Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
31. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Sep;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
32. Ricchio K, Lyter-Antonneau P, Palao J. Reliability of Fitness Trackers at Different Prices for Measuring Steps and Heart Rate: a Pilot Study. Central European Journal of Sport Sciences and Medicine 2018;24:57-64. [doi: 10.18276/cej.2018.4-06]
33. Paradiso C, Colino F, Liu S. The Validity and Reliability of the Mi Band Wearable Device for Measuring Steps and Heart Rate. Int J Exerc Sci 2020;13(4):689-701 [FREE Full text] [Medline: 32509127]
34. Kroll RR, McKenzie ED, Boyd JG, Sheth P, Howes D, Wood M, WEARable Information Technology for hospital INpatients (WEARIT-IN) study group. Use of wearable devices for post-discharge monitoring of ICU patients: a feasibility study. J Intensive Care 2017;5:64 [FREE Full text] [doi: 10.1186/s40560-017-0261-9] [Medline: 29201377]
35. Duncan MJ, Vandelanotte C, Trost SG, Rebar AL, Rogers N, Burton NW, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. BMC Public Health 2016 Dec 30;16:670 [FREE Full text] [doi: 10.1186/s12889-016-3256-x] [Medline: 27473327]
36. Tudor-Locke C, Barreira TV, Schuna JM, Mire EF, Chaput J, Fogelholm M, ISCOLE Research Group. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). Int J Behav Nutr Phys Act 2015 Feb 11;12:11 [FREE Full text] [doi: 10.1186/s12966-015-0172-x] [Medline: 25881074]
37. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? Med Care 1996 Jan;34(1):73-84. [doi: 10.1097/00005650-199601000-00006] [Medline: 8551813]
38. Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. Gerontologist 1970;10(1):20-30. [Medline: 5420677]
39. Venables W, Ripley B. Modern Applied Statistics with S, 4th Ed. New York: Springer-Verlag; 2002.
40. National Information Board. Personalised Health and Care 2020. Digital Health. 2014 Nov. URL: https://www.digitalhealth.net/includes/images/news0254/PDF/0172_NHS_England_NIB_Report_WITH_ADDITIONAL_MATERIAL_S8.pdf [accessed 2018-10-17] [WebCite Cache ID 73EQIHsy7]
41. Kelly S, O'Brien I, Smuts K, O'Sullivan M, Warters A. Prevalence of frailty among community dwelling older adults in receipt of low level home support: a cross-sectional analysis of the North Dublin Cohort. BMC Geriatr 2017 Jun 07;17(1):121 [FREE Full text] [doi: 10.1186/s12877-017-0508-2] [Medline: 28592236]
42. Avila-Funes JA, Helmer C, Amieva H, Barberger-Gateau P, Le Goff M, Ritchie K, et al. Frailty among community-dwelling elderly people in France: the three-city study. J Gerontol A Biol Sci Med Sci 2008 Oct;63(10):1089-1096 [FREE Full text] [doi: 10.1093/gerona/63.10.1089] [Medline: 18948560]

43. St John PD, Montgomery PR, Tyas SL. Social position and frailty. Can J Aging 2013 Sep;32(3):250-259. [doi: 10.1017/S0714980813000329] [Medline: 23962524]
44. Cavanaugh JT, Coleman KL, Gaines JM, Laing L, Morey MC. Using step activity monitoring to characterize ambulatory activity in community-dwelling older adults. J Am Geriatr Soc 2007 Jan;55(1):120-124. [doi: 10.1111/j.1532-5415.2006.00997.x] [Medline: 17233695]
45. McNelly AS, Rawal J, Shrikrishna D, Hopkinson NS, Moxham J, Harridge SD, et al. An Exploratory Study of Long-Term Outcome Measures in Critical Illness Survivors: Construct Validity of Physical Activity, Frailty, and Health-Related Quality of Life Measures. Crit Care Med 2016 Jun;44(6):e362-e369. [doi: 10.1097/CCM.0000000000001645] [Medline: 26974547]
46. Theou O, Jakobi JM, Vandervoort AA, Jones GR. A comparison of physical activity (PA) assessment tools across levels of frailty. Arch Gerontol Geriatr 2012;54(3):e307-e314. [doi: 10.1016/j.archger.2011.12.005] [Medline: 22209454]
47. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
48. Schwenk M, Mohler J, Wendel C, D'Huyvetter K, Fain M, Taylor-Piliae R, et al. Wearable sensor-based in-home assessment of gait, balance, and physical activity for discrimination of frailty status: baseline results of the Arizona frailty cohort study. Gerontology 2015;61(3):258-267 [FREE Full text] [doi: 10.1159/000369095] [Medline: 25547185]
49. Espiritu JRD. Aging-Related Sleep Changes. Clinics in Geriatric Medicine 2008 Feb;24(1):1-14. [doi: 10.1016/j.cger.2007.08.007]
50. Gallicchio L, Kalesan B. Sleep duration and mortality: a systematic review and meta-analysis. J Sleep Res 2009 Jun;18(2):148-158 [FREE Full text] [doi: 10.1111/j.1365-2869.2008.00732.x] [Medline: 19645960]
51. Tamakoshi A, Ohno Y, JACC Study Group. Self-reported sleep duration as a predictor of all-cause mortality: results from the JACC study, Japan. Sleep 2004 Feb 01;27(1):51-54. [Medline: 14998237]
52. Burazeri G, Gofin J, Kark J. Over 8 hours of sleep--marker of increased mortality in Mediterranean population: follow-up population study. Croat Med J 2003 Apr;44(2):193-198 [FREE Full text] [Medline: 12698511]
53. Kassambara A. Machine Learning Essentials: Practical Guide in R. eBook: STHDA; 2008.
54. Fasano A, Mancini M. Wearable-based mobility monitoring: the long road ahead. Lancet Neurol 2020 May;19(5):378-379. [doi: 10.1016/S1474-4422(20)30033-8] [Medline: 32059810]
55. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017 Feb;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
56. Barlow J, Singh D, Bayer S, Curry R. A systematic review of the benefits of home telecare for frail elderly people and those with long-term conditions. J Telemed Telecare 2016 Jun 23;13(4):172-179. [doi: 10.1258/135763307780908058]
57. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [Medline: 8970487]

Abbreviations

AUROC: area under the receiver operating characteristics curve
CCI: Charlson comorbidity index

Edited by G Eysenbach; submitted 30.04.20; peer-reviewed by S Tedesco, M Bardus; comments to author 25.05.20; revised version received 04.07.20; accepted 14.07.20; published 03.09.20.

Please cite as:
Kim B, McKay SM, Lee J
Consumer-Grade Wearable Device for Predicting Frailty in Canadian Home Care Service Clients: Prospective Observational Proof-of-Concept Study
J Med Internet Res 2020;22(9):e19732
URL: https://www.jmir.org/2020/9/e19732
doi:10.2196/19732
PMID:32880582

Original Paper

Relationship Between Chronic Stress and Heart Rate Over Time Modulated by Gender in a Cohort of Office Workers: Cross-Sectional Study Using Wearable Technologies

Alex Wilhelmus Jacobus van Kraaij[1,2,3], MSc; Giuseppina Schiavone[1,2], PhD; Erika Lutin[4,5], MSc; Stephan Claes[6], Prof Dr; Chris Van Hoof[1,2,4,5], Prof Dr

[1]Holst Centre, imec-the Netherlands, Eindhoven, Netherlands

[2]OnePlanet Research Center, imec-the Netherlands, Wageningen, Netherlands

[3]Faculty of Natural Sciences, Math and Informatics (FNWI), Radboud University, Nijmegen, Netherlands

[4]Electrical Engineering-ESAT, KU Leuven, Leuven, Belgium

[5]imec-Belgium, Heverlee, Belgium

[6]University Psychiatric Center & Department of Neurosciences, KU Leuven, Leuven, Belgium

Corresponding Author:
Alex Wilhelmus Jacobus van Kraaij, MSc
OnePlanet Research Center
imec-the Netherlands
Bronland 10
Wageningen, 6708 WH
Netherlands
Phone: 31 404020400
Email: alex.vankraaij@imec.nl

Abstract

Background: Chronic stress is increasing in prevalence and is associated with several physical and mental disorders. Although it is proven that acute stress changes physiology, much less is known about the relationship between physiology and long-term stress. Continuous measurement of vital signs in daily life and chronic stress detection algorithms could serve this purpose. For this, it is paramount to model the effects of chronic stress on human physiology and include other cofounders, such as demographics, enabling the enrichment of a population-wide approach with individual variations.

Objective: The main objectives of this study were to investigate the effect of chronic stress on heart rate (HR) over time while correcting for weekdays versus weekends and to test a possible modulation effect by gender and age in a healthy cohort.

Methods: Throughout 2016 and 2017, healthy employees of technology companies were asked to participate in a 5-day observation stress study. They were required to wear two wearables, of which one included an electrocardiogram sensor. The derived HR was averaged per hour and served as an output for a mixed design model including a trigonometric fit over time with four harmonics (periods of 24, 12, 8, and 6 hours), gender, age, whether it was a workday or weekend day, and a chronic stress score derived from the Perceived Stress Scale (PSS) as predictors.

Results: The study included 328 subjects, of which 142 were female and 186 were male participants, with a mean age of 38.9 (SD 10.2) years and a mean PSS score of 13.7 (SD 6.0). As main effects, gender (χ^2_1=24.02, P<.001); the hour of the day (χ^2_1=73.22, P<.001); the circadian harmonic (χ^2_2=284.4, P<.001); and the harmonic over 12 hours (χ^2_2=242.1, P<.001), over 8 hours (χ^2_2=23.78, P<.001), and over 6 hours (χ^2_2=82.96, P<.001) had a significant effect on HR. Two three-way interaction effects were found. The interaction of age, whether it was a workday or weekend day, and the circadian harmonic over time were significantly correlated with HR (χ^2_2=7.13, P=.03), as well as the interaction of gender, PSS score, and the circadian harmonic over time (χ^2_2=7.59, P=.02).

Conclusions: The results show a relationship between HR and the three-way interaction of chronic stress, gender, and the circadian harmonic. The modulation by gender might be related to evolution-based energy utilization strategies, as suggested in related literature studies. More research, including daily cortisol assessment, longer recordings, and a wider population, should

be performed to confirm this interpretation. This would enable the development of more complete and personalized models of chronic stress.

(J Med Internet Res 2020;22(9):e18253) doi:10.2196/18253

KEYWORDS

chronic stress; heart rate; circadian rhythm; gender; age; wearable device

Introduction

Background

The number of individuals having chronic stress and stress-related mental disorders, such as depression, is increasing globally [1]. While in most cases, the stress response of the human body protects the body in harmful environments, when stress occurs for a prolonged period, it can have several negative health effects [2]. As a matter of fact, chronic stress is known to increase the risk of developing a range of mental and physical disorders [3]. Examples of mental disorders are burnout, depression, and anxiety disorders, while related physical disorders include gastrointestinal disorders, obesity, diabetes, and heart diseases [3]. Moreover, chronic stress is often referred to as an economical problem. In 2018, a review was published on the costs of illness for work-related stress, suggesting that total costs per country could range from US $221 million to US $187 billion each year [4]. For these reasons, it is of critical importance to successfully monitor and manage chronic stress in the entire population, without losing individual-based focus as in current one-on-one psychotherapy.

As described by Kaplan [5], there are many challenges that researchers as well as doctors and therapists face to assess and monitor chronic stress levels in daily life. Although these challenges are already known for a long time, they have not been overcome. Stress questionnaires are the most common and convenient tools used for assessing chronic stress. Questionnaires, such as the Perceived Stress Scale (PSS) that is designed to assess the long-term stress effect (ie, stress over the last month), have been proven to have adequate reliability across different cohorts [6-8]. Nevertheless, they remain subjective measurements that are subject to recall errors and unable to capture the impact of stress on normal physiological functioning [9]. Over the past years, many algorithms and applications have been developed for acute stress detection based on physiological signals [10]. Predictors used are, for example, heart rate (HR), HR variability, skin conductance, and skin temperature [10-12]. Some algorithms can detect acute stress responses with high accuracy (>90%) in controlled conditions [11,12]. Detection algorithms based on real-time monitoring of physiological parameters could improve the objectivity of stress detection, facilitate capturing early signs of chronic stress, and support stress management in clinical practice, functioning both as awareness and feedback tools for users and therapists.

Most of the efforts in this field have focused, so far, on acute stress detection algorithms, showing the difficulty of generalization and the importance of individual-based models [13,14]. Literature on complete physiological modeling of chronic stress in humans remains poor owing to the complexity and high cost of collecting longitudinal real-world data. This complexity needs to be addressed first by linking long-term physiology measurements to chronic stress. Van Uum et al [15] found elevated hair cortisol levels in patients with severe chronic pain and higher PSS scores. Using a repeated measurement design, Schulz et al [16] reported that the awakening salivary cortisol level should be considered a possible biological correlate of chronic stress, as it was found to be elevated in participants with higher chronic stress levels. Schulz et al also found a gender difference in their study, showing larger increases of morning salivary cortisol levels in women than in men. This gender difference in physiological responses to chronic and acute stress has also been investigated by Jones et al [17]. In this previous study, it was found that chronic stress, measured according to the PSS score, modulates both the cortisol stress response and the cardiovascular stress response differently for male and female individuals. Female individuals had a lower HR at rest for higher PSS scores, whereas among male individuals, this correlation was not found. However, male individuals did show a lower HR while in acute stress for higher PSS scores. This second correlation was not found among female individuals.

When studying effects on physiology over multiple days, the circadian rhythm is used to describe diurnal physiological fluctuations [18]. For example, Cho et al [19] developed a model for mood prediction in patients with mood disorders, using the properties of the cosine curve in HR as one of the predictors. Tsanas et al [20] also stressed the importance of the circadian rhythm in monitoring under free-living conditions. Morelli et al [21] described this circadian rhythm in HR as four harmonics with periods of 24, 12, 8, and 6 hours. Their accurate approximation of the resting HR using these four harmonics makes this modeling suitable for mapping long-term effects on HR. Another important aspect to consider while studying effects on physiology over time is the weekday-weekend difference. In the study by Schlotz et al [22], it was shown that the cortisol response, often related to stress, is different for weekdays versus weekends. Chronic work overload and worrying was found to be related to this rise in cortisol after awakening on workdays, but not on weekend days. This effect was independent of gender. Pantzar et al [23] also found dissimilarities in stress patterns between weekend days and workdays when studying HR in response to stress. This difference was modulated by gender, and there was a smaller effect size by age [23].

By combining the approaches and findings from the studies described above in our study, we analyzed the long-term HR response to chronic stress and its bias for gender and age. Different from previous studies, we used wearable technologies to capture circadian variation of HR over multiple days.

Objectives of the Study

The main objectives of this study were as follows:

- To investigate the effect of chronic stress on HR over time, while correcting for weekdays versus weekends. For this, chronic stress has been defined as long-term perceived stress.
- To test possible modulation effects by gender and/or age. All interaction effects of these predictors will be included.

The outcomes of this analysis could be integrated into novel computational models of chronic stress detection, according to vital signs collected daily, for example, using widely adopted fitness trackers.

Methods

Recruitment

This study is part of the Stress in Work Environment (SWEET) study conducted by imec, which has been described previously [13]. Participants were recruited via email in 11 technology-oriented companies and were all office workers. They were included if they were active employees at the time of the study. No other inclusion or exclusion criteria were applied. Participants did not receive any compensation for participating in the study apart from having a chance at winning a restaurant or travel voucher. Vital signs of the participants were continuously measured for 5 consecutive days, starting on Thursday morning and ending on Monday evening, using two wearables devices (a chest patch and a wristband). All participants provided informed consent before participating in the study.

This observational study was approved by the Research Ethical Committee of UZ Leuven (S57916).

Data Collection

Before the start of the experiment, participants completed an intake questionnaire. The first part inquired about personal information, such as age, gender, health problems, work situation, and lifestyle. Thereafter, four psychological questionnaires were used to assess stress, depression, anxiety, sleep, and general health levels. For this study only perceived chronic stress, measured with the PSS, was considered. The questionnaires were distributed via a dedicated and protected web-based platform. On Thursday morning or afternoon, the participants received and started wearing the two wearable devices, including a chest patch [24] that obtained an electrocardiogram (ECG) at 256 Hz and a Chillband (wristband) that measured skin conductance at 256 Hz and skin temperature at 1 Hz. Both devices also measured three-dimensional accelerometer (ACC) signals at 32 Hz to control for movement artifacts. For this study, only the ECG and ACC data of the wearable positioned on the chest was used and only participants with complete ECG data from Friday 12 AM to Monday 4 PM were included for the analysis. This chest patch has been regulatory approved. Participants were advised to wear the chest patch the entire day and night and were asked to remove the chest patch during vigorous physical activities and to shower with a protective cover, since the chest patch is not waterproof. The battery life of the sensors exceeded the duration of the experiment. Data were recorded and stored on the devices' internal secure digital cards and uploaded to an internal secure cluster at the end of the experiment. For more details on the entire data collection in the SWEET study, refer to the report by Smets et al [13].

PSS as a Reference for Chronic Stress

The PSS measures "the degree to which situations in one's life are appraised as stressful" [25] and represents a measure of the global level of perceived stress over the past month. The PSS was designed for use in community samples of individuals with at least a junior high school education. Three versions of the PSS exist (the subscales PSS-14 items, PSS-10 items, and PSS-4 items). Psychometric properties, namely internal consistency reliability, test-retest reliability, and construct validity, of the PSS have been reviewed across studies by Lee et al in 2012 [26], where it was suggested that the subscale PSS-10 should be used to measure perceived stress both in practice and research. Therefore, in this study, the PSS-10 was used.

Data Preprocessing

Feature Calculation

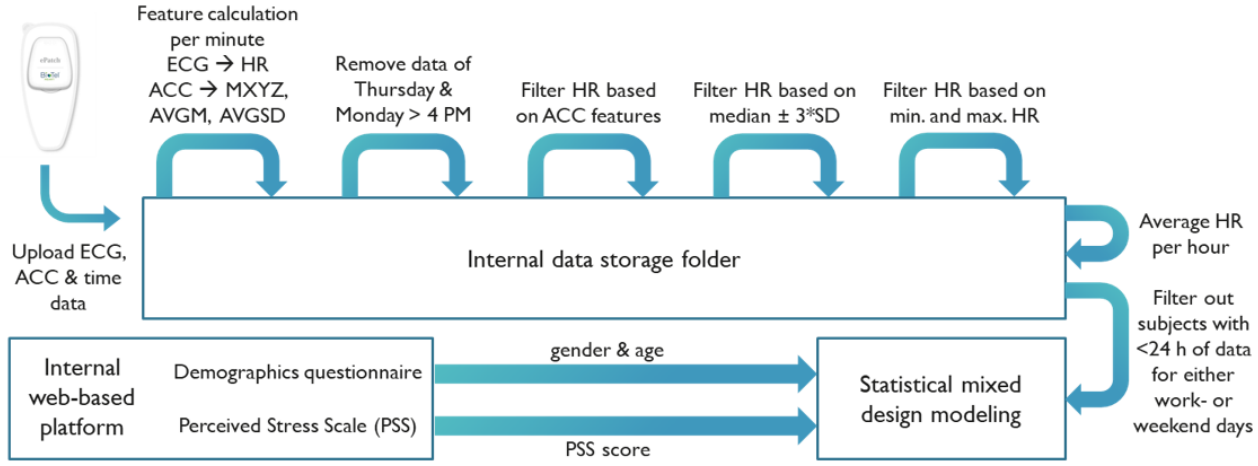
Based on the ECG data, the HR was derived. Refer to the study by Smets et al [13] for the methods on the quality indicator and peak detection algorithms. The ACC data were used to calculate the magnitude in the x, y, and z directions; the average mean of the x, y, and z coordinates; and the average SD of the x, y, and z coordinates. The absolute differences over time of these three features were summed to retrieve an indicator that was used to identify technical malfunctions. The formulas used to calculate each of the features can be found in Multimedia Appendix 1.

Filtering Methods

After feature calculation, a data set was obtained with values for the discussed features per minute. First, because all participants started and ended the assessments at different times on Thursday and Monday, the data of Thursday and the data after Monday 4 PM were removed. The data of every participant finally ranged from Friday 12 AM to Monday 4 PM. Second, technical malfunctions (ie, sensor saturation and poor patch adherence) were empirically detected based on the ACC data. For every minute, the indicator retrieved from three ACC features over the 10 closest minutes was calculated. If this value was less than an empirically selected threshold of 0.0005, the patch was probably either removed from the chest during this period or not measuring correctly. These data points were not included in further analysis. To filter out the values affected by artifacts from the remaining HR, two filters were used. First, the individual's median value ±3 times the SD were used as limits. Second, the minimum HR was set to 30 beats per minute, according to previous studies using HR filters [27,28]. The maximum HR was calculated based on the formula reported by Tanaka et al [29] shown in Multimedia Appendix 1. To assess low-frequency changes in HR, all data were summarized per hour. For the weekday against weekend comparison, Friday and Monday were labeled as no, and Saturday and Sunday were labeled as yes for the variable called weekend. Subjects who

did not have 24 or more hours of data for both a weekday and weekend day were excluded from the analysis. Figure 1 shows the data extraction, feature calculation, and filtering methods.

Figure 1. Diagram showing all steps from data upload to model development. ACC: accelerometer; AVGM: average mean acceleration of the x, y, and z coordinates; AVGSD: average standard deviation in acceleration of the x, y, and z coordinates; ECG: electrocardiogram; HR: heart rate; MXYZ: magnitude of acceleration in the x, y, and z directions. See Multimedia Appendix 1 for calculation of the features.



Statistical Analysis

Model Development

The retrieved average HR per hour functioned as the outcome of a mixed design model implemented in RStudio using the "lmer" method from the "lme4" Rpackage. As predictors, the hour of the day, harmonics over time, weekend (yes/no), gender, age, and PSS score were added in the respective order. The subject number was used as a random between-subject intercept and weekend was used as a random within-subject intercept. The time of the day was used in multiple ways based on previous studies on HR over time. First, according to the study by Morelli et al [21], the correlation between HR and time of the day is best described by a four-harmonic fit with periods of 24, 12, 8, and 6 hours. Second, Field et al [30] stated that because of the autoregressive property of physiological signals over time, the "corAR1" function is used as a correlational matrix, which includes the correlations within the hour of the day with random intercepts for each participant and workdays versus weekend days. Therefore, random slopes were used both for the hour of the day as a linear continuous variable and for the four harmonics in time that interacted with the weekend because of the differences in the pattern of HR between workdays and weekend days, as reported previously [31]. The method of comparison was set to "ML" (maximum likelihood), which, according to Field et al [30], is the best way to compare multilevel models. In "control," the maximum number of iterations was set to 200 and "returnObject" was set to true to make sure the models are not removed because of optimization issues. Not applicable (NA) values were excluded.

Gender, age, and the PSS score were added as fixed effects separately. All two-way interactions and three-way interactions were included. Interaction effects were included either if they were significant or, for the two-way interaction effects, if the three-way interaction effects with both the main effects were significant. Three-way interactions were only included when all related two-way interactions were included as well. Since the study focused on long-term effects, for the interactions with between-subject variables and time, only interactions with the circadian harmonic (the harmonic over 24 hours) were included. For the interactions between weekend and time, all harmonics were included because the random slopes differed for workdays and weekend days.

Model Comparison

Initially, the median and interquartile range (IQR) were calculated for the age, PSS score and HR. The HR, age, and PSS score were compared between male and female participants using a Mann-Whitney-Wilcoxon test, while HR on workdays and weekend days was compared using a Wilcoxon signed-rank test. In the final model, we included only interactions that significantly improved the model, which were related to a drop in the Akaike information criterion (AIC) value and an increase in the log likelihood (LogLik) value. The final model was compared to a similar model without the main PSS effect and interaction effects including the PSS score. The comparison was performed using the "Anova" method from the "lmerTest" Rpackage. To analyze the fixed effects and interaction effects, the chi-square test statistic (χ^2_{df}) and *P* value were calculated.

Results

Population Characteristics

Table 1 shows the population characteristics for all participants included in the analysis. A total of 328 participants fulfilled the criteria of having complete ECG data from Friday 12 AM to Monday 4 PM. There were slightly more male participants (n=186, 56.7%) than female participants (n=142, 43.3%) in the analysis. The Mann-Whitney-Wilcoxon test showed that between male and female participants, there were significant differences in the PSS score ($W=1.61e^4$, $P<.001$) and the median HR ($W=1.67e^4$, $P<.001$), but not in age ($W=1.46e^4$, $P=.10$). The Wilcoxon signed-rank test showed that there was a

significant difference in median HR over 24 hours for workdays compared with weekend days (V=2.23e^4, *P*=.007).

Table 1. Descriptive statistics for the population characteristics (N=328).

Characteristics	Median (IQR)	Comparison test result[a]
Age (years)		W=1.46e^4, *P*=.10
Male (n=186)	37.0 (±16.8)	
Female (n=142)	39.0 (±15.8)	
Total (n=328)	38.0 (±17.0)	
PSS[b] score		W=1.61e^4, *P*<.001
Male (n=186)	13.0 (±9.0)	
Female (n=142)	15.0 (±9.0)	
Total (n=328)	14.0 (±9.0)	
HR[c] (bpm[d])		
Male (n=186)	72.5 (±12.0)	W=1.67e^4, *P*<.001
Female (n=142)	75.8 (±10.0)	
Workdays (n=328)	73.6 (±12.3)	V=2.23e^4, *P*=.007
Weekend days (n=328)	73.7 (±10.9)	
Total (n=328)	73.5 (±10.8)	

[a]For comparing male and female participants, the Mann-Whitney-Wilcoxon test was used. For comparing workdays and weekend days, the Wilcoxon signed-rank test was used.

[b]PSS: Perceived Stress Scale.

[c]HR: heart rate.

[d]bpm: beats per minute.

Final Model

In the final model, 15,699 out of 15,744 observations (99.71%) were included after exclusion of NA values. The predictors included in the final model (AIC=100379.1, LogLik=−50043.6) were (1) the hour of the day, (2) the circadian harmonic, (3) the 12-hour harmonic, (4) the 8-hour harmonic, and (5) the 6-hour harmonic as time components; (6) whether it was the weekend as a within-subject variable; and (7) gender, (8) the PSS score, and (9) age as between-subject variables.

The two-way interaction effects included the following: whether it was the weekend × every harmonic over time; gender × 24-hour harmonic; age × 24-hour harmonic; PSS score × 24-hour harmonic; gender × PSS score; and age × whether it was the weekend.

The following two three-way interaction effects were included in the final model: gender × PSS score × 24-hour harmonic; and age × whether it was the weekend × 24-hour harmonic. The final model, including these predictors, had the lowest AIC value compared with all other models possible with the same main effects. In Multimedia Appendix 2, the built of the final model is shown with AIC and LogLik values for every layer. The final model significantly outperformed the same model without the PSS score and its interactions as predictors (AIC=100381.0, LogLik=−50050.51), according to an Anova model comparison (χ^2_6=13.88, *P*=.03).

Random Intercepts and Slopes

The random intercepts per participant (χ^2_1=6948.7, *P*<.001) and for workdays and weekend days (χ^2_1=233.9, *P*<.001) both significantly improved the model. Adding random slopes for the time of the day (χ^2_4=391.6, *P*<.001); the circadian effect of time (χ^2_{14}=1294.3, *P*<.001); and the harmonic over 12 hours (χ^2_{22}=668.7, *P*<.001), over 8 hours (χ^2_{30}=148.3, *P*<.001), and over 6 hours (χ^2_{38}=361.9, *P*<.001) significantly improved the model of HR. Adding an autoregressive correlation of HR over time also significantly improved the model (χ^2_1=12097.4, *P*<.001).

Main Effects of Time, Weekend, Gender, Age, and PSS Score

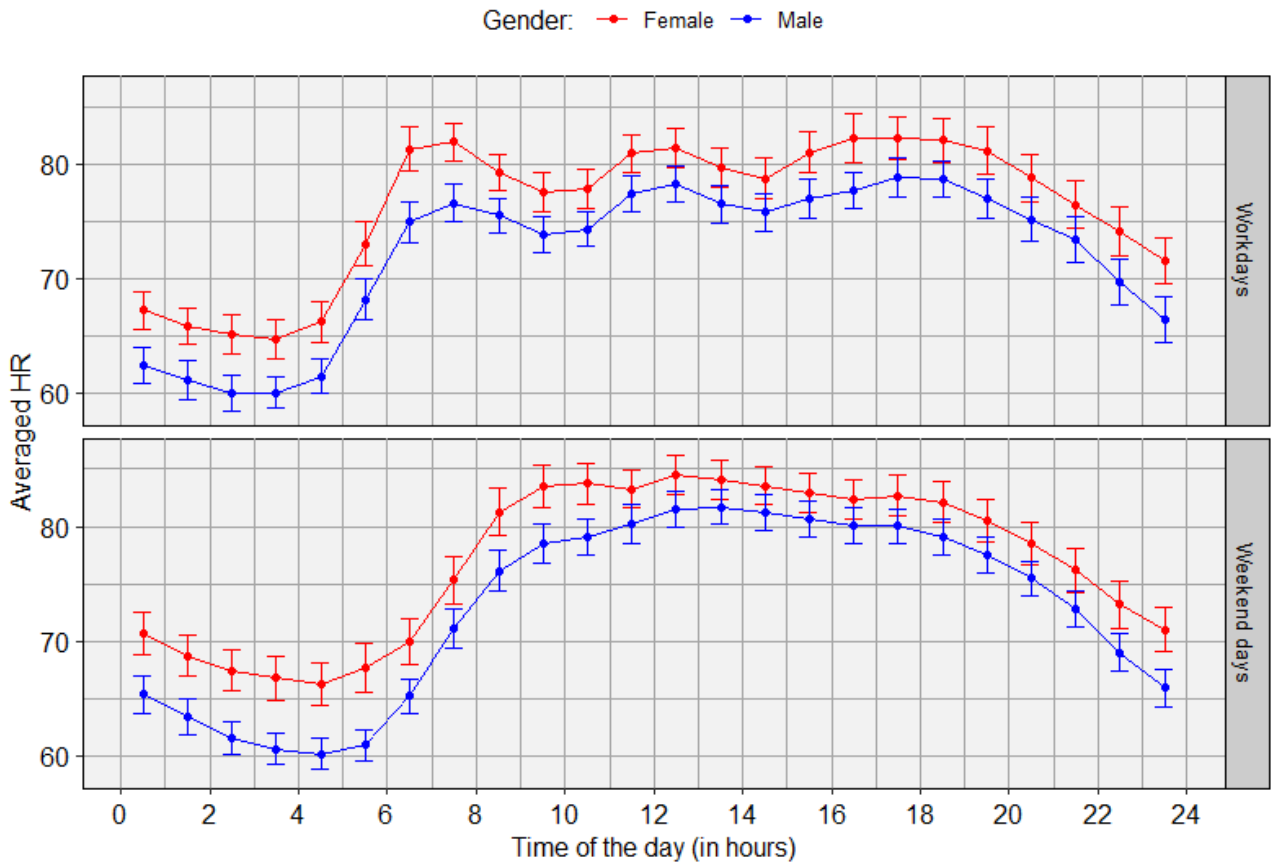
There were significant correlations with the HR for the main effects of the hour of the day (χ^2_1=73.22, *P*<.001); the circadian harmonic (χ^2_2=284.4, *P*<.001); the harmonic over 12 hours (χ^2_2=242.1, *P*<.001), over 8 hours (χ^2_2=23.78, *P*<.001), and over 6 hours (χ^2_2=82.96, *P*<.001); and gender (χ^2_1=24.02, *P*<.001). Table 1 shows a significant difference in HR for gender. However, whether it was a workday or weekend day was not significantly associated with HR (χ^2_1=1.13, *P*=.29). The age (χ^2_1=0.019, *P*=.89) and PSS score (χ^2_1=0.42, *P*=.52)

of the participant also did not have a significant correlation with HR.

Two-Way Interaction Effects With Circadian Harmonic Over Time

There were significant correlations with HR for the interaction effects of weekend and the circadian harmonic (χ^2_2=24.98, P<.001), weekend and the harmonic over 12 hours (χ^2_2=112.5, P<.001), weekend and the harmonic over 8 hours (χ^2_2=94.07, P<.001), and weekend and the harmonic over 6 hours (χ^2_2=131.4, P<.001). In Figure 2, HR is shown over the time of the day separately for gender and workdays and weekend days.

Figure 2. The heart rate (HR) over the time of the day, split for gender and workdays and weekend days. Since the data are summarized per hour, data points are provided in the middle of the hours. The error bars represent the 95% confidence intervals.



Furthermore, a significant correlation with HR was found for the interaction effect between gender and the circadian harmonic (χ^2_2=17.01, P<.001) and between age and the circadian harmonic (χ^2_2=16.69, P<.001). Although the interaction effect between the PSS score and the circadian harmonic was not significantly correlated with HR (χ^2_2=2.62, P=.27), it was included in the final model because of a significant higher order interaction effect.

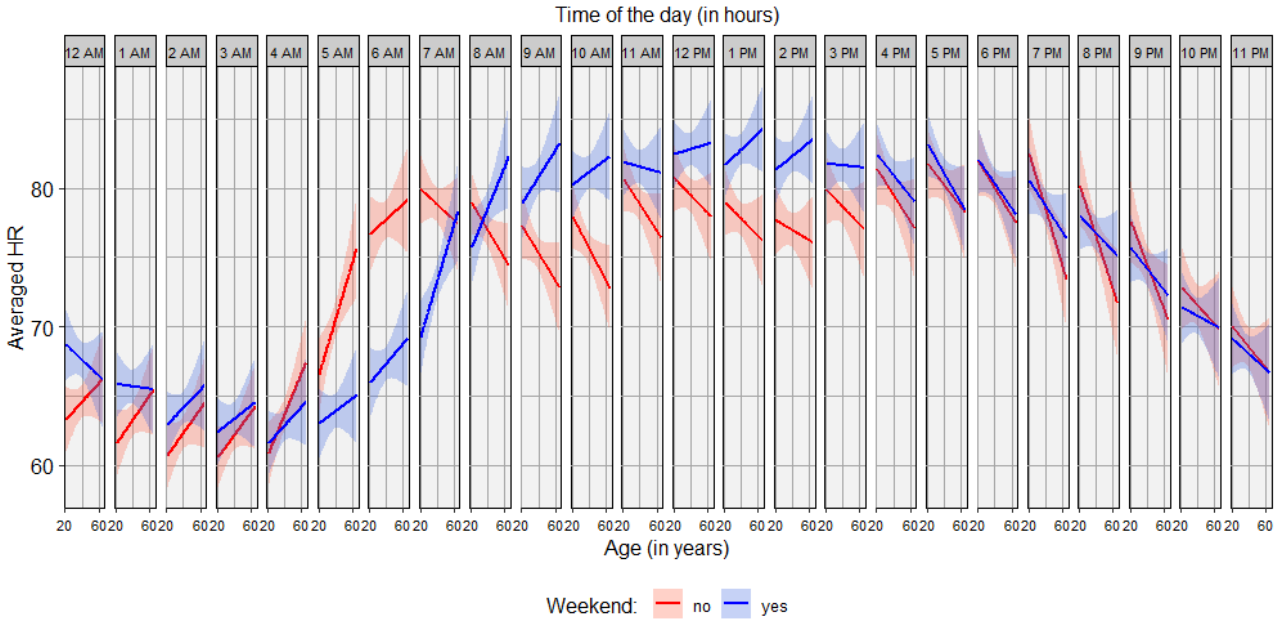
Two-Way Interaction Effect Between Age and Weekend, and Between Gender and the PSS Score

The correlation between HR and the interaction effect of gender and the PSS score of the participant was slightly nonsignificant (χ^2_1=3.63, P=.06). It was still included in the model, because of a significant higher order interaction effect. Although there was no significant correlation with HR for the interaction effect between age and whether it was the weekend (χ^2_1=0.34, P=.56), the interaction effect was included in the final model, because of a higher order interaction effect.

Three-Way Interaction Effect Among Age, Weekend, and Circadian Harmonic

There was a significant correlation between HR and the three-way interaction effect among age, whether it was a workday or weekend day, and the circadian harmonic (χ^2_2=7.13, P=.03). Figure 3 shows the effect of age on HR for every hour of the day, split for workdays and weekend days. On workdays, the correlation between age and HR switched from positive to negative at around 7 AM, whereas on weekend days, the same correlation switched from positive to negative at around 3 PM.

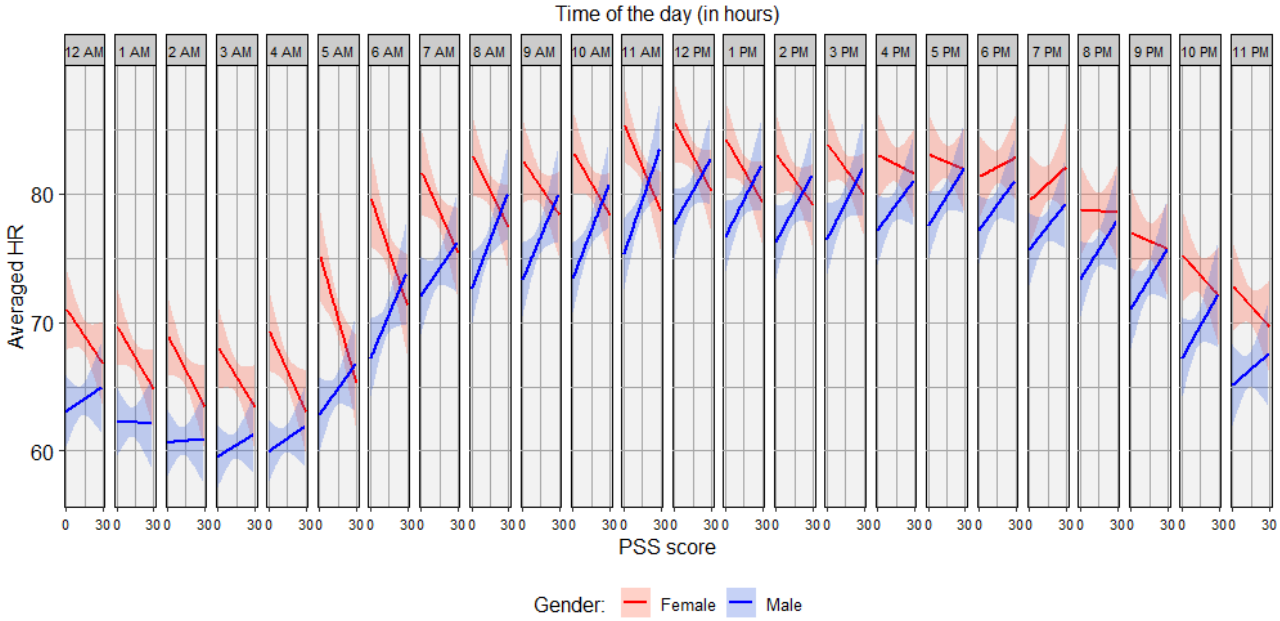
Figure 3. The heart rate (HR) over age per hour of the day, split for workdays and weekend days. Since the data are summarized per hour, the labels provided are the start of the hour.



Three-Way Interaction Effect Among Gender, PSS Score, and Circadian Harmonic

The three-way interaction among gender, the PSS score, and the circadian harmonic was found to be significantly associated with HR (χ^2_2=7.59, *P*=.02). Figure 4 shows the effect of the PSS score on HR for every hour of the day, split for female and male participants. The positive correlation between the PSS score and HR for male participants was the strongest at around 11 AM and flattened in the night. The negative correlation between the PSS score and HR for female participants was the strongest in the night/early morning and flattened or even switched to a slightly positive correlation at around 6 PM to 8 PM.

Figure 4. The heart rate (HR) over the Perceived Stress Scale (PSS) score for every hour of the day, split for female and male participants. The time above each subsection provides the starting time of the hour for which the HR is shown.



Discussion

Principal Findings

In this observational study, the circadian rhythm was defined as a combination of the 24-hour harmonic, 12-hour harmonic, 8-hour harmonic, and 6-hour harmonic, which were all found to be predictors of HR over time. The median HR was found to be different for male and female participants and for weekdays versus weekends. A relationship between HR and the three-way interaction of age, the circadian harmonic, and whether it was the weekend was found. Moreover, HR was

found to be related to the three-way interaction of the PSS score, gender, and the circadian harmonic.

The results confirm the validity of a four-harmonic circadian rhythm in HR as described by Morelli et al [21]. This is shown by the random slopes in harmonics over 24, 12, 8, and 6 hours and by the main effects of the four harmonics. However, four-harmonic circadian HR fluctuations were different between workdays and weekend days, as shown in Figure 2. Moreover, the random intercept for weekends and the difference in HR between workdays and weekend days, as shown in Table 1, suggested a different average HR for workdays and weekend days. The fact that there was no main effect of weekend is likely related to the inclusion of weekend as a random intercept. These findings match the findings of Cavallari et al [31] and suggest that specific HR models for workdays and weekend days are needed. Jones et al [17] related circadian HR fluctuations to circadian cortisol fluctuations. An increase in cortisol gives rise to appetite, resulting in food intake and therefore an increase in energy and HR. After eating, the parasympathetic nervous system is activated, which stimulates the gastrointestinal tract for food processing [32]. This activation is accompanied by a decrease in HR [33]. Figure 2 shows that for workdays, the eating times were very similar for everyone, with breakfast between 6 AM and 8 AM, lunch between 11 AM and 2 PM, and dinner between 5 PM and 8 PM. On weekends, the eating times are likely to be less congruent among participants, resulting in a smoothened average over the day. The study population consisted of office workers. All working people follow a very similar work rhythm (go to work, sit at work, have lunch, sit at work, and go home at roughly similar times), whereas on weekends, there is much more variability among people. It needs to be verified how this difference between workdays and weekend days scales to a different kind of working population, for example, shift workers having different daily patterns. Another clear difference in HR between workdays and weekend days was the timing of the HR morning awakening response. For workdays, HR rises earlier than for weekend days, which could be related to the earlier awakening times on workdays [34].

The average HR as well as the four-harmonic circadian HR fluctuations were different for not only workdays and weekend days but also participants, as indicated by the random intercept and slopes per participant. This suggests a need for personalized circadian HR models, which has been discussed previously by Fijorek et al [35]. This personalization is also suggested based on the higher average HR for female participants, as shown in Table 1, and the difference between male and female participants in circadian fluctuations in HR, as shown in Figure 2. According to Sandstede et al [36], the higher average HR for female participants is related to their average lower cardiac mass. Figure 2 shows that at daytime, the HRs of male and female participants are closer to each other than at night, which was also observed by Fijorek et al [35] and Bonnemeier et al [37]. According to Gregoire et al [38], the lower day-night difference for female individuals is probably related to the lower activity of the sympathetic nervous system in women, especially during the daytime.

Since the maximum HR decreases linearly with age [29], a correlation between age and HR might have been expected as well. However, this effect was not found to be relevant. This could be related to participant inclusion as a random intercept, explaining a part of the variance caused by differences in age. Nevertheless, age interacts with the circadian rhythm in HR differently for workdays and weekend days, as shown in Figure 3. There was a decrease in amplitude of HR fluctuations with age for workdays but not for weekend days. Hood & Shimon [39] have published a review on the aging clock and its circadian rhythms, in which they describe a lower amplitude of both cortisol and waking activity fluctuations, resulting in similar lower amplitudes of circadian HR fluctuations, described as flattening of the sinusoidal curve, as age increases. This explains the positive correlation between age and HR until 7 AM for workdays, which turns negative and then turns positive again around 12 AM, as can be seen in Figure 3. A higher nocturnal nadir of cortisol levels for higher ages was found by Van Cauter et al [40] and Sharma et al [41]. The translation of cortisol levels to long-term HR levels was not made in these studies but is used more often in stress studies [17,42]. On weekends, the positive correlation between age and HR turned negative at around 3 PM and then positive again at around 2 AM. The shift of the switch from negative to positive from 12 AM on workdays to 2 AM on weekend days could be related to later sleeping times during the weekend [43]. In general, it seems that the average HR during work nights (12 AM to 4 AM) is slightly lower than that during weekend nights, because people go to sleep later during the weekend. However, this does not explain why the switch from a positive correlation of age and HR to a negative correlation shifted from 7 AM on workdays to 3 PM on weekend days. Between 7 AM and 3 PM, there might be the most variance in activity among participants, since these are free hours during weekend days and work hours during workdays. Activity trackers could provide more understanding on the different directions of the correlation between age and HR for workdays and weekend days.

Lastly, no correlations between HR and the main effect of the PSS or HR and the two-way interaction effects of the PSS and other predictors were found. However, a correlation was found between HR and the three-way interaction of gender, the PSS score, and the circadian harmonic. This suggests that the effect of chronic stress can only be captured by including the dynamics of HR over time. Figure 4 shows that the correlation of the PSS and HR was exactly opposite for male and female participants throughout the day, except for the periods around 1 AM to 3 AM and 6 PM to 8 PM. In the first period, a flattening of the positive correlation between the PSS score and HR was noted for male participants, while in the second period, the negative correlation between the PSS score and HR for female participants switched to a slightly positive correlation. Table 1 shows that beside the difference in HR between male and female participants, there was a difference in the PSS score, which was found by Remor [6]. However, this does not explain the more complex three-way interaction effect including circadian fluctuations. Jones et al [17] described similar results, as seen in Figure 4, only for acute stress responses and explained it as a gender difference in evolution-based energy utilization strategies. Male individuals have a well-studied stress response,

often called the fight-or-flight response [44,45]. However, for female individuals, Taylor et al [46] introduced a different response, called the tend-or-befriend response. For long-term stress, the fight-or-flight response means stacking up resources to increase energy that might be needed for an actual fight-or-flight situation. In terms of biology, this means an increase in cortisol, accompanied by an increase in appetite, which results in an increase in food intake and energy, resulting in an increased HR [17]. This male-specific chronic stress response could be related to the correlation of job strain and obesity in male individuals but not female individuals, which was found by Brunner et al [47]. The tend-or-befriend response in female individuals would stimulate nurturing and caring behaviors, making sure that, for example, their children have enough resources [46]. This is related to a decrease in food intake, causing the depletion of fat stores and a decrease in HR among female individuals [17]. In the literature, this female evolutionary behavior has also been described as the biological reason why female individuals have, in general, higher fat storage than male individuals [48]. The theory explains the positive correlation between the PSS and HR for male participants and the negative correlation between the PSS and HR for female participants. In Figure 4, the period from 6 PM to 8 PM, when the correlation between the PSS and HR switched from negative to positive for female participants, could be explained by dinner time, which includes food intake among female individuals as well. The period from 1 AM to 3 AM, when the positive correlation between the PSS score and HR for male participants flattened, can be explained by the absence of food intake at night and no increased HR associated with chronic stress during that period. This theory should be studied in more detail in relation to circadian HR, comparing the circadian physiology of male and female individuals for different stress levels and linking HR to cortisol, sleep, and mealtime as well.

Limitations

This study only involved two workdays (Monday and Friday) and two weekend days (Saturday and Sunday) per participant. For the workdays, Monday stopped at 4 PM. The effect of this limitation can easily be seen in Figure 4, where the CIs in the plots seem to increase from 4 PM. It would be interesting to include more workdays and weekend days of multiple weeks, so an average over multiple days could provide a more consistent long-term measurement of HR fluctuations.

Future Work

The autonomic stress response includes several physiological changes, of which we investigated the HR response. Our longitudinal analysis should be applied to other physiological signals, such as skin temperature, galvanic skin response, and blood pressure [5,13,14]. These physiological measurements are known to have circadian fluctuations as well. It is also recommended to include longitudinal cortisol measurements to better understand the interaction across individual characteristics, hormonal mechanisms, and physiology in chronic stress. Information on sport activities, food intake times, and sleep times could also be considered to provide more context regarding physiological changes.

Conclusions

This study confirmed previous findings on the circadian rhythm of HR, its difference between workdays and weekend days, and its interactions with gender and age. The main discovery is the relationship between HR and the three-way interaction of chronic stress, gender, and the circadian harmonic. Our findings suggest that chronic stress prediction models and objective chronic stress measurements based on continuous HR detection should include interaction effects with circadian harmonics and gender to explain more subject variability. The development of these prediction models would enable continuous monitoring of long-term stress levels that could support therapists and psychologists to better understand patient progress and well-being.

Acknowledgments

This work was supported by a PhD fellowship from the Research Foundation-Flanders (FWO) awarded to EL (1SB4719N). Furthermore, we would like to thank Emmanuel Rios Velazquez and Vojkan Mihajlovic for carefully reviewing our manuscript.

OnePlanet Research Center acknowledges the support of the Province of Gelderland.

Authors' Contributions

Conceptualization: AK, GS, and EL; data curation: GS; formal analysis: AK, GS, and EL; investigation: AK, GS, and EL; methodology: AK, GS, and EL; project administration: AK and GS; resources: CH; supervision: GS and CH; visualization: AK and GS; writing-original draft preparation: AK and GS; writing-review and editing: AK, GS, EL, SC, and CH; guarantor: CH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Formulas.
[DOCX File , 20 KB - jmir_v22i9e18253_app1.docx]

Multimedia Appendix 2
Models.
[DOCX File , 23 KB - jmir_v22i9e18253_app2.docx]

References

1. Mojtabai R, Olfson M, Han B. National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults. Pediatrics 2016 Dec;138(6) [FREE Full text] [doi: 10.1542/peds.2016-1878] [Medline: 27940701]
2. Yaribeygi H, Panahi Y, Sahraei H, Johnston T, Sahebkar A. The impact of stress on body function: A review. EXCLI J 2017;16:1057-1072 [FREE Full text] [doi: 10.17179/excli2017-480] [Medline: 28900385]
3. McGonagle K, Kessler R. Chronic stress, acute stress, and depressive symptoms. Am J Community Psychol 1990 Oct;18(5):681-706. [doi: 10.1007/BF00931237] [Medline: 2075897]
4. Hassard J, Teoh KR, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: A systematic review. J Occup Health Psychol 2018 Jan;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]
5. Kaplan H. Psychosocial stress: Perspectives on structure, theory, life-course, and methods. San Diego: Academic Press; 1996.
6. Remor E. Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). Span J Psychol 2006 May;9(1):86-93. [doi: 10.1017/s1138741600006004] [Medline: 16673626]
7. Lesage F, Berjot S, Deschamps F. Psychometric properties of the French versions of the Perceived Stress Scale. Int J Occup Med Environ Health 2012 Jun;25(2):178-184 [FREE Full text] [doi: 10.2478/S13382-012-0024-8] [Medline: 22528542]
8. Ramírez MT, Hernández RL. Factor structure of the Perceived Stress Scale (PSS) in a sample from Mexico. Span J Psychol 2007 May;10(1):199-206. [doi: 10.1017/s1138741600006466] [Medline: 17549893]
9. Taylor JM. Psychometric analysis of the Ten-Item Perceived Stress Scale. Psychol Assess 2015 Mar;27(1):90-101. [doi: 10.1037/a0038100] [Medline: 25346996]
10. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. J Biomed Inform 2019 Apr;92:103139 [FREE Full text] [doi: 10.1016/j.jbi.2019.103139] [Medline: 30825538]
11. Castaldo R, Montesinos L, Melillo P, Massaro S, Pecchia L. To What Extent Can We Shorten HRV Analysis in Wearable Sensing? A Case Study on Mental Stress Detection. In: Eskola H, Väisänen O, Viik J, Hyttinen J, editors. EMBEC & NBC 2017, IFMBE Proceedings. Singapore: Springer; 2018:643-646.
12. Boonnithi S, Phongsuphap S. Comparison of heart rate variability measures for mental stress detection. 2011 Presented at: Computing in Cardiology; September 18-21, 2011; Hangzhou, China.
13. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67 [FREE Full text] [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
14. Yu H, Klerman E, Picard R, Sano A. Personalized wellbeing prediction using behavioral, physiologicalweather data. 2019 Presented at: International Conference on Biomedical & Health Informatics (BHI); May 2019; Chicago, USA. [doi: 10.1109/bhi.2019.8834456]
15. Van Uum SH, Sauvé B, Fraser LA, Morley-Forster P, Paul TL, Koren G. Elevated content of cortisol in hair of patients with severe chronic pain: a novel biomarker for stress. Stress 2008 Nov;11(6):483-488. [doi: 10.1080/10253890801887388] [Medline: 18609301]
16. Schulz P, Kirschbaum C, Prüßner J, Hellhammer D. Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. Stress Med 1998 Apr;14(2):91-97 [FREE Full text] [doi: 10.1002/(sici)1099-1700(199804)14:2<91::aid-smi765>3.0.co;2-s]
17. Jones A, Pruessner JC, McMillan MR, Jones RW, Kowalik GT, Steeden JA, et al. Physiological adaptations to chronic stress in healthy humans - why might the sexes have evolved different energy utilisation strategies? J Physiol 2016 Aug 01;594(15):4297-4307 [FREE Full text] [doi: 10.1113/JP272021] [Medline: 27027401]
18. Cornelissen G. Cosinor-based rhythmometry. Theor Biol Med Model 2014 Apr 11;11:16 [FREE Full text] [doi: 10.1186/1742-4682-11-16] [Medline: 24725531]
19. Cho C, Lee T, Kim M, In HP, Kim L, Lee H. Mood Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study. J Med Internet Res 2019 Apr 17;21(4):e11029 [FREE Full text] [doi: 10.2196/11029]
20. Tsanas A, Woodward E, Ehlers A. Objective Characterization of Activity, Sleep, and Circadian Rhythm Patterns Using a Wrist-Worn Actigraphy Sensor: Insights Into Posttraumatic Stress Disorder. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e14306 [FREE Full text] [doi: 10.2196/14306] [Medline: 32310142]
21. Morelli D, Bartoloni L, Rossi A, Clifton DA. A computationally efficient algorithm to obtain an accurate and interpretable model of the effect of circadian rhythm on resting heart rate. Physiol Meas 2019 Sep 30;40(9):095001. [doi: 10.1088/1361-6579/ab3dea] [Medline: 31437825]
22. Schlotz W, Hellhammer J, Schulz P, Stone AA. Perceived work overload and chronic worrying predict weekend-weekday differences in the cortisol awakening response. Psychosom Med 2004;66(2):207-214. [doi: 10.1097/01.psy.0000116715.78238.56] [Medline: 15039505]

23. Pantzar M, Ruckenstein M, Mustonen V. Social rhythms of the heart. Health Sociol Rev 2017 Jan 02;26(1):22-37 [FREE Full text] [doi: 10.1080/14461242.2016.1184580] [Medline: 28163655]
24. ePatch extended holter monitoring. BioTelemetry Inc. URL: https://www.gobio.com/heart-monitoring/14dayepatch/ [accessed 2020-06-05]
25. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385-396 [FREE Full text] [doi: 10.2307/2136404]
26. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
27. Kostis JB, Moreyra AE, Amendo MT, Di Pietro J, Cosgrove N, Kuo PT. The effect of age on heart rate in subjects free of heart disease. Studies by ambulatory electrocardiography and maximal exercise stress test. Circulation 1982 Jan;65(1):141-145. [doi: 10.1161/01.cir.65.1.141] [Medline: 7198013]
28. Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE J Biomed Health Inform 2015 May;19(3):832-838. [doi: 10.1109/JBHI.2014.2338351] [Medline: 25069129]
29. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. Journal of the American College of Cardiology 2001 Jan;37(1):153-156 [FREE Full text] [doi: 10.1016/s0735-1097(00)01054-8]
30. Field A, Miles J, Field Z. Discovering Statistics Using R. London: Sage Publications; 2012.
31. Cavallari JM, Fang SC, Mittleman MA, Christiani DC. Circadian variation of heart rate variability among welders. Occup Environ Med 2010 Oct;67(10):717-719 [FREE Full text] [doi: 10.1136/oem.2010.055210] [Medline: 20798005]
32. D'Alessio DA. Activation of the Parasympathetic Nervous System Is Necessary for Normal Meal-Induced Insulin Secretion in Rhesus Macaques. Journal of Clinical Endocrinology & Metabolism 2001 Mar 01;86(3):1253-1259 [FREE Full text] [doi: 10.1210/jc.86.3.1253]
33. Glick G, Braunwald E, Lewis RM. Relative Roles of the Sympathetic and Parasympathetic Nervous Systems in the Reflex Control of Heart Rate. Circulation Research 1965 Apr;16(4):363-375 [FREE Full text] [doi: 10.1161/01.res.16.4.363]
34. Robbins R, Affouf M, Seixas A, Beaugris L, Avirappattu G, Jean-Louis G. Four-Year Trends in Sleep Duration and Quality: A Longitudinal Study Using Data from a Commercially Available Sleep Tracker. J Med Internet Res 2020 Feb 20;22(2):e14735 [FREE Full text] [doi: 10.2196/14735] [Medline: 32078573]
35. Fijorek K, Patel N, Klima ?, Stolarz-Skrzypek K, Kawecka-Jaszcz K, Polak S. Age and gender dependent heart rate circadian model development and performance verification on the proarrhythmic drug case study. Theor Biol Med Model 2013 Feb 09;10:7 [FREE Full text] [doi: 10.1186/1742-4682-10-7] [Medline: 23394137]
36. Sandstede J, Lipke C, Beer M, Hofmann S, Pabst T, Kenn W, et al. Age- and gender-specific differences in left and right ventricular cardiac function and mass determined by cine magnetic resonance imaging. Eur Radiol 2000;10(3):438-442. [doi: 10.1007/s003300050072] [Medline: 10756992]
37. Bonnemeier H, Richardt G, Potratz J, Wiegand UK, Brandes A, Kluge N, et al. Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. J Cardiovasc Electrophysiol 2003 Aug;14(8):791-799. [doi: 10.1046/j.1540-8167.2003.03078.x] [Medline: 12890036]
38. Gregoire J, Tuck S, Yamamoto Y, Hughson RL. Heart rate variability at rest and exercise: influence of age, gender, and physical training. Can J Appl Physiol 1996 Dec;21(6):455-470. [doi: 10.1139/h96-040] [Medline: 8959312]
39. Hood S, Amir S. The aging clock: circadian rhythms and later life. J Clin Invest 2017 Feb 01;127(2):437-446 [FREE Full text] [doi: 10.1172/JCI90328] [Medline: 28145903]
40. Van Cauter E, Leproult R, Kupfer DJ. Effects of gender and age on the levels and circadian rhythmicity of plasma cortisol. J Clin Endocrinol Metab 1996 Jul;81(7):2468-2473. [doi: 10.1210/jcem.81.7.8675562] [Medline: 8675562]
41. Sharma M, Palacios-Bois J, Schwartz G, Iskandar H, Thakur M, Quirion R, et al. Circadian rhythms of melatonin and cortisol in aging. Biol Psychiatry 1989 Feb 01;25(3):305-319. [doi: 10.1016/0006-3223(89)90178-9] [Medline: 2914154]
42. Kudielka BM, Schommer NC, Hellhammer DH, Kirschbaum C. Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. Psychoneuroendocrinology 2004 Sep;29(8):983-992. [doi: 10.1016/j.psyneuen.2003.08.009] [Medline: 15219648]
43. Gubin DG, Weinert D, Rybina SV, Danilova LA, Solovieva SV, Durov AM, et al. Activity, sleep and ambient light have a different impact on circadian blood pressure, heart rate and body temperature rhythms. Chronobiol Int 2017;34(5):632-649. [doi: 10.1080/07420528.2017.1288632] [Medline: 28276854]
44. Collu R, Gibb W, Ducharme JR. Effects of stress on the gonadal function. J Endocrinol Invest 2014 Apr 4;7(5):529-537 [FREE Full text] [doi: 10.1007/bf03348463]
45. Lee J, Harley VR. The male fight-flight response: A result of SRY regulation of catecholamines? Bioessays 2012 Mar 08;34(6):454-457 [FREE Full text] [doi: 10.1002/bies.201100159]
46. Taylor SE, Klein LC, Lewis BP, Gruenewald TL, Gurung RA, Updegraff JA. Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. Psychol Rev 2000 Jul;107(3):411-429. [doi: 10.1037/0033-295x.107.3.411] [Medline: 10941275]
47. Brunner EJ, Chandola T, Marmot MG. Prospective effect of job strain on general and central obesity in the Whitehall II Study. Am J Epidemiol 2007 Apr 01;165(7):828-837. [doi: 10.1093/aje/kwk058] [Medline: 17244635]

48. Vague J. The degree of masculine differentiation of obesities: a factor determining predisposition to diabetes, atherosclerosis, gout, and uric calculous disease. Am J Clin Nutr 1956;4(1):20-34. [doi: 10.1093/ajcn/4.1.20] [Medline: 13282851]

Abbreviations

ACC: accelerometer
AIC: Akaike information criterion
ECG: electrocardiogram
HR: heart rate
LogLik: log likelihood
NA: not applicable
PSS: Perceived Stress Scale
SWEET: Stress in Work Environment

Edited by G Eysenbach; submitted 14.02.20; peer-reviewed by M Crisan-Vida, BA Barbara arnoldussen, T Muto, Y Kim, K Alipasandi; comments to author 20.04.20; revised version received 12.06.20; accepted 07.07.20; published 09.09.20.

Please cite as:
van Kraaij AWJ, Schiavone G, Lutin E, Claes S, Van Hoof C
Relationship Between Chronic Stress and Heart Rate Over Time Modulated by Gender in a Cohort of Office Workers: Cross-Sectional Study Using Wearable Technologies
J Med Internet Res 2020;22(9):e18253
URL: http://www.jmir.org/2020/9/e18253/
doi:10.2196/18253
PMID:32902392

Original Paper

Social Reminiscence in Older Adults' Everyday Conversations: Automated Detection Using Natural Language Processing and Machine Learning

Andrea Ferrario[1], PhD; Burcu Demiray[2,3,4], PhD; Kristina Yordanova[5,6,7], Dr; Minxia Luo[2,3], MSSc; Mike Martin[2,3,4,8], Prof Dr

[1]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Department of Psychology, University of Zurich, Zurich, Switzerland

[3]University Research Priority Program, University of Zurich, Zurich, Switzerland

[4]Collegium Helveticum, Zurich, Switzerland

[5]Institute of Computer Science, University of Rostock, Rostock, Germany

[6]Institute of Visual & Analytic Computing, University of Rostock, Rostock, Germany

[7]Interdisciplinary Faculty Ageing of Individuals and Society, University of Rostock, Rostock, Germany

[8]School of Psychology, Faculty of Health and Behavioural Sciences, University of Queensland, Brisbane, Australia

Corresponding Author:
Andrea Ferrario, PhD
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 44 632 86 24
Email: aferrario@ethz.ch

Abstract

Background: Reminiscence is the act of thinking or talking about personal experiences that occurred in the past. It is a central task of old age that is essential for healthy aging, and it serves multiple functions, such as decision-making and introspection, transmitting life lessons, and bonding with others. The study of social reminiscence behavior in everyday life can be used to generate data and detect reminiscence from general conversations.

Objective: The aims of this original paper are to (1) preprocess coded transcripts of conversations in German of older adults with natural language processing (NLP), and (2) implement and evaluate learning strategies using different NLP features and machine learning algorithms to detect reminiscence in a corpus of transcripts.

Methods: The methods in this study comprise (1) collecting and coding of transcripts of older adults' conversations in German, (2) preprocessing transcripts to generate NLP features (bag-of-words models, part-of-speech tags, pretrained German word embeddings), and (3) training machine learning models to detect reminiscence using random forests, support vector machines, and adaptive and extreme gradient boosting algorithms. The data set comprises 2214 transcripts, including 109 transcripts with reminiscence. Due to class imbalance in the data, we introduced three learning strategies: (1) class-weighted learning, (2) a meta-classifier consisting of a voting ensemble, and (3) data augmentation with the Synthetic Minority Oversampling Technique (SMOTE) algorithm. For each learning strategy, we performed cross-validation on a random sample of the training data set of transcripts. We computed the area under the curve (AUC), the average precision (AP), precision, recall, as well as F1 score and specificity measures on the test data, for all combinations of NLP features, algorithms, and learning strategies.

Results: Class-weighted support vector machines on bag-of-words features outperformed all other classifiers (AUC=0.91, AP=0.56, precision=0.5, recall=0.45, F1=0.48, specificity=0.98), followed by support vector machines on SMOTE-augmented data and word embeddings features (AUC=0.89, AP=0.54, precision=0.35, recall=0.59, F1=0.44, specificity=0.94). For the meta-classifier strategy, adaptive and extreme gradient boosting algorithms trained on word embeddings and bag-of-words outperformed all other classifiers and NLP features; however, the performance of the meta-classifier learning strategy was lower compared to other strategies, with highly imbalanced precision-recall trade-offs.

Conclusions: This study provides evidence of the applicability of NLP and machine learning pipelines for the automated detection of reminiscence in older adults' everyday conversations in German. The methods and findings of this study could be relevant for designing unobtrusive computer systems for the real-time detection of social reminiscence in the everyday life of older adults and classifying their functions. With further improvements, these systems could be deployed in health interventions aimed at improving older adults' well-being by promoting self-reflection and suggesting coping strategies to be used in the case of dysfunctional reminiscence cases, which can undermine physical and mental health.

(J Med Internet Res 2020;22(9):e19133) doi:10.2196/19133

KEYWORDS

aging; dementia; reminiscence; real-life conversations; electronically activated recorder (EAR); natural language processing; machine learning; imbalanced learning

Introduction

Reminiscence and Healthy Aging

The world's population is rapidly aging. With its first world report on aging and health [1], the World Health Organization (WHO) promoted a global paradigm shift in aging research by moving from a disease-focused model to a dynamic, contextualized, person-focused model of "healthy aging" [1]. This model emphasizes the interplay of personal characteristics (eg, abilities), environments, and their interactions in producing functioning. Activities represent the interaction between person characteristics and environments; they are understudied in traditional aging research. The novel WHO model encourages aging researchers to step outside the lab and into the real world to examine activities in everyday life [2], aiming to empower individuals to observe, measure, and take earlier action for their own health [3,4]. In this study, we embrace the healthy aging model by examining one such real-life activity: reminiscence.

Reminiscence [5] is the "naturally occurring act of thinking about or telling others about personally meaningful past experiences" [6,7]. These experiences may refer to specific events (eg, the first kiss), repeated ones (eg, going to the gym every Friday), extended ones (eg, a Christmas trip), or even long periods of life (eg, living in a foreign country for some years) [8]. Recalling or sharing valuable life experiences with third parties can support decision-making, bonding with others, and self-understanding [9]. Reminiscing can be a volitional or nonvolitional process recollecting memories [7], an activity that may be private or involve others [7]; in the latter case, we refer to social reminiscence. Many disciplines are interested in the study of reminiscence, such as nursing, social work, education, theology, psychology, and gerontology [10], with a strong focus on reminiscence in the context of aging. Researchers who study aging emphasize a cognitive activity in old age such as reminiscence to be an essential part of healthy aging [11]; in fact, the use of memory interventions and reminiscence in therapies for older adults is common, emphasizing the relation between self-positive functions of reminiscence and well-being [12], according to Webster and Cappeliez's tripartite model of reminiscence [13].

Naturalistic Observation as a New Approach to the Study of Reminiscence

The study of reminiscence in older adults has traditionally focused on (1) reflective self-reporting and life reviews [11] and (2) automated reminiscence therapy, that is, "a nonpharmacological intervention involving the prompting of past memories, […] for therapeutic benefits, such as the facilitation of social interactions or the increase of self-esteem" [14], especially for dementia patients.

The use of self-reporting has potential limitations, such as recall biases, response styles, demand characteristics, social desirability, and limitations to introspection [15]. Moreover, the self-report method provides researchers only with the average frequency of an activity over a certain period of time [16]. Studies with a focus on automated reminiscence therapy, in contrast, typically aim at eliciting reminiscence from users (eg, with the remote assistance of a therapist), rather than collecting spontaneous reminiscence events during everyday life settings [17-19].

In 2017, Demiray et al [6] were the first to examine reminiscence using a naturalistic observation method to enable investigating reminiscence in the real world. As opposed to those used in the reminiscence therapies, this method does not rely on self-report and tracks objective behaviors, such as speech in everyday life, with no elicitation of reminiscence events. It can also involve older adults in scientific investigations who would otherwise be excluded from real-life studies relying on self-reporting (eg, older adults who are intimidated by technology or therapy are unable to use a smartphone to complete surveys or to self-report due to worsened eyesight, and/or are part of a clinical population). Furthermore, this method allows for microlongitudinal study designs with many measurement points per participant and both within- and between-persons perspectives in analyses.

However, to understand what kinds of reminiscence patterns are predictive of maintaining healthy aging and quality of life [2], each instance of reminiscing has to be reliably detected in everyday life contexts. In fact, reminiscing is a context-dependent, real-world cognitive activity [20]; if it can be reliably and accurately assessed, it is possible to study the effects of real-world cognitive activities on health outcomes, such as cognitive abilities or cognitive impairments [6,12]. A rich body of literature exists on the link between reminiscence therapies and cognitive and well-being benefits in aging populations: a recent review [21] based on 22 randomized controlled studies showed evidence for the positive effects of reminiscence therapy on quality of life, cognition, and communication in dementia patients. Subramaniam and Woods [22] specifically reviewed the effects of individual-based therapy

and concluded that it shows immediate benefits on well-being and cognition. The challenge, however, is to find automated ways to extract and disambiguate the cognitive activity information from data streams collected from many persons in real-world situations [23]. Tracking, detecting, and prompting functional as well as positive reminiscence behaviors in everyday life should allow researchers to design digital interventions to enhance the quality of life of healthy older adults and other patient populations. To these ends, natural language processing (NLP) and machine learning methodologies allow researchers to explore the possibility of reliably predicting reminiscence in combination with naturalistic observation methods and in real time.

Using NLP and Machine Learning for Reminiscence

Yordanova et al [24] were the first to investigate the applicability of NLP and machine learning methodologies on data from a naturalistic observation study by Demiray et al [6]; they introduced an NLP pipeline and machine learning routines to automatically code the social behaviors and interactions (eg, talking to a partner or daughter/son, giving advice, receive support, etc) in the transcripts of recorded conversations. As coding is a manual process that involves much effort and time, their results showed that the use of NLP and machine learning automation on transcripts of recorded conversations enabled reliable coding of social behaviors and interactions, reducing effort and time. However, they did not consider detecting reminiscence. Their proposed pipeline included data augmentation procedure to cope with highly imbalanced classes, feature engineering based on linguistic, contextual and statistical approaches, and supervised learning with classifiers such as decision trees, random forests (RF), and support vector machines (SVM). As the extreme imbalance between classes poses a central problem in automated coding of textual data, other works propose training classifiers with annotated data sets and later performing manual evaluation to correct misclassified instances by experts [25]. This second manual step is expensive and time-consuming, but it shows that NLP and machine learning deliver promising results with respect to the automated analysis of textual data in social science applications. In fact, the use of NLP and machine learning methods has shown great potential for analyzing social media posts for social affect and behavior, identifying trends in society and demographics, as well as generating predictions of society-changing events, such as diseases [26-28].

Motivated by the availability of reminiscence data from the naturalistic observation study [6] and the results of Yordanova et al [24], this study aims to develop pipelines with NLP and machine learning strategies to automatically detect reminiscence in older adults' everyday conversations in German using their written transcripts. To do so, we introduce various NLP features (bag-of-words models, part-of-speech [POS] tagging, and pretrained word embeddings) to preprocess written transcripts, which are fed into four families of machine learning algorithms (RF, adaptive boosting [ADA], extreme gradient boosting [XGB], and SVM); multiple learning strategies are setup to cope with class imbalance in data. The methods of this study support the understanding of a key activity of the healthy aging model, that is, reminiscence, in a real-world setting by leveraging transcriptions of everyday life conversations. Moreover, they could support the design of computer systems to detect social reminiscence in the everyday lives of older adults in real time and classify different reminiscence functions. These systems lie at the core of digital health intervention programs [29] aimed at improving older adults' well-being by promoting self-reflection, as suggested by the healthy aging model. They also provide users with coping strategies for dysfunctional reminiscence [13], which has a negative emotional valence and affects physical and mental health [30-32]. Even when the real-time recording of older adults' memory activities contained in daily conversations in a research study is shown to be feasible, there are multiple challenges to scaling up this intervention to a large population. However, current findings suggest that despite potential concerns about privacy and data protection issues, there are now a number of technical [33] and analytical solutions for privacy-preserving machine learning for such data [34]. Additionally, a majority of older adults is willing to share portable data collections with researchers [3].

Methods

Overview of the Study Design

This study comprises the following steps: (1) collecting data from a naturalistic observation study [6], (2) preprocessing data with NLP methodologies, and (3) training and validating machine learning models to detect reminiscence in a given corpus of transcriptions, by implementing three distinct learning strategies to cope with class imbalance in the data.

Data Collection: Older Adults' Everyday Conversations and Reminiscence

Data Source

The aim of the naturalistic observation study by Demiray et al [6] was to collect everyday conversations of older adults and examine social reminiscence behavior. Random snippets of older adults' everyday conversations were collected using the Electronically Activated Recorder (EAR) [35]. They generated 13,275 audio files from 48 older adults (22 men and 26 women) residing in Zurich, Switzerland, over a period of 4 days. The average age of the sample was 70.54 years (SD 4.65, range 62-83 years). To be eligible for the study, participants were required to have a minimum score of 27 on the Mini Mental State Examination [36]. The participants had an average of 10.5 years of education (SD 3.0, range 7-25 years), and they all spoke Swiss German. The study included an introductory session, an observation period, and a feedback session. In the introductory session of the study, participants signed informed consent and received an iPhone in which the EAR was installed. They were informed that the EAR would randomly record a few seconds of audio multiple times per day, except for an automatically inactivated period from every midnight to 6 AM the next day, during the whole observation period. They were told to carry the iPhone and continue daily living, and they were informed that they would not notice when the EAR was recording.

The observation period started the day after the introductory session and lasted 4 consecutive days, during which the audio file recording occurred. After the observation period, the

participants were invited to the feedback session, where they returned the phones, completed further questionnaires, and provided their feedback about their experiences of carrying the EAR. They received password-protected CDs containing all of their audio files. All study procedures were approved by the Ethics Research Institute of the Department of Philosophy at the University of Zurich.

The EAR App

The EAR (version 2.3.0) [35,37] was installed on each iPhone. It was set to randomly record 30-second audio files 72 times over 4 days. Thus, each participant was recorded 288 times and for a total 144 minutes each. Each iPhone was set to "airplane mode" and locked with a screen-lock password. The participants were instructed to charge the iPhone in the evenings. At the end of the study, the participants reported that the EAR did not affect their daily activities or way of speaking [20].

Data Generation: Transcribing and Coding Audio Files

For each audio file, all utterances by the study participants were transcribed by two research assistants, who were fluent in Swiss German and standard written German. Swiss German is an Alemannic dialect spoken in the German-speaking part of Switzerland, which does not have a standard written form. Thus, the Swiss-German dialect in the audio files was translated word-by-word into standard written German and then transcribed. Coders generated binary variables (with values 0 and 1), indicating whether the participant was talking or not and whether he or she was reminiscing or not. The function(s) of reminiscence [8] and participants' conversation partners (ie, partner/spouse, daughter/son, other family members, etc) were also coded.

Coding Reminiscence

Two coders performed the manual coding of reminiscence in each audio file independently. The interrater reliability for the coding of reminiscence, that is, the percentage of audio files with the same coding assigned by both coders, was 95%. All discrepancies in coding were solved by relistening and recoding through discussion. To classify reminiscence, the coders generated a binary variable with values 0 and 1, which correspond to a general conversation and a reminiscence case, respectively.

Preparing Data

Of the 13,275 audio files generated during the study by Demiray et al [6], 2214 contained conversations of older adults. The data set used for this study comprises these 2214 transcripts, of which 109 (4.9%) were coded as reminiscence. Before applying NLP to the transcripts, we preprocessed the data with regular expressions to remove coding artifacts like "xxx," "YYY," "xxxx," etc, denoting utterances from the audio recording that were not possible to transcribe as well as leading and trailing whitespaces.

Natural Language Processing of Transcripts

To use written transcripts as inputs of machine learning models, we preprocessed them by computing the NLP features (1) "bag-of-" models on both words and POS tags and (2) real-valued embeddings for each transcript in the data set using pretrained German word embeddings.

Bag-of-Words Models

Bag-of-words models [38,39] represent transcripts as real-valued vectors by tokenizing all transcripts in the provided data set, collecting unique tokens, and counting their occurrences before applying normalization (eg, term frequency–inverse document frequency [tf–idf] [38,39]). Bag-of-words models do not consider the order of words in transcripts and generate high-dimensional representation of textual data. In fact, bag-of-words models represent each transcript by a real-valued vector whose dimension is equal to the size of the vocabulary of the whole data set of transcripts. They are widely used for text classification tasks, including studies in digital health [29,40-42]. We computed bag-of-words models using the TfIdfVectorizer() function in the Python sklearn library [43].

Bag-of-POS Models

POS tagging is the process of assigning a POS tag to each word in a given corpus [38,39]; the algorithm that performs the tagging is called a POS tagger; a set of all tags is called a tagset. POS tagging enables including information from a word's context (ie, its relationships with close and related words in a document) in text classification tasks [29]. In this study, we used the POS tagger provided in the core model for the German language "de_core_news_sm," which is available in the Python library SpaCy [44], to generate the POS tags for all tokens retrieved from the corpus of 2214 transcripts. The SpaCy POS tagger has a tagset comprising 17 distinct tags; similar to bag-of-words models, a bag-of-POS model extracts all POS tags (instead of words) from a transcript and counts their occurrences before applying normalization, such as tf–idf. With bag-of-POS models, one can encode information on the linguistic structure of each transcript in a real-valued, low-dimension representation.

Word Embeddings

Word embeddings are real-valued representations of textual data encoding the "distributional hypothesis" [45] about language and words: words that occur in similar contexts tend to be closer to each other as real-valued vectors. Moreover, word embeddings are generally real-vector representations of textual data of much lower dimension than, for example, those in bag-of-words models. They have emerged as a common technique to compute representations of textual data, including studies in digital health [29,40]. In this study, given the limited number of available transcripts, we opted for pretrained German word embeddings using the SpaCy core model for the German language "de_core_news_sm." The model is "German multi-task CNN trained on the TIGER and WikiNER corpus" [46] and each word embedding has 300 dimensions. The TIGER corpus [47] is curated by the Institute for Natural Language Processing at the University of Stuttgart; it comprises 900,000 tokens from sentences of German text, taken from the Frankfurter Rundschau newspaper [48]. On the other hand, WikiNER [49] is a corpus for multilingual named entity recognition from Wikipedia. In this study, the embedding of each transcript is a real vector of

300 dimensions resulting from averaging the embeddings of all words generated from the given transcript after its tokenization.

Machine Learning on Transcripts: Learning Strategies

Class Imbalance

The data set of 2214 transcripts has 109 records coded as reminiscence; therefore, it has a class imbalance with a ratio of 20:1. A class imbalance [50-53] is a common phenomenon in machine learning and, in particular, in textual data [24,40,54,55]. Since most machine learning algorithms are biased toward the majority class [53], researchers have proposed various solutions to cope with learning with imbalanced data sets [50,51], such as class-based weighting of misclassification errors or data resampling techniques aimed at reducing imbalance [53]. Research has also promoted the use of performance measures, which consider the presence of class imbalance [53] to evaluate machine learning models.

Machine Learning Methods

In this study, to cope with class imbalance during the training of machine learning models to detect reminiscence, we implemented the following learning strategies: (1) class-weighted learning (CWL), (2) meta-classifier (MC) learning, and (3) learning in the presence of data augmentation with the Synthetic Minority Oversampling Technique (SMOTE) algorithm [56]. For all the learning strategies, we trained machine learning classifiers [57] using the RF, ADA, XGB [58], and SVM algorithms fed on all NLP features: bag-of-words, bag-of-POS NLP models, and pretrained word embeddings. In the cases of RF or boosting algorithms trained on either bag-of-words or bag-of-POS features, we computed feature importance using the sklearn property .feature_importances_. Table 1 summarizes all the NLP features, classifiers, and learning strategies used in this study.

Table 1. Summary of all natural language processing (NLP) features, machine learning algorithms (ie, classifiers), and learning strategies considered in this study.

Features, algorithms, and strategies	Cases
NLP features	• Bag-of-words • Bag-of-POS[a] • Pretrained German word embeddings
Algorithm (classifier)	• Random forests • Adaptive boosting • Extreme gradient boosting • Support vector machines
Learning strategy	• Class-weighted training • Meta-classifier training • Data augmentation with SMOTE[b]

[a]POS: part-of-speech.

[b]SMOTE: Synthetic Minority Oversampling Technique.

Evaluation Metrics

For all the learning strategies, the performance of the machine learning models on test data was computed using the area under curve (AUC), precision, recall (or sensitivity), the average precision (AP), specificity, and the F1 score measures. The AUC summarizes in a single number the performance of the classifier shown in the receiver operating curve [59,60], which plots the true positive rate versus the false positive rate, at various classifier thresholds settings. Precision is the number of true positives (ie, transcripts containing reminiscence, which are correctly predicted by the machine learning model) divided by the number of transcripts, which are predicted to contain reminiscence by the model. Recall (or sensitivity) is the number of true positives divided by the number of transcripts containing reminiscence. The sklearn implementation of the AP [61] summarizes the precision-recall curve [62] as the weighted mean of precisions achieved at each threshold by the classifier, the increase in recall from the previous threshold used as the weighting. We chose to report the AUC and AP, as they provide a global overview of the classifier performance, for all possible classification thresholds. Specificity is the number of true negatives divided by the total number of negative instances, while the F1 score is the harmonic mean of precision and recall measures [38]; it is a common evaluation metric in the presence of imbalanced data.

The formulas for AP and F1 score are as follows:

In the AP formula, R_n denotes the recall computed at step n (similarly for R_{n-1}), while P_n is the precision at step n.

Error Analysis

To identify errors in detecting reminiscence, we analyzed the false positives with the highest predicted probabilities (10% of all cases) and the false negatives with the lowest predicted probabilities (10% of all cases) computed for all the models presented in Table 2.

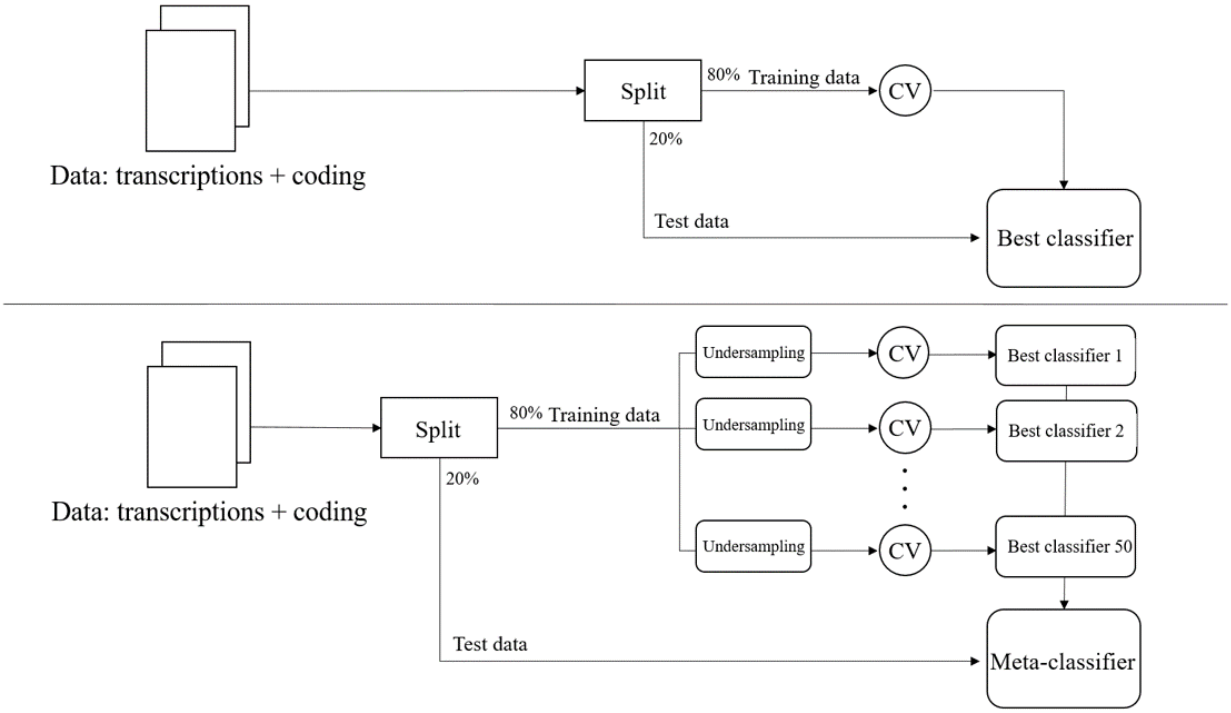
Experimental Setting

Class-Weighted Learning

We show the CWL strategy in Figure 1, top panel. After an initial partition of the data into train and test subsets with a 80:20 ratio, a 5-fold cross-validation routine was applied to the

1771 training data (87 reminiscence) to select the best model for all NLP features and families of classifiers shown in Table 1. We use the AUC to measure performance on the validation folds. The CWL strategy does not modify the imbalanced class distribution of training data, but reweighs them according to their class during the training of the machine learning algorithm at hand [40,52,54], penalizing the cost of misclassifying data points from the minority class [52,63,64]. For example, for the RF, ADA, and SVM algorithms, we selected the parameter "*class_weight=balanced*" [65] in their sklearn implementations. Similar considerations held for XGB algorithms, where we selected "*scale_pos_weight=weight*," where *weight* denoted the ratio of the number of negative class samples to the positive class (ie, reminiscence) [66]. The best model resulting from the cross-validation was evaluated on test data, by computing AUC, AP, and F1.

Figure 1. Top panel: Class-weighted learning and data augmentation learning strategies. We performed a single data partition into train and test sets; the best classifier emerged from 5-fold cross-validation (CV). Bottom panel: meta-classifier learning strategy; we undersampled the training data 50 times, collecting 50 distinct models in a voting ensemble after CV. We applied the three strategies for all the combinations of natural language processing features and machine learning algorithms shown in Table 1.



Meta-Classifier Training

We show the MC strategy in Figure 1 (bottom panel). The MC is a majority voting ensemble classifier [67] comprising 50 equally weighted distinct models resulting from 50 runs of 5-fold cross-validation on 50 randomly undersampled (with 1:1 ratio) training data sets. Undersampling is a common technique in imbalanced learning [41,68]; the use of voting allows reducing the bias of a single undersampling of training data. For each run, the cross-validation routine was performed on 174 training transcripts (87 reminiscence). The AUC and AP of the MC were computed on test data.

Data Augmentation With SMOTE

Figure 1 (top panel) shows the strategy involving data augmentation with the SMOTE algorithm: it follows the same steps as the one implemented for CWL. However, during the 5-fold cross-validation, the results of all NLP pipelines on transcripts are preprocessed with SMOTE before being fed into the machine learning classifiers. SMOTE [50] is an algorithm that generates synthetic examples of the minority class (in this study, the reminiscence class) in imbalanced data sets; given a minority class data point, SMOTE generates synthetic examples along line segments connecting the given data point to its K nearest neighbors (K=5 is the default value). SMOTE is widely used to perform data augmentation in the presence of imbalanced learning, including clinical studies [40,69,70]. The Python implementation of the SMOTE algorithm also allows controlling the oversampling quota, that is, the ratio between the minority class (after resampling) and the majority class. For example, using SMOTE on training data to reach a 1:1 ratio among classes, we ran the cross-validation routines in Figure 1 (top panel) on a total of 3368 data points (1684 reminiscence).

Cross-Validation

For all three learning strategies, as shown in Figure 1, we performed cross-validation by tuning the hyperparameters of (1) the NLP pipelines computing bag-of-words and bag-of-POS features, together with those of (2) the machine learning algorithms before retrieving the best classifier (MC in the case of the voting learning strategy).

NLP Preprocessing Pipelines

For the bag-of-words features, we tuned the n-grams, German stopword list (ie, its removal or not), minimum document frequency, maximum document frequency, and maximum

number of feature hyperparameters. After preprocessing, in case of 1-grams we have 6596 tokens, and 38,347 in the case of 2-grams. The preprocessing of the POS features follows the one for bag-of-words, with the exception of the German stopword list; in case of 1-grams we obtained 16 tokens, and 270 in the case of 2-grams.

Machine Learning Models: Hyperparameters

For ADA classifiers, we tuned the number of estimators and the learning rate; for XGB, we tuned the number of estimators, the learning rate, and the maximum tree depth; and for RF, we tuned the number of trees in the ensemble, their depth, and the number of features considered at each tree split during training. In addition, when considering the data augmentation learning strategy, we tuned the SMOTE number of nearest neighbors and oversampling quota hyperparameters. The NLP preprocessing pipelines are the same for all machine learning models.

Results

Machine Learning Modeling

Multimedia Appendix 1 displays the results of the machine learning modeling, reporting the best performing classifier for all NLP features and learning strategies, together with its AUC, AP, precision, recall, F1, and specificity performance measures, which we computed on test data (443 transcripts, 22 reminiscence).

Table 2 displays the two best performing classifiers (considering the F1 score as performance measure) for each learning strategy. We choose two best models per learning strategy to show how different combinations of machine learning algorithms and NLP features may result in different precision-recall trade-offs.

Considering the class weighting strategy, SVM outperforms all other classifiers, with the highest F1 score (0.48) when trained on bag-of-words features, with a balanced precision-recall trade-off and very high specificity (0.98). On the other hand, SVM trained on word embeddings shows higher AP (0.61) and recall (0.77); however, its lower precision (0.21) results in a lower F1 score (0.33), together with a lower specificity (0.85).

The best performing classifiers for the MC strategy, which are trained on word embeddings and bag-of-words features, show F1 scores (0.30 for both ADA and XGB classifiers) lower than those in the class-weighted and data augmentation strategies. This is due to low precision; similarly, they show low specificities. However, they reach very high recall (1.00 for ADA and 0.86 for XGB classifiers) and high AUCs.

Finally, for the data augmentation strategy, SVM on word embeddings shows highest F1 score (0.44), with recall (0.59) higher than precision (0.35) and high specificity (0.94). On the other hand, ADA classifiers trained on word embeddings show a more balanced precision-recall trade-off, with a higher precision (0.43) but lower recall (0.41) than the SVM classifier, resulting in a slightly lower F1 score (0.42) but higher specificity (0.97).

Table 2. Summary of best models for each learning strategy, considering the F1 score.

Learning strategy and NLP[a] feature	Classifier family	AUC[b]	AP[c]	Precision	Recall	F1	Specificity
Class weighting							
BOW[d]	SVM[e]	0.91	0.56	0.50	0.45	0.48	0.98
EMB[f]	SVM	0.91	0.61	0.21	0.77	0.33	0.85
Meta-classifier							
BOW	XGB[g]	0.90	0.45	0.18	0.86	0.30	0.79
EMB	ADA[h]	0.92	0.38	0.18	1.00	0.30	0.76
Data augmentation							
EMB	SVM	0.89	0.54	0.35	0.59	0.44	0.94
EMB	ADA	0.84	0.36	0.43	0.41	0.42	0.97

[a]NLP: natural language processing.

[b]AUC: area under the curve.

[c]AP: average precision.

[d]BOW: bag-of-words.

[e]SVM: support vector machines.

[f]EMB: word embeddings.

[g]XGB: extreme gradient boosting.

[h]ADA: adaptive boosting.

Feature Analysis

Overall, considering Table 2 and the F1 score, the use of word embeddings outperformed other NLP features for the data augmentation strategy; it delivers performance equal to the one with bag-of-words features for the MC strategy; on the other hand, for class weighting, bag-of-words features deliver the highest performance with the SVM classifier.

Considering bag-of-words features, for the data augmentation strategy XGB outperformed the other classifiers (with 1-grams, and no German stopword removal) with 300 boosting iterations, shallow trees (ie, a depth of 1), and the number of neighbors used by SMOTE equal to K=13. The words with the highest feature importance were German stopwords (eg, “gewesen,” “und,” “wir,” “ist,” and “ich” [“been,” “and,” “we,” “is,” and “I”]). Training the same classifier but removing the stopwords led to a strong decline in performance (AUC=0.79, AP=0.22, F1 = 0.27); in this case, the words with the highest feature importance comprised adverbs (eg, “aber,” and “einfach” [“but” and “simply”]) and past participles (eg, “gesagt” and “gehabt” [“said” and “had”]), among others.

Considering POS features and the class weighting strategy, XGB delivered the highest performance in the presence of 1- and 2-grams, with few boosting iterations (ie, 50) and shallow trees (ie, a depth of 1). The 1-grams with the highest feature importance were auxiliary verb forms (eg, imperative, infinitive, perfect participle of “sein,” “haben,” and “werden” [ie, “to be,” “to have,” and “to become”]), conjunctions, and adverbs, while the 2-grams comprised verbs and auxiliary verbs (eg, the combinations of a noun and the past participle of “sein,” “haben,” and “werden”), adverbs and verbs (eg, “aber weisst,” “dann sagt,” or “mehr gegessen”).

On the other hand, considering the data augmentation strategy and POS features, the XGB algorithm presented in Multimedia Appendix 1 performs on 10 boosting iterations, in the presence of shallow trees, 2-grams, and K=5 neighbors used by SMOTE. In the case of the class weighted learning strategy and POS features, it performs on 50 boosting iterations, shallow trees, and 2-grams.

The POS 2-grams with the highest feature importance comprised punctuation signs followed by a conjunction (eg, “. Und,” “. Aber,” or “, oder”), adverbs and auxiliary verbs (eg, “dann habe/n” and “da hat/ben”), prepositions and determinative articles (eg, “mit einer/m” and “mit der/m,” [ie, “with a(n)” and “with the”]) and adverbs followed by verbs (eg, “aber weisst,” “mehr gegessen,” and “selber gefahren”).

Finally, considering word embeddings, SVM with radial basis kernel (and scaling) on augmented data outperformed all other classifiers, with K=5 neighbors used by SMOTE (F1=0.44), followed by ADA algorithms (F1=0.42) on K=9 SMOTE neighbors. The model with SVM shows much higher recall (0.59), AUC, and AP, while the model with ADA improves precision (0.43) and consequently specificity (0.97).

Discussion

Principal Results

The primary purpose of this study was to leverage NLP features and machine learning strategies to detect reminiscence in the conversations of older adults in German in a naturalistic observation study. We used the written transcripts of the conversations and a manually coded variable (reminiscence or not reminiscence) as the basis for the prediction. We considered a wide array of methodologies, including different NLP features (ie, bag-of-words, POS tagging, and pretrained German word embeddings), multiple machine learning algorithms, and learning strategies to handle class imbalance. Results indicate that selected combinations (see below) of learning strategies, NLP features, and machine learning models show the potential to detect reminiscence. We argue that their performance can be further improved through feature engineering, by combining NLP features, using NLP-driven data augmentation techniques and collecting more data.

Learning Strategies

Class weighting SVM outperforms others machine learning models, with the highest performance seen when trained on bag-of-words (F1=0.48) with a balanced precision-recall trade-off, high specificity (0.98), AUC=0.91, and AP=0.56. MC strategies show lower F1 scores than class weighted and data augmentation ones, with highly imbalanced precision-recall trade-offs, resulting in low precision and, consequently, low F1 score and specificity. On the other hand, data augmentation with SMOTE delivers performance comparable to the one of class weighting, although only when using word embeddings. However, SMOTE is a purely computational approach to data augmentation, as it generates new data points from any numerical representation of transcripts (eg, on word embeddings). On the other hand, data augmentation algorithms, such as replacing with synonyms, random inserting, swapping, or deleting words [71], process transcripts directly and have improved text classification performance across multiple data sets [71]. We will further investigate the use of NLP-driven data augmentation methodologies and feature engineering [72] in forthcoming studies.

Machine Learning Classifiers

Overall, SVM delivered the highest performance across the class-weighted and data augmentation learning strategies, and boosting methodologies for the MC strategy. SVM proved to exhibit competitive performance in NLP tasks with imbalanced data sets, such as detecting offensive language (including German) [55,73,74], while the XGB algorithm proved its effectiveness in a vast array of machine learning problems [75,76]. Consequently, RF learning methods were outperformed by all learning strategies and NLP features.

NLP Features

Bag-of-words and word embeddings delivered the highest performance of all learning strategies; machine learning models trained on POS features delivered low performance due to their low precision (combined with low recall, such as in the case of ADA for class weighting and data augmentation strategies).

Considering models trained on bag-of-words features, German stopwords related to past tense (eg, "gewesen"), personal pronouns (eg, "ich" and "wir"), and connecting words (eg, "und" and "aber") showed high feature importance, suggesting that they can encode the time, personal narrative, and structure of reminiscence. Models trained on POS tags, however, use both 1-gram and 2-grams; in particular, 2-grams comprising punctuation signs followed by a conjunction, such as ". Und" and ". Aber," suggest that reminiscing is a multisentence act during which the speaker pauses and uses forms of emphasis to elevate certain clauses to positions of more influence and importance. We will further investigate these points in future studies, introducing linguistic measures for reminiscence transcripts like other studies [24,77] to improve bag-of-words and POS models with additional sets of features to encode the phrase structure of a sentence and combining bag-of-words and POS features together.

Error Analysis

Based on an error analysis conducted on the false positives and false negatives for all the models in Table 2, the models tend to predict long transcripts with multiple sentences referring to the past incorrectly as reminiscence (ie, false positives). This results in low precision, affecting the F1 score of all classifiers. On the other hand, false negatives (ie, transcripts incorrectly classified as not reminiscence) are typically short transcripts with few words; more coded transcriptions will help to reduce both errors, in particular the false positives, enlarging the corpus of conversations and improving vocabulary richness.

In this study, we considered only transcripts in German; therefore, we have no specific insight into the specific challenges of detecting reminiscence in another language, such as English. However, earlier studies [24] have compared the application of NLP methods on transcripts of daily conversations in both German and English. Evaluated on the same set of methods and analysis pipeline, the results have suggested no significant difference in performance between the two languages. The use of different NLP features, classifiers, and learning strategies discussed in this study seems promising to develop a system for the real-time detection of reminiscence in everyday conversations in German of older adults. Such a system could leverage audio-to-text software [78] of advanced methods from automated coding [24] to automate the transcription of conversations before NLP preprocessing and the computation of machine learning predictions.

Limitations

This study has several limitations. The data set of conversations used to classify reminiscence had a limited number of records due to the short duration of the naturalistic observation study (4 days) [6]. Moreover, as we considered data from a single observation study, we cannot infer the generalizability of the presented results to other data sources. The results are based on a single partitioning of data into train and test sets. In this study, we used SpaCy pretrained German word embeddings; clearly, these embeddings are trained on corpora that may not fully encode the linguistic specificities characterizing conversations among older adults. In addition, due to the small data set available for this study, we refrained from training embeddings as well as performing machine learning with data-intensive deep learning models [79].

Comparison With Prior Work

Previous research focused on reminiscence as a therapy against dementia have shown the efficacy of information and communication technologies [14]; in particular, studies addressing the topic of automated reminiscence therapy aim at eliciting reminiscence from the users rather than detecting the presence of reminiscence in everyday life [17,18]. These studies have assumed that reminiscence is elicited in a setting with a human companion (eg, a therapist) or through a digital companion device using the Wizard of Oz technique [19]. Few studies have attempted to identify the presence of reminiscence with methods from pattern recognition and machine learning. For example, Naini et al [80] proposed a machine learning model for ranking posts from social media to create life summaries and retain memorable Facebook posts, that is, posts in a user's timeline worth remembering. A retention model based on the learning-to-rank RankSVM algorithm [81] selects posts. Alternatively, Kikhia et al [82] proposed a method to identify places of importance in lifelogging events using clustering methods and locational sensor data. The use of lifelogs allows individuals to reminisce by recalling memories, experiences, and valuable past events for fun, as a personal diary, or as a support for people with memory problems [83].

Data from Demiray et al's observational study [6] have been recently used in studies using NLP methodologies and machine learning. For example, one study [24] used their data to address the more general problem of coding social and cognitive variables from the transcripts of older adults' daily conversations. They trained multiple families of classifiers (ie, SVM, decision trees, and RF) on NLP features, including latent Dirichlet allocations and data augmentation, achieving good performance in different categories. Another study [77] used their data to compute linguistic features (ie, entropy and number of clauses) to examine the effects of age on the use of real-life language in contexts with social interactions. However, to our knowledge, this is the first study investigating the detection of reminiscence directly from the transcripts of daily conversations. Due to the presence of imbalanced data and the type of language (German) used in the conversations, this study addresses problems that are only partially present in prior scientific works in the reminiscence literature.

Conclusions

This work provides evidence to support the use of NLP and machine learning in detecting reminiscence based on written transcripts of older adults' everyday conversations in German. These results represent a novelty in the literature on reminiscence. The proposed methodology can be applied to larger natural observation studies to investigate the differences in reminiscing among cultures, countries, and social contexts using, in particular, lexical features and linguistic measures.

The aim of the WHO healthy aging model is to empower individuals to observe, measure, monitor, and take earlier action for their own health; reminiscence is an essential activity for promoting healthy aging among older adults. Real-time detection

of reminiscence in everyday conversations can offer valuable information for older adults to understand their own health behaviors in everyday life and to support their autonomy in health maintenance.

Therefore, the training of high-performance classifiers supports the design of digital health interventions to improve older adults' quality of life by supporting their healthy aging through the real-time monitoring of reminiscence events in everyday conversations. The digital interventions may consist of a daily diary function to support the contextualization of identified reminiscing events and suggest reminiscence-related positive activities, such as replaying positive life stories and prompting more active social interactions. Another possible implementation is a conversational agent–based digital support function to support users in detecting dysfunctional reminiscence events by suggesting coping strategies. The effectiveness of the interventions could be measured by quantifying their effect on depression, quality of life, and social behavior scores, therefore assessing the impact of reminiscence therapies on healthy aging in everyday life contexts.

Acknowledgments

AF thanks Dominik Stammbach for fruitful conversations on NLP. BD thanks Marianne Mischler for the manual coding of reminiscence and reminiscence functions. KY is funded by the European Social Fund, project number ESF/14-BM-A55-0018/19. MM is partially supported by Velux Stiftung Grant 917.

Authors' Contributions

AF is the corresponding author for this work; he designed and evaluated the NLP and machine learning pipelines for this study, and drafted the manuscript. BD was responsible of the original EAR study design and data collection. BD, ML, and MM provided important intellectual inputs on reminiscence theory, the healthy aging model, and the naturalistic observational study, which helped refine the NLP pipelines. KY provided important intellectual inputs on the NLP and machine learning pipelines, as well as revising them. All authors contributed substantially to paper revisions and its finalization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Machine learning modeling: all results.
[DOCX File , 16 KB - jmir_v22i9e19133_app1.docx]

References

1. World report on ageing and health. World Health Organization. 2015. URL: http://www.who.int/ageing/publications/world-report-2015/en/ [accessed 2020-03-28]
2. Martin M, Weibel R, Röcke C, Boker SM. Semantic Activity Analytics for Healthy Aging: Challenges and Opportunities. IEEE Pervasive Comput 2018 Jul;17(3):73-77. [doi: 10.1109/mprv.2018.03367738]
3. Seifert A, Christen M, Martin M. Willingness of Older Adults to Share Mobile Health Data with Researchers. GeroPsych 2018 Mar;31(1):41-49. [doi: 10.1024/1662-9647/a000181]
4. Cesari M, Araujo de Carvalho I, Amuthavalli Thiyagarajan J, Cooper C, Martin FC, Reginster J, et al. Evidence for the Domains Supporting the Construct of Intrinsic Capacity. J Gerontol A Biol Sci Med Sci 2018 Nov 10;73(12):1653-1660. [doi: 10.1093/gerona/gly011] [Medline: 29408961]
5. Butler RN. The life review: an interpretation of reminiscence in the aged. Psychiatry 1963 Feb;26:65-76. [doi: 10.1080/00332747.1963.11023339] [Medline: 14017386]
6. Demiray B, Mischler M, Martin M. Reminiscence in Everyday Conversations: A Naturalistic Observation Study of Older Adults. J Gerontol B Psychol Sci Soc Sci 2019 Jun 14;74(5):745-755. [doi: 10.1093/geronb/gbx141] [Medline: 29190392]
7. Bluck S, Levine L. Reminiscence as autobiographical memory: a catalyst for reminiscence theory development. Ageing and Society 1998 Mar 01;18(2):185-208. [doi: 10.1017/s0144686x98006862]
8. O'Rourke N, Cappeliez P, Claxton A. Functions of reminiscence and the psychological well-being of young-old and older adults over time. Aging Ment Health 2011 Mar;15(2):272-281. [doi: 10.1080/13607861003713281] [Medline: 21140308]
9. Bluck S, Alea N, Demiray B. You get what you need: The psychosocial functions of remembering. In: Mace JH, editor. New perspectives in cognitive psychology. The act of remembering: Toward an understanding of how we recall the past. Hoboken, New Jersey: Wiley-Blackwell; 2010:284-307.
10. Webster JD, Bohlmeijer ET, Westerhof GJ. Mapping the Future of Reminiscence: A Conceptual Guide for Research and Practice. Res Aging 2010 May 24;32(4):527-564. [doi: 10.1177/0164027510364122]
11. Westerhof G, Bohlmeijer E, Webster J. Reminiscence and mental health: a review of recent progress in theory, research and interventions. Ageing and Society 2010 Mar 01;30(4):697-721. [doi: 10.1017/s0144686x09990328]

12. Pinquart M, Forstmeier S. Effects of reminiscence interventions on psychosocial outcomes: a meta-analysis. Aging Ment Health 2012;16(5):541-558. [doi: 10.1080/13607863.2011.651434] [Medline: 22304736]
13. Webster JD, Cappeliez P. Reminiscence and Autobiographical Memory: Complementary Contexts for Cognitive Aging Research. Developmental Review 1993 Mar;13(1):54-91. [doi: 10.1006/drev.1993.1003]
14. Lazar A, Thompson H, Demiris G. A systematic review of the use of technology for reminiscence therapy. Health Educ Behav 2014 Oct;41(1 Suppl):51S-61S [FREE Full text] [doi: 10.1177/1090198114537067] [Medline: 25274711]
15. Schwarz N. Why researchers should think "real-time": A cognitive rationale. In: Mehl MR, Conner TS, editors. Handbook of research methods for studying daily life. US: The Guilford Press; 2012:22-42.
16. Bielak AAM. How can we not 'lose it' if we still don't understand how to 'use it'? Unanswered questions about the influence of activity participation on cognitive performance in older age--a mini-review. Gerontology 2010;56(5):507-519 [FREE Full text] [doi: 10.1159/000264918] [Medline: 19996570]
17. Caros M, Garolera M, Radeva P, Giro-i-Nieto X. Automatic Reminiscence Therapy for Dementia. ArXiv. URL: http://arxiv.org/abs/1910.11949 [accessed 2019-10-25]
18. Nikitina S, Callaioli S, Baez M. Smart conversational agents for reminiscence. In: SE4COG '18: Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services. New York, NY, USA: Association for Computing Machinery; 2018 Presented at: ICSE '18: 40th International Conference on Software Engineering; May 2018; Gothenburg Sweden p. 52-57. [doi: 10.1145/3195555.3195567]
19. Vardoulakis L, Ring L, Barry B, Sidner C, Bickmore T. Designing Relational Agents as Long Term Social Companions for Older Adults. In: Nakano Y, Neff M, Paiva A, Walker M, editors. IVA'12: Proceedings of the 12th international conference on Intelligent Virtual Agents. Berlin, Heidelberg: Springer; 2012:289-302.
20. Demiray B, Luo M, Tejeda-Padron A, Mehl MR. Sounds of Healthy Aging: Assessing Everyday Social and Cognitive Activity from Ecologically Sampled Ambient Audio Data. In: Hill P, Allemand M, editors. Personal Healthy Aging Adulthood New Dir Tech. Cham: Personal Healthy Aging Adulthood New Dir Tech Internet; 2020:111-132.
21. Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2018 Mar 01;3:CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120.pub3] [Medline: 29493789]
22. Subramaniam P, Woods B. The impact of individual reminiscence therapy for people with dementia: systematic review. Expert Rev Neurother 2012 May;12(5):545-555. [doi: 10.1586/ern.12.35] [Medline: 22550983]
23. Seyfert-Margolis V. The evidence gap. Nat Biotechnol 2018 Mar 06;36(3):228-232. [doi: 10.1038/nbt.4097] [Medline: 29509737]
24. Yordanova K, Demiray B, Mehl MR, Martin M. Automatic Detection of Everyday Social Behaviours and Environments from Verbatim Transcripts of Daily Conversations. In: Proc IEEE Int Conf Pervasive Comput Commun. Proc IEEE Int Conf Pervasive Comput Commun Internet Kyoto, Japan: IEEE; 2019 Presented at: Pervasive Computing and Communications (PerCom), 2019 IEEE International Conference on; March; Kyoto, Japan p. 242-251. [doi: 10.1109/percom.2019.8767403]
25. Liew J, McCracken N, Zhou S, Crowston K. Optimizing Features in Active Machine Learning for Complex Qualitative Content Analysis. USA: Association for Computational Linguistics; 2014 Presented at: Proc ACL 2014 Workshop Lang Technol Comput Soc Sci; June; Baltimore, MD, USA p. 44-48. [doi: 10.3115/v1/w14-2513]
26. Colneric N, Demsar J. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. IEEE Trans. Affective Comput 2020 Jul 1;11(3):433-446. [doi: 10.1109/taffc.2018.2807817]
27. Joshi A, Sparks R, McHugh J, Karimi S, Paris C, MacIntyre CR. Harnessing Tweets for Early Detection of an Acute Disease Event. Epidemiology 2020;31(1):90-97. [doi: 10.1097/ede.0000000000001133]
28. Alburez-Gutierrez D, Zagheni E, Aref S, Gil-Clavel S, Grow A, Negraia D. Demography in the Digital Era: New Data Sources for Population Research. SocArXiv. 2019. URL: https://osf.io/preprints/socarxiv/24jp7/
29. Funk B, Sadeh-Sharvit S, Fitzsimmons-Craft EE, Trockel MT, Monterubio GE, Goel NJ, et al. A Framework for Applying Natural Language Processing in Digital Health Interventions. J Med Internet Res 2020 Feb 19;22(2):e13855 [FREE Full text] [doi: 10.2196/13855] [Medline: 32130118]
30. Cappeliez P, O'Rourke N. Empirical validation of a model of reminiscence and health in later life. J Gerontol B Psychol Sci Soc Sci 2006 Jul 01;61(4):P237-P244. [doi: 10.1093/geronb/61.4.p237] [Medline: 16855036]
31. Korte J, Bohlmeijer ET, Westerhof GJ, Pot AM, Pot AM. Reminiscence and adaptation to critical life events in older adults with mild to moderate depressive symptoms. Aging Ment Health 2011 Jul 01;15(5):638-646. [doi: 10.1080/13607863.2010.551338] [Medline: 21815856]
32. O'Rourke N, King DB, Cappeliez P. Reminiscence functions over time: consistency of self functions and variation of prosocial functions. Memory 2016 May 04;25(3):403-411. [doi: 10.1080/09658211.2016.1179331] [Medline: 27145425]
33. Boker SM, Brick TR, Pritikin JN, Wang Y, von Oertzen T, Brown D, et al. Maintained Individual Data Distributed Likelihood Estimation (MIDDLE). Multivariate Behav Res 2015;50(6):706-720 [FREE Full text] [doi: 10.1080/00273171.2015.1094387] [Medline: 26717128]
34. Al-Rubaie M, Chang JM. Privacy-Preserving Machine Learning: Threats and Solutions. IEEE Secur Privacy 2019 Mar;17(2):49-58. [doi: 10.1109/msec.2018.2888775]

35. Mehl MR, Pennebaker JW, Crow DM, Dabbs J, Price JH. The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. Behav Res Methods Instrum Comput 2001 Nov;33(4):517-523. [doi: 10.3758/bf03195410] [Medline: 11816455]
36. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
37. Mehl MR. The Electronically Activated Recorder (EAR): A Method for the Naturalistic Observation of Daily Social Behavior. Curr Dir Psychol Sci 2017 Apr;26(2):184-190 [FREE Full text] [doi: 10.1177/0963721416680611] [Medline: 28529411]
38. Jurafsky D, Martin JH. Speech and Language Processing (2nd edition). Upper Saddle River, NJ, USA: Prentice-Hall, Inc; 2009.
39. Manning C, Schütze H. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press; 1999.
40. Chen J, Lalor J, Liu W, Druhl E, Granillo E, Vimalananda VG, et al. Detecting Hypoglycemia Incidents Reported in Patients' Secure Messages: Using Cost-Sensitive Learning and Oversampling to Reduce Data Imbalance. J Med Internet Res 2019 Mar 11;21(3):e11990 [FREE Full text] [doi: 10.2196/11990] [Medline: 30855231]
41. Leroy G, Gu Y, Pettygrove S, Galindo MK, Arora A, Kurzius-Spencer M. Automated Extraction of Diagnostic Criteria From Electronic Health Records for Autism Spectrum Disorders: Development, Evaluation, and Application. J Med Internet Res 2018 Nov 07;20(11):e10497 [FREE Full text] [doi: 10.2196/10497] [Medline: 30404767]
42. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VV. Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study. J Med Internet Res 2018 Jun 29;20(6):e231 [FREE Full text] [doi: 10.2196/jmir.9702] [Medline: 29959110]
43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
44. spaCy | Industrial-strength Natural Language Processing in Python. URL: https://spacy.io/ [accessed 2020-04-06]
45. Harris ZS. Distributional Structure. WORD 1954 Aug;10(2-3):146-162. [doi: 10.1080/00437956.1954.11659520]
46. German | spaCy Models Documentation. URL: https://spacy.io/models/de [accessed 2020-04-06]
47. Brants S, Dipper S, Eisenberg P, Hansen-Schirra S, König E, Lezius W, et al. TIGER: Linguistic Interpretation of a German Corpus. Res Lang Comput 2004 Dec 1;2(4):597-620. [doi: 10.1007/s11168-004-7431-3]
48. Nachrichten aus Deutschland und der Welt | Frankfurter Rundschau. URL: https://www.fr.de/ [accessed 2020-04-06]
49. Nothman J, Ringland N, Radford W, Murphy T, Curran JR. Learning multilingual named entity recognition from Wikipedia. Artificial Intelligence 2013 Jan 01;194:151-175. [doi: 10.1016/j.artint.2012.03.006]
50. Fernández A, García S, Galar M, Prati R, Krawczyk B, Herrera F. Learning from Imbalanced Data Sets 1st edition. New York, NY: Springer; 2018.
51. He H, Ma Y. Imbalanced Learning: Foundations, Algorithms, and Applications, 1st edition. Hoboken, New Jersey: Wiley-IEEE Press; 2013.
52. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 2016 Apr 22;5(4):221-232. [doi: 10.1007/s13748-016-0094-0]
53. López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences 2013 Nov 20;250:113-141. [doi: 10.1016/j.ins.2013.07.007]
54. Padurariu C, Breaban ME. Dealing with Data Imbalance in Text Classification. Procedia Computer Science 2019 Jan 01;159:736-745. [doi: 10.1016/j.procs.2019.09.229]
55. GermEval 2018 – Germeval Task 2, 2019 — Shared Task on the Identification of Offensive Language. URL: https://projects.fzai.h-da.de/iggsa/germeval-2018/ [accessed 2020-04-06]
56. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
57. Hastie T, Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. New York, NY, USA: Springer Series in Statistics Springer New York Inc; 2001.
58. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco California USA p. 785-794. [doi: 10.1145/2939672.2939785]
59. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006 Jun 01;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]
60. Jin Huang, Ling C. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng 2005 Mar;17(3):299-310. [doi: 10.1109/TKDE.2005.50]
61. sklearn.metrics.average_precision_score — scikit-learn 0.22.2 documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html [accessed 2020-04-06]
62. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015 Mar 04;10(3):e0118432 [FREE Full text] [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]

63. López V, Fernández A, Moreno-Torres JG, Herrera F. Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. Expert Systems with Applications 2012 Jun;39(7):6585-6608. [doi: 10.1016/j.eswa.2011.12.043]
64. Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Predicting Hospital Readmission via Cost-Sensitive Deep Learning. IEEE/ACM Trans Comput Biol Bioinform 2018 Dec;15(6):1968-1978. [doi: 10.1109/TCBB.2018.2827029] [Medline: 29993930]
65. sklearn.ensemble.RandomForestClassifier — scikit-learn 0.22.2 documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html [accessed 2020-04-06]
66. XGBoost Parameters — xgboost 1.1.0 documentation. URL: https://xgboost.readthedocs.io/en/latest/parameter.html [accessed 2020-04-06]
67. Rokach L. Ensemble-based classifiers. Artif Intell Rev 2009 Nov 19;33(1-2):1-39. [doi: 10.1007/s10462-009-9124-7]
68. Siers M, Islam M. Software defect prediction using a cost sensitive decision forest and voting, and a potential solution to the class imbalance problem. Information Systems 2015 Jul;51:62-71 [FREE Full text] [doi: 10.1016/j.is.2015.02.006]
69. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F, Khalili D. The Impact of Oversampling with SMOTE on the Performance of 3 Classifiers in Prediction of Type 2 Diabetes. Med Decis Making 2016 Jan;36(1):137-144. [doi: 10.1177/0272989X14560647] [Medline: 25449060]
70. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 2013 Mar 22;14(1):106 [FREE Full text] [doi: 10.1186/1471-2105-14-106] [Medline: 23522326]
71. Wei J, Zou K. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).: Association for Computational Linguistics; 2019 Presented at: EMNLP | IJCNLP; November 2019; Hong Kong, China p. 6382-6388 URL: https://www.aclweb.org/anthology/D19-1670/ [doi: 10.18653/v1/d19-1670]
72. Nobata C, Tetreault J, Thomas A, Mehdad Y, Chang Y. Abusive Language Detection in Online User Content. In: Proc 25th Int Conf World Wide Web. 2016 Presented at: WWW '16: 25th International World Wide Web Conference; April; Montréal, Québec, Canada. [doi: 10.1145/2872427.2883062]
73. Schmidt A, Wiegand M. A Survey on Hate Speech Detection using Natural Language Processing. In: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media. Proc Fifth Int Workshop Nat Lang Process Soc Media Internet Valencia, Spain: Association for Computational Linguistics; 2017 Presented at: SocialNLP | WS; April; Valencia, Spain p. 1-10. [doi: 10.18653/v1/w17-1101]
74. Wiegand M, Ruppenhofer J, Schmidt A, Greenberg C. Inducing a Lexicon of Abusive Words – a Feature-Based Approach. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers).: Association for Computational Linguistics; 2018 Presented at: NAACL; June 2018; New Orleans, Louisiana p. 1046-1056 URL: https://www.aclweb.org/anthology/N18-1095/ [doi: 10.18653/v1/n18-1095]
75. The Present and the Future of the KDD Cup Competition: an Outsider's Perspective. URL: https://www.linkedin.com/pulse/present-future-kdd-cup-competition-outsiders-ron-bekkerman/ [accessed 2020-04-06]
76. Ferrario A, Hämmerli R. On Boosting: Theory and Applications. Rochester, NY: Social Science Research Network; 2019. URL: https://papers.ssrn.com/abstract=3402687 [accessed 2020-04-06]
77. Luo M, Schneider G, Martin M, Demiray B. Cognitive Aging Effects on Language Use in Real-Life Contexts: A Naturalistic Observation Study. In: University of Zurich. 2019 Presented at: 41st Annual Conference of the Cognitive Science Society; July 24-27, 2019; Montreal URL: https://www.zora.uzh.ch/id/eprint/174992/ [doi: 10.5167/uzh-174992]
78. Ziman K, Heusser AC, Fitzpatrick PC, Field CE, Manning JR. Is automatic speech-to-text transcription ready for use in psychological experiments? Behav Res Methods 2018 Dec 23;50(6):2597-2605. [doi: 10.3758/s13428-018-1037-4] [Medline: 29687235]
79. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016.
80. Naini KD, Kawase R, Kanhabua N, Niederée C, Altingovde IS. Those were the days: learning to rank social media posts for reminiscence. Inf Retrieval J 2018 Aug 11;22(1-2):159-187. [doi: 10.1007/s10791-018-9339-9]
81. Joachims T. Optimizing search engines using clickthrough data. In: Proc Eighth ACM SIGKDD Int Conf Knowl Discov Data Min. Edmonton, Alberta, Canada: Association for Computing Machinery; 2002 Presented at: KDD02: The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; July; Edmonton, Alberta, Canada p. 133-142. [doi: 10.1145/775047.775067]
82. Kikhia B, Boytsov A, Hallberg J, ul Hussain Sani Z, Jonsson H, Synnes K. Structuring and Presenting Lifelogs Based on Location Data. In: Cipresso P, Matic A, Lopez G, editors. Pervasive Computing Paradigms for Mental Health. Cham: Springer International Publishing; 2014:133-144.
83. Kikhia B, Hallberg J, Bengtsson JE, Savenstedt S, Synnes K. Building digital life stories for memory support. IJCIH 2010;1(2):161. [doi: 10.1504/ijcih.2010.037460]

Abbreviations

ADA: adaptive boosting
AP: average precision
AUC: area under the curve
CLW: class-weighted learning
EAR: Electronically Activated Recorder
NLP: natural language processing
POS: part-of-speech
RF: random forests
SMOTE: Synthetic Minority Oversampling Technique
SVM: support vector machines
tf–idf: term frequency–inverse document frequency
WHO: World Health Organization
XGB: extreme gradient boosting

Edited by G Eysenbach; submitted 06.04.20; peer-reviewed by A Kulakli, C Basch, J Chen; comments to author 28.04.20; revised version received 27.05.20; accepted 11.08.20; published 15.09.20.

Please cite as:
Ferrario A, Demiray B, Yordanova K, Luo M, Martin M
Social Reminiscence in Older Adults' Everyday Conversations: Automated Detection Using Natural Language Processing and Machine Learning
J Med Internet Res 2020;22(9):e19133
URL: http://www.jmir.org/2020/9/e19133/
doi:10.2196/19133
PMID:32866108

Original Paper

Associations Between Substance Use and Instagram Participation to Inform Social Network–Based Screening Models: Multimodal Cross-Sectional Study

Brandon G Bergman[1*], PhD; Weiyi Wu[2*], MSc; Lisa A Marsch[3], PhD; Benjamin S Crosier[4], PhD; Timothy C DeLise[5], MSc; Saeed Hassanpour[4], PhD

[1]Recovery Research Institute, Center for Addiction Medicine, Massachusetts General Hospital, & Harvard Medical School, Boston, MA, United States
[2]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH, United States
[3]Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States
[4]Departments of Biomedical Data Science, Computer Science, and Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH, United States
[5]Department of Mathematics and Statistics, Universite de Montreal, Montreal, QC, Canada
[*]these authors contributed equally

Corresponding Author:
Brandon G Bergman, PhD
Recovery Research Institute
Center for Addiction Medicine
Massachusetts General Hospital, & Harvard Medical School
151 Merrimac Street, 6th Floor
Boston, MA, 02114
United States
Phone: 1 617 643 7563
Email: bgbergman@mgh.harvard.edu

Abstract

Background: Technology-based computational strategies that leverage social network site (SNS) data to detect substance use are promising screening tools but rely on the presence of sufficient data to detect risk if it is present. A better understanding of the association between substance use and SNS participation may inform the utility of these technology-based screening tools.

Objective: This paper aims to examine associations between substance use and Instagram posts and to test whether such associations differ as a function of age, gender, and race/ethnicity.

Methods: Participants with an Instagram account were recruited primarily via Clickworker (N=3117). With participant permission and Instagram's approval, participants' Instagram photo posts were downloaded with an application program interface. Participants' past-year substance use was measured with an adapted version of the National Institute on Drug Abuse Quick Screen. At-risk drinking was defined as at least one past-year instance having "had more than a few alcoholic drinks a day," drug use was defined as any use of nonprescription drugs, and prescription drug use was defined as any nonmedical use of prescription medications. We used logistic regression to examine the associations between substance use and any Instagram posts and negative binomial regression to examine the associations between substance use and number of Instagram posts. We examined whether age (18-25, 26-38, 39+ years), gender, and race/ethnicity moderated associations in both logistic and negative binomial models. All differences noted were significant at the .05 level.

Results: Compared with no at-risk drinking, any at-risk drinking was associated with both a higher likelihood of any Instagram posts and a higher number of posts, except among Hispanic/Latino individuals, in whom at-risk drinking was associated with a similar number of posts. Compared with no drug use, any drug use was associated with a higher likelihood of any posts but was associated with a similar number of posts. Compared with no prescription drug use, any prescription drug use was associated with a similar likelihood of any posts and was associated with a lower number of posts only among those aged 39 years and older. Of note, main effects showed that being female compared with being male and being Hispanic/Latino compared with being White were significantly associated with both a greater likelihood of any posts and a greater number of posts.

Conclusions: Researchers developing computational substance use risk detection models using Instagram or other SNS data may wish to consider our findings showing that at-risk drinking and drug use were positively associated with Instagram participation,

while prescription drug use was negatively associated with Instagram participation for middle- and older-aged adults. As more is learned about SNS behaviors among those who use substances, researchers may be better positioned to successfully design and interpret innovative risk detection approaches.

(J Med Internet Res 2020;22(9):e21916) doi:10.2196/21916

KEYWORDS

substance use; social network sites; health risk; screening; machine learning; social media; Instagram; alcohol; drug

Introduction

Enhancing the Utility of Technology-Based Strategies to Detect Substance Use

More than 1 in 10 US adults meet diagnostic criteria for alcohol use disorder [1], and 25% engage in binge drinking (4+ drinks for women and 5+ for men in 2 hours) [2]. In addition, 3 in 10 adults use other drugs, like cannabis, stimulants, and opioids [2], and 4% meet diagnostic criteria for a drug use disorder [3]. Together, alcohol and other drug use (ie, substance use) are major public health burdens [4,5] and cost the United States $500 billion annually [6,7]. Only 10% of those with substance use disorder (SUD) seek specialty treatment, and of those who do not seek SUD treatment, 95% do not perceive a treatment need [8]. The reach of public health approaches to boost clinical screening for substance use problems [9] may be enhanced by developing and testing novel, technology-based strategies to identify individuals who use substances, including but not limited to those engaging in harmful use [10]. If effective at identifying such individuals, technology-based strategies could then be paired with alternative and innovative interventions, which may be scaled up and deployed to help individuals with milder SUD variants who may be less likely to seek formal care [11], as well as individuals who might benefit from harm reduction–based psychoeducation. One such technology-based strategy leverages data from commonly used social network sites (SNSs) such as Facebook, Twitter, and Instagram. More than 70% of US adults have at least one SNS account, including nearly 70% who use Facebook, 37% who use Instagram, and 22% who use Twitter [12]. Among young adults specifically, who are disproportionately represented among individuals with SUD and hazardous drinking [2], Instagram is nearly as popular as Facebook, with 67% versus 79%, respectively, having ever used these platforms [12].

As detailed below, emerging research suggests that individuals' SNS posts can be leveraged to detect substance use [10,13,14]. However, these computational strategies function more efficiently and reliably with increasing amounts of data [15]. Thus, if certain types of individuals are less likely to participate or have lower levels of participation on certain SNSs, then it may be more difficult to detect substance use in these individuals, even if models are ostensibly well designed. As such, a greater understanding of the associations among individuals' demographic characteristics, SNS participation, and substance use can help inform the utility of these computational screening tools. In the current study, we examined associations between substance use, including at-risk drinking and drug use, and Instagram posts and tested whether such associations differ as a function of age, gender, and race/ethnicity.

SNS-Based Computational Strategies to Detect Substance Use Risk

In a prior study, our team [10] employed a novel deep neural network (ie, deep learning) framework [16] capable of processing data points with varying dimensions, including text and pictures of various sizes, to detect at-risk drinking and drug use from Instagram images, captions, and comments among a community sample of adults aged 18 and older, recruited primarily through Clickworker (Clickworker GmbH). Findings showed the deep learning model significantly predicted at-risk drinking, defined as having "had more than a few alcoholic drinks a day" at least once in the past year. In this model, the area under the receiver operating characteristic curve (AUROC) was 0.65, meaning there was a 65% chance that this novel classification model would correctly assign a higher score to a random positive example (participant reported at-risk drinking at least once in the past year) than a random negative example (participant reported no at-risk drinking in the past year). While the consensus is that AUROCs less than 0.7 provide low discrimination [17], the proof-of-concept study showed that a combination of visual and text-based SNS data could be used to detect self-reported at-risk drinking. The deep learning model was unable to detect other drug use any better than chance. Of note, the combination of images, captions, and comments evidenced superior detection of at-risk drinking compared with other combinations of Instagram content. To our knowledge, this study represents the only published research to date that has leveraged machine learning to detect individuals' substance use from their SNS data in a nonclinical sample.

Studies showing that online forum content can be used to detect risk in clinical SUD samples are also instructive. Kornfield et al [13,14] used natural language processing in 2 studies, one of which also employed machine learning, to predict risk for negative substance use outcomes among individuals with both alcohol and other drug use disorders participating in a smartphone app online recovery forum. In the first study [14], they examined the utility of natural language processing with a Linguistic Inquiry and Word Count approach to predict binge drinking (5+ drinks for men and 4+ for women in 2 hours) in participants with alcohol use disorder. Individuals in this secondary analysis of a randomized controlled trial testing a smartphone-based intervention after residential treatment [18] were both randomized to receive the smartphone app intervention and participated in the application's online forum between baseline (ie, residential treatment discharge) and the 4-month follow-up. Controlling for individual and system use characteristics (eg, number of messages posted), a greater

percentage of words capturing swearing, negative affect, inhibition/control, and love and a lower percentage of words capturing higher-order cognitive processes (eg, insight) and achievement predicted past-month binge drinking at the 12-month follow-up [14]. In a related study including both participants from Gustafson et al [18] and participants with a range of SUDs recruited from primary care in Quanbeck et al [19], Kornfield et al [13] showed that decision tree machine learning algorithms can be used to detect recovery problems from online forum content. Algorithms included both Linguistic Inquiry and Word Count as well as Bag of Words natural language processing approaches. Recovery problems were determined by the research team based on a codebook informed by forum moderators' perspectives regarding which posts warranted concern or intervention.

It is worth acknowledging the large and growing body of literature on the use of aggregated SNS big data to help surveil public health trends in drug use [20,21]. This big data surveillance literature and the use of SNS data to detect substance use risk share the use of SNS data as a proxy or marker of human behavior, which can then be downloaded for analysis. Aims of the current study, however, are intended to inform strategies that detect substance use at the individual level. Such a risk detection approach, when determined to be sufficiently reliable, would ultimately then be paired with interventions as mentioned above. When using SNS data in the context of drug use surveillance, data are analyzed in aggregate to identify macro-level trends and are thus outside the scope of this study of individual behaviors.

Associations Between SNS Participation and Substance Use

Following from studies of computational methods using SNS data to detect substance use, it is also important to examine whether, and in what contexts, individuals who use substances also engage with SNSs to inform the utility of these technology-based screening tools. In a meta-analysis of 17 studies targeting adolescents and young adults, Curtis et al [22] found that self-reported alcohol consumption, including both general consumption and measures of risky drinking, is moderately associated with self-reported and hand-coded SNS engagement, including both alcohol-related SNS posts as well as exposure to alcohol-related SNS posts of others (overall r=0.36). In a related meta-analysis of 7 studies [22] authors similarly found that alcohol-related problems, as measured, for example, by the Alcohol Use Disorders Identification Test [23], were also moderately associated with SNS engagement (overall r=0.37). Thus, younger individuals may provide enough SNS data for these computational strategies to detect at-risk drinking.

Most studies examining the association between SNS participation and substance use to date have focused on the association between Facebook engagement and drinking. In an exception, Instagram participation was negatively related to past-month days of cannabis use but positively related to alcohol use in emerging adult (ie, aged 18-29 years) Instagram users recruited through Amazon mTurk [24]. Given the literature's focus on Facebook and drinking among youth, studies that include individuals across age groups and that examine the association between substance use, including both alcohol and drug use, and popular platforms other than Facebook, such as Instagram, may help build on this emerging scientific literature.

Demographic Factors That Moderate SNS–Substance Use Relationships

Knowledge about whether demographic characteristics moderate associations between SNS participation and substance use might further improve research on computational detection strategies. If, for example, there was an association between at-risk drinking and SNS participation only for younger individuals, this might reduce the utility of an SNS data–based model in detecting at-risk drinking for older individuals. To date, however, there are no existing studies that examine whether demographic characteristics moderate the association between substance use and SNS participation in community (ie, nonclinical) samples.

In the absence of prior work that might inform whether demographic characteristics moderate SNS–substance use relationships in the current study, it is worth mentioning that, consistent with the general population [12], emerging adults (aged 18-29 years) in SUD treatment report greater SNS participation compared with middle- and older-aged adults [25]. Similarly, in a nationally representative sample of US adults who resolved a substance use problem, emerging adults were more likely than their older counterparts to have used online resources, including but not limited to SNS platforms, to address their substance use or enhance their SUD recovery [26]. Also, in this same nationally representative recovery sample, Hispanic race/ethnicity relative to White race was related to a greater likelihood of recovery-related use of online resources, though men and women reported similar rates of this online help-seeking behavior. While these data are based on treatment and recovery samples, whereas the current study focuses on substance use in nonclinical samples, they suggest that any observed relationships between substance use and SNS participation might differ by age and race/ethnicity.

Summary and Current Study

Emerging technology-based strategies to detect substance use with SNS data hold promise as scalable health risk screening tools. These computational strategies, of course, can only be effective among individuals who participate on SNSs. Even among those who participate, these screening tools are more powerful and, therefore, more useful, with increasing amounts of data. Thus, the real-world utility of these tools can be informed by research examining the associations between substance use, SNS participation, and demographic factors that moderate these relationships. Existing research suggests that substance use, including but not limited to hazardous use, is associated with greater SNS participation, but studies have focused primarily on drinking among young people on Facebook. To expand on this work, the current study targeted Instagram participation among adults of all ages and had the following aims: (1) to examine whether at-risk drinking and drug use is related to Instagram posts and (2) to examine whether these relationships between substance use and Instagram posts are moderated by age, gender, and race/ethnicity. We hypothesized that at-risk drinking would be related to more

Instagram posts; we made no a priori hypotheses about drug use. We also hypothesized that the relationship between at-risk drinking and Instagram posts would be greater among younger individuals; we made no a priori hypotheses about other potential moderating demographic characteristics.

Methods

Procedure

As detailed in Hassanpour et al [10], study participants were recruited in the winter of 2016, primarily through the Clickworker crowdsourcing platform, which compensates individuals directly for study participation. Participants were also recruited via word of mouth and SNS advertisements on social media. Each recruitment avenue directed participants to the study website, where, following online consent, they completed an online survey consisting of demographic information and the National Institute on Drug Abuse's (NIDA's) Quick Screen substance use screener (see the "Measures" section below for more details of the NIDA Quick Screen) [27].

Instagram permitted the use of their application program interface (API) to collect participants' data with individuals' permission. Specifically, upon completion of the survey, the online study site linked participants to the Instagram gateway where they could give their permission to allow an application developed by our team to communicate with the Instagram API. If participants granted permission, their posts were downloaded onto a secure server and stored under an anonymized unique identifier for restricted use in this study. The team piloted the procedure and application, facilitating download from Instagram's API, on 81 individuals, whose data were not included in the final, analyzed sample.

Measures

Demographic Characteristics

Participants indicated their gender (male, female, transgender, or other), age in years, and race/ethnicity (Asian, Black, Hispanic/Latino, Native American/Alaskan Native, Native Hawaiian/Pacific Islander, White). We categorized age into groups of 18 to 25 years, 26 to 38 years, and 39 years or older, with cut-off points approximating 1 SD above and below the mean, while accounting for the theoretically important life stage of "emerging adulthood," sometimes operationalized as ages 18 to 25 [28].

Instagram Posts

Instagram is a photo- and video-based SNS, accessible traditionally by smartphone app but also accessible with limited features via the website www.instagram.com. All participants had an Instagram account, though a subset had no content in their account. Participants were dichotomized into any versus no Instagram photo posts. For those with any posts, we included a count variable measuring the total number of posts. Video posts were excluded, as this content could not be analyzed with machine learning architecture used in the current study (see Hassanpour et al [10] for a detailed description of the machine learning approach).

Substance Use

Adapted from the NIDA Quick Screen [27], individuals reported frequency in the past year—never, once or twice, monthly, weekly, and daily or almost daily—of having "had more than a few alcoholic drinks a day," which we refer to as at-risk drinking, using illegal drugs (ie, cannabis, stimulants, opioids, etc), which we refer to here simply as drug use, and using prescription drugs for nonmedical reasons (ie, opioid painkillers, benzodiazepines, and stimulants for attention-deficit/hyperactivity disorder), which we refer to here simply as prescription drug use. Consistent with NIDA's guidance [27] and our past work [10], individuals who indicated "once or twice" or more frequently for each substance use category were counted as "positive" in screening for at-risk drinking, drug use, and prescription drug use. Of note, the NIDA Quick Screen assesses alcohol use with a criterion of 4+ drinks and 5+ drinks in 1 day for women and men, respectively. In the original study from which these data were derived [10], the research team decided to use having "had more than a few alcoholic drinks a day" instead of the NIDA criterion to reduce recall burden in the context of a very brief questionnaire. Also, while the NIDA Quick Screen includes an item for tobacco products, the current study focused on alcohol and drug use only.

Participants

Out of a total of 3117 participants, 861 (27.62%) had an account with 0 posts and 2256 (72.37%) had an account with one or more posts. For those with one or more posts, the average number of posts was 201.9 (SD 394.7), ranging from 1 to 4999 posts. For substance use, in the past year, 55.34% (1725/3117) engaged in at-risk drinking, 22.30% (695/3117) used drugs, and 15.27% (476/3117) used prescription drugs.

With respect to demographic characteristics, participants were aged 29.9 years on average (SD 9.86; range 18-73 years), with 40.52% (1263/3117) aged between 18 and 25 years, 42.38% (1321/3117) between 26 and 38 years, and 17.10% (533/3117) 39 years and older. For gender, 37.12% (1157/3117) identified as male, 62.14% (1937/3117) as female, 0.38% (12/3117) as transgender, and 0.35% (11/3117) as other gender. For race/ethnicity, 6.48% (202/3117) identified as Asian, 16.27% (507/3117) as Black, 10.27% (320/3117) as Hispanic/Latino, 62.59% (1951/3117) as White, and 4.40% (137/3117) as other (including Alaskan Native/Native American or Hawaiian Native/Pacific Islander).

Analysis Plan

For aim 1, we used logistic regression to examine the main effects of demographic characteristics and substance use variables on having an account with 0 versus any Instagram posts. We used negative binomial regression to examine the main effects of demographic and substance use variables on the number of Instagram posts for individuals. Since Instagram posts are count data and its distribution is overdispersed, we considered both Poisson regression, which assumes that the variance and mean are equal in the dependent variable, and negative binomial regression. Upon examination of the quantile-quantile plot (Multimedia Appendix 1), the distribution

of Instagram posts was similar to a negative binomial distribution, supporting the use of a negative binomial regression to examine those factors associated with the number of Instagram posts [29].

For aim 2, we used the same logistic regression and negative binomial regression models as in aim 1 but examined the interaction effects between each of the demographic characteristics and at-risk drinking, drug use, and prescription drug use in the prediction of Instagram posts.

Of note, negative binomial regression models can yield inflated observed significance levels when there are high numbers of 0-count variables. In addition, the negative binomial regression models were unable to discriminate "true zeros" (ie, individuals who had Instagram accounts but 0 posts) from "artificial zeros" (ie, individuals who had 0 posts because they created accounts to participate in the study). In order to test whether removing those with 0 Instagram posts from the model would alter the pattern of findings, we conducted sensitivity analyses with zero-truncated negative binomial models, as in aims 1 and 2, with only individuals who had accounts with one or more Instagram posts (n=2256). The pattern of results for these sensitivity analyses was nearly identical, with the same significance testing results for all effects. Thus, we present only the primary analyses but include these tabulated sensitivity analysis results in Multimedia Appendix 2.

We used R 3.5.3 (The R Foundation) to conduct study analyses. All analyses tested significance at the .05 level. The institutional review board at Dartmouth College approved all study procedures. Data collection for this research project was conducted with informed consent from all participants and complied with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects.

Results

Associations Between Substance Use and Instagram Posts

Alcohol

At-risk drinking was significantly associated with a greater likelihood of any posts (Table 1) and greater number of posts (Table 2) compared with no at-risk drinking. Controlling for effects of demographic characteristics as well as drug use and prescription drug use, at-risk drinking was uniquely associated with 51.6% greater likelihood of any posts and 88.1% more posts.

Table 1. Logistic regression examining the associations between demographic characteristics, substance use, and the interaction between demographic characteristics and substance use with the likelihood of any Instagram posts.

Explanatory variables	Odds ratio (95% CI)	*P* value
Intercept	0.861 (0.655-1.133)	.29
Age (years)		
18-25	1.596 (1.205-2.114)	.001
26-38	1.000	N/A[a]
39+	0.56 (0.411-0.762)	<.001
Gender[b]		
Male	1.000	N/A
Female	2.677 (2.077-3.452)	<.001
Race		
White	1.000	N/A
Asian	0.933 (0.608-1.433)	.75
Black	1.676 (1.209-2.323)	.002
Hispanic/Latino	1.653 (1.074-2.543)	.02
Other	1.525 (0.831-2.797)	.17
At-risk drinking		
No	1.000	N/A
Yes	1.516 (1.048-2.192)	.03
Drug use		
No	1.000	N/A
Yes	1.771 (1.089-2.881)	.02
Prescription drug use		
No	1.00	N/A
Yes	1.153 (0.676-1.968)	.60
Interaction terms		
Age 26-38 × at-risk drinking	1.000	N/A
Age 18-25 x at-risk drinking	1.292 (0.856-1.950)	.22
Age 39+ × at-risk drinking	0.661 (0.418-1.046)	.08
Age 26-38 × drug use	1.000	N/A
Age 18-25 × drug use	1.238 (0.711-2.154)	.45
Age 39+ × drug use	0.824 (0.411-1.655)	.59
Age 26-38 × prescription drug use	1.000	N/A
Age 18-25 × prescription drug use	0.733 (0.407-1.322)	.30
Age 39+ × prescription drug use	0.613 (0.306-1.227)	.17
Male × at-risk drinking	1.000	N/A
Female × at-risk drinking	1.391 (0.966-2.004)	.08
Male × drug use	1.000	N/A
Female × drug use	0.823 (0.497-1.364)	.45
Male × prescription drug use	1.000	N/A
Female × prescription drug use	0.627 (0.369-1.065)	.08
White × at-risk drinking	1.000	N/A
Asian × at-risk drinking	1.029 (0.471-2.25)	.94

Explanatory variables	Odds ratio (95% CI)	*P* value
Black × at-risk drinking	0.709 (0.431-1.166)	.18
Hispanic/Latino × at-risk drinking	0.992 (0.532-1.848)	.98
Other × at-risk drinking	0.428 (0.17-1.076)	.07
White × drug use	1.000	N/A
Asian × drug use	2.127 (0.536-8.431)	.28
Black × drug use	0.779 (0.404-1.503)	.46
Hispanic/Latino × drug use	0.732 (0.321-1.668)	.46
Other × drug use	2.268 (0.58-8.87)	.24
White × prescription drug use	1.000	N/A
Asian × prescription drug use	1.531 (0.464-5.05)	.49
Black × prescription drug use	1.072 (0.527-2.182)	.85
Hispanic/Latino × prescription drug use	0.745 (0.322-1.727)	.49
Other × prescription drug use	1.46 (0.429-4.97)	.55

[a]N/A: not applicable.

[b]Although participants could report nonbinary gender, logistic regression models excluded other gender and transgender due to small cell sizes.

Table 2. Negative binomial regression examining the associations between demographic characteristics, substance use, and the interaction between demographic characteristics and substance use with the number of Instagram posts.

Explanatory variables	Probability ratio (95% CI)	*P* value
Intercept	55.339 (42.202-72.566)	<.001
Age (years)		
26-38	1.000	N/A[a]
18-25	0.911 (0.702-1.181)	.48
39+	0.45 (0.331-0.612)	<.001
Gender[b]		
Male	1.000	N/A
Female	2.967 (2.326-3.785)	<.001
Race		
White	1.000	N/A
Asian	0.917 (0.602-1.395)	.68
Black	0.905 (0.67-1.221)	.51
Hispanic/Latino	2.005 (1.353-2.969)	<.001
Other	0.822 (0.481-1.404)	.47
At-risk drinking		
No	1.000	N/A
Yes	1.881 (1.316-2.688)	<.001
Drug use		
No	1.000	N/A
Yes	1.289 (0.832-1.997)	.26
Prescription drug use		
No	1.000	N/A
Yes	0.915 (0.559-1.499)	.72
Interaction terms		
Age 26-38 × at-risk drinking	1.000	N/A
Age 18-25 × at-risk drinking	0.81 (0.571-1.148)	.24
Age 39+ × at-risk drinking	0.802 (0.513-1.253)	.33
Age 26-38 × drug use	1.000	N/A
Age 18-25 × drug use	0.959 (0.624-1.476)	.85
Age 39+ × drug use	1.401 (0.716-2.741)	.33
Age 26-38 × prescription drug use	1.000	N/A
Age 18-25 × prescription drug use	0.775 (0.474-1.269)	.31
Age 39+ × prescription drug use	0.242 (0.123-0.476)	<.001
Male × at-risk drinking	1.000	N/A
Female × at-risk drinking	0.82 (0.589-1.143)	.24
Male × drug use	1.000	N/A
Female × drug use	1.073 (0.712-1.618)	.74
Male × prescription drug use	1.000	N/A
Female × prescription drug use	1.1 (0.693-1.747)	.69
White × at-risk drinking	1.000	N/A
Asian × at-risk drinking	1.246 (0.634-2.449)	.52

Explanatory variables	Probability ratio (95% CI)	*P* value
Black × at-risk drinking	1.224 (0.794-1.888)	.36
Hispanic/Latino × at-risk drinking	0.479 (0.284-0.81)	.006
Other × at-risk drinking	1.211 (0.556-2.637)	.63
White × drug use	1.000	N/A
Asian × drug use	0.749 (0.3-1.873)	.54
Black × drug use	1.048 (0.609-1.803)	.87
Hispanic/Latino × drug use	1.375 (0.712-2.655)	.34
Other × drug use	0.835 (0.333-2.093)	.70
White × prescription drug use	1.000	N/A
Asian × prescription drug use	2.076 (0.795-5.421)	.14
Black × prescription drug use	0.738 (0.395-1.377)	.34
Hispanic/Latino × prescription drug use	0.753 (0.361-1.574)	.45
Other × prescription drug use	1.814 (0.679-4.847)	.24

[a]N/A: not applicable.

[b]Although participants could report nonbinary gender, negative binomial models excluded other gender and transgender due to small cell sizes.

Drug Use

Drug use was significantly associated with a greater likelihood of any posts (Table 1) but a similar number of posts (Table 2) compared with no drug use. Controlling for demographic characteristics and other substance use measures, drug use was uniquely associated with a 77% greater likelihood of any posts.

Prescription Drug Use

Prescription drug use was not associated with the likelihood of any posts (Table 1) or number of posts (Table 2).

Moderation of the Instagram–Substance Use Associations by Demographics

Age

The association between prescription drug use and number of posts was moderated by age. For those aged 39 years or older, there was a significantly more negative association between prescription drug use and number of posts than for those aged 26 to 38 years (Table 2). Specifically, for those aged 39 years or older, prescription drug use was associated with 77.9% fewer posts compared with no prescription drug use, a significant effect, but for those aged 26 to 38 years, the association between prescription drug use and number of posts was nonsignificant. Associations for both at-risk drinking and drug use were not significantly moderated by age, for either any or number of Instagram posts.

Of note, the main effects of age showed that being aged 18 to 25 years was associated with a significantly greater likelihood of any posts (Table 1) but a similar number of posts (Table 2) compared with those aged 26 to 38 years. Both being aged 18 to 25 years and 26 to 38 years, compared with being aged 39 years or older, were associated with a significantly greater likelihood of any posts (Table 1) and a greater number of posts (Table 2).

Gender

Gender did not moderate any of the associations between substance use and any posts (Table 1) or number of posts (Table 2). Of note, there was a main effect of gender, such that being female was significantly associated with a greater likelihood of any posts (Table 1) and greater number of posts (Table 2) compared with being male.

Race/Ethnicity

The association between at-risk drinking and number of posts was moderated by race/ethnicity. For those identifying as Hispanic/Latino, the association between at-risk drinking and number of posts was nonsignificant, but for those identifying as White, at-risk drinking was associated with 88.1% more posts, a significant effect. Of note, however, main effects of race/ethnicity showed that compared with White identification, Black identification was significantly associated with a greater likelihood of any posts (Table 1), while Hispanic/Latino identification was significantly associated with a greater likelihood of any posts (Table 1) and greater number of posts (Table 2). Asian identification and other racial/ethnic identification was associated with a similar likelihood of any posts (Table 1) and similar number of posts (Table 2) compared with White identification.

Discussion

Summary of Findings

A greater understanding of the associations between substance use and SNS participation, as well as the demographic characteristics that moderate these associations, may help inform the utility of SNS data–based substance use screening tools [10]. In the present study, we showed that past-year at-risk drinking is associated with a greater likelihood of any Instagram posts. As hypothesized, at-risk drinking was associated with a greater number of posts, too, except for those identifying as Hispanic/Latino. Contrary to hypotheses, the relationship

between at-risk drinking and Instagram posts did not differ as a function of age. Drug use (eg, cannabis, cocaine, heroin, etc) was associated with a greater likelihood of any posts but a similar number of posts compared to no drug use. Relative to no prescription drug use, any prescription drug use (eg, nonmedical use of opioids, benzodiazepines, stimulants, etc) was associated with fewer posts only among those aged 39 years and older, but it was associated with a similar likelihood of any posts, more generally. We outline the implications of these findings for researchers developing and testing strategies that employ SNS data to detect substance use.

Association Between At-Risk Drinking and Instagram Posts

Findings showed that at-risk drinking, defined here as having "had more than a few alcoholic drinks a day" at least once in the past year, is uniquely associated with 88% more Instagram posts. For example, for an individual with demographic characteristics mapping onto all reference groups—male, aged 26 to 38 years, White, and no at-risk substance use—the model predicts 55 Instagram posts, which then increases to a predicted 104 posts if at-risk drinking is reported. Our data, which included adults aged 18 to 73 years and focused on Instagram participation, add to the body of literature showing that more drinking, as well as drinking problems, are associated with Facebook engagement among youth [22]. While needing to be replicated in other community samples, individuals with at-risk drinking may provide more data than their non–at-risk drinking counterparts, which can be leveraged in computational models of SNS data–based substance use risk detection.

We used a liberal definition of at-risk drinking, as recommended by NIDA Quick Screen guidelines, resulting in 55.34% (1725/3117) of the sample meeting at-risk drinking criteria. More than 70% of those defined as at-risk drinkers reported at-risk drinking ("more than a few alcoholic drinks a day") one or two times in the past year [10]. While we did not assess for overall health, it is unlikely that alcohol consumption at this level would cause substantial physical consequences or map onto clinically significant alcohol use disorder. Thus, given that greater SNS participation may reflect greater social capital [30,31], that is, the social resources people can bring to bear on navigating challenges and problem-solving, it is possible that greater social capital may be associated with an increased likelihood of at-risk drinking when defined more liberally [32], as was the case here. In another study of Instagram users, for example, greater Instagram participation was associated with a composite of overall drinking and at-risk drinking (4+ and 5+ drinks in one day for women and men, respectively) only for those with the highest levels of peer belongingness [24]. Individuals with alcohol use disorder, on the other hand, are more likely to have reduced social involvement relative to those without SUD, based on the diagnostic criteria (eg, continuing drinking despite giving up hobbies, occupational or educational consequences, interpersonal difficulties, physical and mental health–related harms, etc). We might hypothesize, therefore, that individuals with alcohol use disorder would produce fewer SNS posts, potentially reducing the sensitivity of SNS data–based computational models. This type of curvilinear, inverted U-shaped relationship between at-risk drinking and SNS posts is speculative and should be tested in future work.

There was no association between at-risk drinking and number of Instagram posts for Hispanic/Latino individuals, but such individuals did have twice as many Instagram posts, overall, relative to White individuals. As such, it seems unlikely that using an SNS data–based method to screen for substance use would be any more challenging in Hispanic/Latino individuals relative to other racial/ethnic groups.

Association Between Drug Use and Instagram Posts

Consistent with the NIDA Quick Screen [27], we analyzed drug use (ie, nonprescription drug use), such as cannabis, heroin, and cocaine, separately from prescription drug use (ie, nonmedical use of prescription medications), such as opioid painkillers, benzodiazepines, and stimulants, with disparate findings. Although drug use was associated with a 77% greater likelihood of any posts and similar number of posts, prescription drug use was generally not associated with either the likelihood of any posts or number of posts. For individuals 39 years and older, compared with those aged 26 to 38 years, prescription drug use may be associated with a *lower* number of posts. Thus, computational models that use Instagram data may have fewer posts with which to work if specifically aiming to detect prescription drug use among middle- and older-aged adults.

There are few prior studies of the association between SNS participation and drug use to which the current findings can be compared and contextualized. Exceptions have focused on cannabis, given it is the most widely used drug apart from alcohol [2] and its recreational use is now legal in Canada [33] and several states in the United States [34]. Bergman et al [24], for example, found cannabis use was negatively related to Instagram participation in a community sample of emerging adults aged 18-29 years. As the NIDA Quick Screen [27] queries frequency of use aggregated across drug types, we could not ascertain the association between Instagram posts and specific types of drug use. Studies that examine the utility of technology-based screening tools for opioid use in the context of the opioid overdose crisis [35], for example, may be warranted.

Sample Generalizability

Study findings derive largely from participants recruited via Clickworker, a crowdsourced pay-for-performance site. While observed associations between certain demographic characteristics and Instagram post behaviors are not surprising, they are worth special mention, given their similarity with epidemiological data derived from nationally representative surveys among US adults. Specifically, we found that individuals aged 18 to 25 years were more likely to have at least one post compared with those aged 26 to 38 years, who, in turn, were more likely to have a post than those 39 years and older. Similarly, Pew Research Center [12] reports that 67% of individuals aged 18 to 29 years have an Instagram account, while 47% and 23% of individuals aged 30 to 49 years and 50 to 64 years, respectively, have an account. We found women were twice as likely to have at least one Instagram post compared to men, while Pew reported 43% of women to have

an Instagram account compared to 31% of men. Finally, we found that Black and Hispanic/Latino individuals had the highest rates of Instagram engagement compared with other races/ethnicities, while Pew reported that 40% of Black and 51% of Hispanic/Latino individuals had an Instagram account compared to only 33% of White individuals. Thus, crowdsourced pay-for-performance or microtask sites, such as Clickworker and Amazon mTurk, may be reliable ways to achieve demographically representative groups of Instagram participants.

Limitations

The following methodological limitations may be used to contextualize the study's findings. First, when collecting the data, we did not capture the dates of each Instagram post and, by association, we are unable to determine how long individuals had been using Instagram. Given that computational models using SNS data are generally targeting current substance use, we would ideally be able to examine the association between substance use risk and recent SNS posts but were unable to do so with the current study methods. Second, the reasons that individuals had Instagram accounts with 0 posts remain unclear. For example, it may be that they simply observe the accounts of others (eg, "lurkers"), or they may have created an account for the sole purpose of participating in the current study to obtain compensation. Our analytic approach including both logistic regression (ie, any posts) and negative binomial regression (ie, number of posts) helps minimize the potential for such behavior to impact our pattern of findings. That said, our ability to interpret the real-world implications of the Instagram post outcomes is somewhat limited without this context. Finally, there were a set of limitations related to our substance use assessment. As mentioned above, our decision to target any instances of at-risk alcohol, prescription, and other drug use in the past year was consistent with the NIDA Quick Screen [27] but nevertheless constitutes a highly sensitive approach to identifying risk. While the detection of any substance use may aid critical prevention initiatives, future studies may also disentangle the reach of SNSs, including but not limited to Instagram, in detecting any substance use from their reach in detecting more clearly harmful variants (eg, screening tools for alcohol and other drug use disorder). Such studies might also examine whether the ability of such an SNS-based tool to detect SUD is moderated by number of Instagram posts. In addition, the measure of alcohol consumption used here queried instances of having "had more than a few alcoholic drinks a day" rather than the 4+ or 5+ per day criterion used in the NIDA Quick Screen. The ramifications of this adaptation are unclear.

Summary and Conclusion

Greater knowledge of the association between substance use and SNS participation may inform the development and application of technology-based screening tools. Our findings suggest individuals with at-risk drinking and nonprescription drug use (eg, cannabis, cocaine, heroin, etc) may demonstrate greater participation on Instagram, which could be helpful when developing SNS-based models to detect substance use. On the other hand, the utility of SNS-based models to detect prescription drug use overall, and particularly among middle- and older-age adults, may be more limited, given their lower levels of Instagram participation. As we used a liberal criterion for at-risk drinking and aggregated several drug classes into just two categories, future work might focus on individuals with clinically significant drinking, such as those with alcohol use disorder, and individuals with specific types of drug use (eg, cannabis and opioids). Machine learning technologies that leverage individuals' SNS data to passively screen for substance use may ultimately help reduce the overall burden of SUD and other harmful forms of drinking and drug use. As more is learned about SNS behaviors among those who use substances, researchers may be better positioned to successfully design and interpret these innovative risk detection approaches.

Acknowledgments

This research was supported in part by a National Institute on Drug Abuse grant (P30DA029926) and a pilot grant from the office of the provost at Dartmouth College. BGB's participation was supported by the National Institute on Alcohol Abuse and Alcoholism (K23AA025707).

Authors' Contributions

BGB and WW contributed equally to the study design and interpretation. WW conducted study analyses. BGB wrote an initial manuscript draft, and WW, LAM, and SH helped edit drafts through the final draft. LAM and SH helped interpret findings. SH was the principal investigator on the original study from which the current study data were derived. TCD and BSC edited manuscript drafts and contributed to the original study design and data collection.

Conflicts of Interest

LAM is affiliated with Pear Therapeutics Inc, HealthSim LLC, and Square2 Systems Inc. Conflicts of interest are extensively managed by her academic institution, Dartmouth College. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1

The distribution of number of Instagram posts was similar to a negative binomial distribution, leading us to use negative binomial regression (vs. Poisson regression) when examining associations between demographic characteristics, substance use, and number of Instagram posts.

[PNG File , 20 KB - jmir_v22i9e21916_app1.png]

Multimedia Appendix 2
Sensitivity analysis: zero-truncated negative binomial model examining the associations between demographic characteristics, substance use, and the interaction between demographic characteristics and substance use with the number of Instagram posts.
[DOCX File , 16 KB - jmir_v22i9e21916_app2.docx]

References

1. Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, et al. Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. JAMA Psychiatry 2015 Aug;72(8):757-766 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.0584] [Medline: 26039070]
2. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf [accessed 2020-08-26]
3. Grant BF, Saha TD, Ruan WJ, Goldstein RB, Chou SP, Jung J, et al. Epidemiology of DSM-5 Drug Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions-III. JAMA Psychiatry 2016 Jan;73(1):39-47 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.2132] [Medline: 26580136]
4. Rehm J, Dawson D, Frick U, Gmel G, Roerecke M, Shield KD, et al. Burden of disease associated with alcohol use disorders in the United States. Alcohol Clin Exp Res 2014 Apr;38(4):1068-1077 [FREE Full text] [doi: 10.1111/acer.12331] [Medline: 24428196]
5. Degenhardt L, Whiteford H, Hall WD. The Global Burden of Disease projects: what have we learned about illicit drug use and dependence and their contribution to the global burden of disease? Drug Alcohol Rev 2014 Jan;33(1):4-12. [doi: 10.1111/dar.12088] [Medline: 24251757]
6. US Department of Justice, National Drug Intelligence Center. The economic impact of illicit drug use on American society. 2011. URL: https://www.justice.gov/archive/ndic/pubs44/44731/44731p.pdf [accessed 2020-08-26]
7. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. The Lancet 2009 Jun;373(9682):2223-2233. [doi: 10.1016/S0140-6736(09)60746-7]
8. Park-Lee E, Lipari R, Hedden S, Kroutil L, Porter J. Receipt of services for substance use and mental health issues among adults: Results from the 2016 National Survey on Drug Use and Health.: NSDUH Data Review URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DR-FFR2-2016/NSDUH-DR-FFR2-2016.htm [accessed 2020-08-26]
9. Bray JW, Del Boca FK, McRee BG, Hayashi SW, Babor TF. Screening, Brief Intervention and Referral to Treatment (SBIRT): rationale, program overview and cross-site evaluation. Addiction 2017 Feb;112 Suppl 2:3-11. [doi: 10.1111/add.13676] [Medline: 28074566]
10. Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. Neuropsychopharmacology 2019 Feb;44(3):487-494 [FREE Full text] [doi: 10.1038/s41386-018-0247-x] [Medline: 30356094]
11. Finney JW, Moos RH. Entering treatment for alcohol abuse: a stress and coping model. Addiction 1995 Sep;90(9):1223-1240. [doi: 10.1111/j.1360-0443.1995.tb01092.x]
12. Pew Research Center. Social media fact sheet. URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2020-03-31]
13. Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, et al. Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum. J Med Internet Res 2018 Jun 12;20(6):e10136 [FREE Full text] [doi: 10.2196/10136] [Medline: 29895517]
14. Kornfield R, Toma C, Shah D, Moon T, Gustafson D. What Do You Say Before You Relapse? How Language Use in a Peer-to-peer Online Discussion Forum Predicts Risky Drinking among Those in Recovery. Health Commun. Sep 2018;33(9):1184-1193. [doi: 10.1080/10410236.2017.1350906]
15. Bengio Y. Learning Deep Architectures for AI. FNT in Machine Learning 2009;2(1):1-127. [doi: 10.1561/2200000006]
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
17. Hosmer D, Lemeshow S, Sturdivant R. Assessing the fit of the model. In: Applied Logistic Regression. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013:153-226.
18. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]

19. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a Mobile Health System to Integrate the Treatment of Addiction Into Primary Care: A Hybrid Implementation-Effectiveness Study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
20. Kazemi D, Borsari B, Levine M, Dooley B. Systematic review of surveillance by social media platforms for illicit drug use. J Public Health (Oxf) 2017 Dec 01;39(4):763-776 [FREE Full text] [doi: 10.1093/pubmed/fdx020] [Medline: 28334848]
21. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res 2017 Oct 31;19(10):e353 [FREE Full text] [doi: 10.2196/jmir.6426] [Medline: 29089287]
22. Curtis BL, Lookatch SJ, Ramo DE, McKay JR, Feinn RS, Kranzler HR. Meta-Analysis of the Association of Alcohol-Related Social Media Use with Alcohol Consumption and Alcohol-Related Problems in Adolescents and Young Adults. Alcohol Clin Exp Res 2018 Jun;42(6):978-986 [FREE Full text] [doi: 10.1111/acer.13642] [Medline: 29786874]
23. Saunders J, Aasland O, Babor T, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804 [FREE Full text] [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
24. Bergman BG, Dumas TM, Maxwell-Smith MA, Davis JP. Instagram Participation and Substance Use Among Emerging Adults: The Potential Perils of Peer Belonging. Cyberpsychol Behav Soc Netw 2018 Dec;21(12):753-760 [FREE Full text] [doi: 10.1089/cyber.2018.0109] [Medline: 30499697]
25. Ashford RD, Lynch K, Curtis B. Technology and Social Media Use Among Patients Enrolled in Outpatient Addiction Treatment Programs: Cross-Sectional Survey Study. J Med Internet Res 2018 Mar 06;20(3):e84 [FREE Full text] [doi: 10.2196/jmir.9172] [Medline: 29510968]
26. Bergman BG, Claire Greene M, Hoeppner BB, Kelly JF. Expanding the reach of alcohol and other drug services: Prevalence and correlates of US adult engagement with online technology to address substance problems. Addict Behav 2018 Dec;87:74-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.06.018] [Medline: 29960132]
27. National Institute on Drug Abuse. Resource guide: Screening for drug use in general medical settings.: US Department of Health and Human Services URL: https://d14rmgtrwzf5a.cloudfront.net/sites/default/files/resource_guide.pdf [accessed 2020-04-16]
28. Arnett JJ. The Developmental Context of Substance use in Emerging Adulthood. Journal of Drug Issues 2016 Aug 03;35(2):235-254. [doi: 10.1177/002204260503500202]
29. Zeileis A, Kleiber C, Jackman S. Regression Models for Count Data in R. J Stat Soft 2008;27(8):25. [doi: 10.18637/jss.v027.i08]
30. Ellison NB, Steinfield C, Lampe C. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. J Comput Mediated Commun 2007 Jul;12(4):1143-1168 [FREE Full text]
31. Steinfield C, Ellison NB, Lampe C. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. Journal of Applied Developmental Psychology 2008 Nov;29(6):434-445. [doi: 10.1016/j.appdev.2008.07.002]
32. Seid AK. Social interactions, trust and risky alcohol consumption. Health Econ Rev 2016 Dec;6(1):3 [FREE Full text] [doi: 10.1186/s13561-016-0081-y] [Medline: 26753688]
33. Government of Canada. Cannabis laws and regulations. URL: https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/laws-regulations.html [accessed 2020-04-16]
34. National Conference of State Legislatures. Marijuana overview. URL: http://www.ncsl.org/research/civil-and-criminal-justice/marijuana-overview.aspx [accessed 2020-04-16]
35. Ciccarone D. The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis. Int J Drug Policy 2019 Sep;71:183-188 [FREE Full text] [doi: 10.1016/j.drugpo.2019.01.010] [Medline: 30718120]

Abbreviations

API: application program interface
AUROC: area under the receiver operating characteristic curve
NIDA: National Institute on Drug Abuse
SNS: social network site
SUD: substance use disorder

Edited by G Eysenbach; submitted 28.06.20; peer-reviewed by B Curtis, D Smith; comments to author 14.07.20; revised version received 18.08.20; accepted 18.08.20; published 16.09.20.

Please cite as:
Bergman BG, Wu W, Marsch LA, Crosier BS, DeLise TC, Hassanpour S
Associations Between Substance Use and Instagram Participation to Inform Social Network–Based Screening Models: Multimodal Cross-Sectional Study
J Med Internet Res 2020;22(9):e21916
URL: http://www.jmir.org/2020/9/e21916/
doi:10.2196/21916
PMID:32936081

Original Paper

A New Approach for Detecting Sleep Apnea Using a Contactless Bed Sensor: Comparison Study

Ibrahim Sadek[1,2,3], PhD; Terry Tan Soon Heng[4], MD; Edwin Seet[5], MD; Bessam Abdulrazak[1,2], PhD

[1]AMI-Lab, Computer Science Department, Faculty of Science, University of Sherbrooke, Sherbrooke, QC, Canada

[2]Research Centre on Aging, Sherbrooke, QC, Canada

[3]Biomedical Engineering Dept, Faculty of Engineering, Helwan University, Helwan, Cairo, Egypt

[4]Department of Otolaryngology, Woodlands Health Campus and Khoo Teck Puat Hospital, Singapore, Singapore

[5]Department of Anaesthesia, Khoo Teck Puat Hospital, Singapore, Singapore

Corresponding Author:
Ibrahim Sadek, PhD
AMI-Lab, Computer Science Department, Faculty of Science
University of Sherbrooke
2500 Boulevard de l'Université
Sherbrooke, QC, J1K 2R1
Canada
Phone: 1 819 821 8000 ext 62860
Email: ibrahim.sadek@usherbrooke.ca

Abstract

Background: At present, there is an increased demand for accurate and personalized patient monitoring because of the various challenges facing health care systems. For instance, rising costs and lack of physicians are two serious problems affecting the patient's care. Nonintrusive monitoring of vital signs is a potential solution to close current gaps in patient monitoring. As an example, bed-embedded ballistocardiogram (BCG) sensors can help physicians identify cardiac arrhythmia and obstructive sleep apnea (OSA) nonintrusively without interfering with the patient's everyday activities. Detecting OSA using BCG sensors is gaining popularity among researchers because of its simple installation and accessibility, that is, their nonwearable nature. In the field of nonintrusive vital sign monitoring, a microbend fiber optic sensor (MFOS), among other sensors, has proven to be suitable. Nevertheless, few studies have examined apnea detection.

Objective: This study aims to assess the capabilities of an MFOS for nonintrusive vital signs and sleep apnea detection during an in-lab sleep study. Data were collected from patients with sleep apnea in the sleep laboratory at Khoo Teck Puat Hospital.

Methods: In total, 10 participants underwent full polysomnography (PSG), and the MFOS was placed under the patient's mattress for BCG data collection. The apneic event detection algorithm was evaluated against the manually scored events obtained from the PSG study on a minute-by-minute basis. Furthermore, normalized mean absolute error (NMAE), normalized root mean square error (NRMSE), and mean absolute percentage error (MAPE) were employed to evaluate the sensor capabilities for vital sign detection, comprising heart rate (HR) and respiratory rate (RR). Vital signs were evaluated based on a 30-second time window, with an overlap of 15 seconds. In this study, electrocardiogram and thoracic effort signals were used as references to estimate the performance of the proposed vital sign detection algorithms.

Results: For the 10 patients recruited for the study, the proposed system achieved reasonable results compared with PSG for sleep apnea detection, such as an accuracy of 49.96% (SD 6.39), a sensitivity of 57.07% (SD 12.63), and a specificity of 45.26% (SD 9.51). In addition, the system achieved close results for HR and RR estimation, such as an NMAE of 5.42% (SD 0.57), an NRMSE of 6.54% (SD 0.56), and an MAPE of 5.41% (SD 0.58) for HR, whereas an NMAE of 11.42% (SD 2.62), an NRMSE of 13.85% (SD 2.78), and an MAPE of 11.60% (SD 2.84) for RR.

Conclusions: Overall, the recommended system produced reasonably good results for apneic event detection, considering the fact that we are using a single-channel BCG sensor. Conversely, satisfactory results were obtained for vital sign detection when compared with the PSG outcomes. These results provide preliminary support for the potential use of the MFOS for sleep apnea detection.

(J Med Internet Res 2020;22(9):e18297) doi:10.2196/18297

KEYWORDS

ballistocardiography; sleep apnea; vital signs; eHealth; mobile health; home care

Introduction

Monitoring of Contactless Patients

At present, there are many hurdles confronting health care providers and decision makers, such as the sizable aging patient population, the rising prevalence of chronic diseases, the ever-growing health care spending, and the shortage of clinicians [1,2]. To emphasize, the Association of American Medical Colleges anticipates that the United States could face a shortage of 122,000 physicians by 2032 as the need for physicians outpaces supply [3]. Thus, physicians may not achieve close and continuous monitoring of chronically ill patients on time, thereby increasing their rate of mortality [4]. Apart from ongoing challenges, the existing modalities used to monitor patients at the hospital are too intrusive. They require attaching sensors to the skin or strapping devices to the body. As a result, they will have limited benefits outside hospital rooms. In other words, patients are not monitored before and after being admitted to the hospital.

By comparison, remote and continuous monitoring of patients through contactless sensors can effectively assist physicians in keeping track of their patients' health status while they are at home. More importantly, monitoring and managing patient populations with chronic diseases in a contactless way is essential to avoid additional distress. Contactless monitoring can be achieved largely because of the miniaturization of microprocessors, which allows researchers to integrate sensors into familiar objects, for example, home appliances and mobile devices [5]. Infrared motion sensors, for instance, can capture patients' indoor activities such as being still/moving and moving across rooms. Similarly, contact sensors can capture room, cupboards, and fridge door opening and/or closing. Bed-embedded sensors, also known as ballistocardiogram (BCG) sensors, can deliver noteworthy information about the patient's vital signs, that is, heart rate (HR), breathing, body movements, and quality of sleep [6]. In all, researchers, through contactless sensors, are ultimately trying to predict changes in a patient's health status that can prevent or delay the progression of diseases [7,8]. The hypothesis is that the health status of patients admitted to hospitals is not suddenly deteriorating. Monitoring vital sign trends over time can provide early diagnosis and allow physicians or caregivers to make timely decisions [9]. In this study, we introduce a new approach using a contactless system that is based on the ballistocardiographic principle for detecting abnormal breathing events (ie, apneas and hypopneas) in an effort to address one of today's health care issues.

Sleep Apnea Facts and Diagnoses

The most common form of sleep-disordered breathing is obstructive sleep apnea (OSA). It occurs when a complete or partial closure of the upper airway triggers apnea and hypopnea during sleep [10]. An apnea is a cessation of breathing for at least 10 seconds. Hypopnea is a reduction in airflow for at least 10 seconds by at least 30% accompanied by a drop in oxygen saturation and/or arousal from sleep [11]. Among the general public, OSA affects both men (34%) and women (17%). Nonetheless, it is believed that the prevalence of this syndrome might be underrated. To illustrate, in the United States, estimates showed that 82% of men and 93% of women are underdiagnosed [12].

OSA severity is determined in reference to the apnea-hypopnea index (AHI), that is, the average number of apnea and hypopnea episodes observed per hour of sleep. The severity of OSA is classified as follows: normal (no OSA; AHI <5 events per hour), mild sleep apnea (AHI ≥5 and <15 events per hour), moderate sleep apnea (AHI ≥15 and <30 events per hour), and severe sleep apnea (AHI ≥30 events per hour) [13]. In this regard, patients with moderate or severe apnea are at a higher risk of complications, such as stroke, hypertension, congestive heart failure, and depression. Overall, the late diagnosis of OSA has been shown to double the mortality risk for patients diagnosed with heart failure [14]. The gold standard for evaluating the severity of OSA is polysomnography (PSG). PSG is an overnight controlled sleep study in a specialist sleep laboratory that follows established scoring guidelines for OSA-associated respiratory events.

Through PSG testing, physicians can record different bodily functions. These functions involve HR and rhythms, brain waves, eye movements, leg movements, nasal-oral airflow, thoracoabdominal effort, oxygen saturation, snoring, and body position. The PSG test provides physicians with information about body functions, and therefore, they can diagnose various sleep disorders. However, there are some cons related to the test, for example, high cost, labor intensive, complex, and insufficient privacy. Furthermore, it is not possible to emulate the usual sleep environment in a sleep laboratory. As a consequence, home sleep apnea tests (HSATs) have become alternative possibilities for patients who want to circumvent the in-laboratory PSG. These kinds of tests do not record the full range of signals similar to the PSG. However, they can record up to 7 parameters, including airflow (thermal and nasal pressure), effort (inductive plethysmography), and oximetry [12]. Although such testing is not as reliable as PSG, its portability, affordability, and long-term data collection make it a preferred choice for patients. Recently, off-the-shelf BCG sensors have been investigated by researchers to detect apneic events under the HSAT category. Although the results were encouraging, much work is still needed to reach agreeable results compared with PSG [15]. In this regard, we will discuss later, in brief, the concept of BCG and how it has been employed in the scientific literature to identify apneic events.

Ballistocardiography and Contactless Apnea Detection–Related Work

Ballistocardiography reflects the movement of the center of mass of the body because of cardiovascular activity. The concept of BCG is not new, and there has been a resurgence because of recent improvements in digital electronics reaching the era of microprocessors. Formerly, BCG systems (ie, tables employed by Starr et al [16]) were bulky, heavy, and complicated,

demanding professional mechanical maintenance. Consequently, these systems were principally intended for a single-snapshot recording instead of a long-term data recording [17]. At present, BCG signals are seamlessly being recorded using different sensing modalities, particularly bed-embedded sensors (eg, microbend fiber optic sensors [MFOSs], piezoelectric polyvinylidene sensors, electromechanical film sensors, pneumatic sensors, strain gauges, and hydraulic sensors) [6,18], accelerometers [17,19], and Doppler radar-based sensors [20,21], smart beds [22]. Bed-based sensors, along with accelerometers, can be integrated with everyday objects such as pillows, mattresses, chairs or even installed on the seat of a standard toilet [23]. Moreover, attempts have been made to measure BCG signals via video recording by tracking the motion of facial features [24,25]. Video-based approaches can be practical for surveillance; however, they can impose privacy issues for in-home patient monitoring. Typically, BCG sensors are positioned under the patient's mattress covering the upper half of the body, which allows capturing heart movements, breathing movements, and overall body movements.

Several publications in the literature highlight the extensive use of BCG sensors for both HR and respiratory rate (RR) detection [6,18], which, in turn, attracted researchers to investigate the benefits of BCG signals for more complicated health issues, namely, cardiovascular functions [26] as well as sleep quality [15]. Regarding sleep quality health issues, efforts have been made in the literature to automate the detection of both sleep staging [27,28] and sleep apnea. So far, there have been a few studies that targeted sleep apnea detection through BCG sensors. In this section, we will focus on sleep apnea detection–related work. The study by Sadek et al [6] has more research on HR detection and/or RR detection.

Tenhunen et al [29] investigated the potential of an electromechanical film-based sensor for diagnosing OSA. Although a high sensitivity was reached for detecting apneic events, breathing patterns were analyzed manually by 2 independent scorers, and no contributions were made to computerize the detection process. Hwang et al [30] proposed the use of a polyvinylidene film-based sensor for detecting apneic events. A rule-based framework was implemented to detect apneic events by considering the SD of the sensor signals. Beattie et al [31] tested the effectiveness of using load cells placed under the support of a bed for apnea detection. Although satisfactory results were achieved, the detection process was performed manually by an expert. Waltisberg et al [32] deployed a sensor system that consisted of an array of strain gauges to detect apnea and periodic limb movement events. A supervised learning framework comprising a decision fusion method and a measurement fusion method was applied for the classification process. Similarly, Wang et al [33] used a supervised learning framework to detect apneic events via a micromovement sensitive mattress. Multiple time-domain and frequency-domain features were extracted, which were then fed to different classifiers, that is, k-nearest neighbor, random forest, and support vector machine. Hsu et al [34] sought to detect apneic events by integrating 2 fiber optic–based sensors within a pillow as well as a bed mattress. Apnea detection was achieved by applying 2 methods, that is, a drop degree from the baseline and linear regression models through the percentage of the total duration of the respiratory declination. To compute the model parameters, the empirical mode decomposition (EMD) algorithm was used. However, this signal analysis method is time consuming and precisely for computing the corresponding intrinsic mode functions. Moreover, it is sensitive to the mode-mixing problem [35].

The supervised learning-based approaches (used in the studies mentioned earlier) require a considerable amount of accurately annotated data, which can become quite restrictive for noncontrolled settings [36]. Manual annotation can be considered as an issue because the morphology of BCG signals is highly dependent on the measurement device. BCG signals can differ significantly between studies; besides, they differ within and between subjects. Comparatively, Huysmans et al [37] tested a commercial BCG sensor, that is, Emfit QS, for sleep apnea screening. Unlike the work proposed by Tenhunen et al [29], the authors automated the apnea detection process as follows: 2 Emfit sensors were employed, that is, one sensor was placed below the thorax of the patient and the second was placed under the topper. The detection was then completed via an unsupervised clustering method. The assumption was that during abnormal breathing events, there will always be substantial variations in the signal due to chest motions; thus, by locating these artifacts, they could detect apneic events. This approach avoided the limitation of supervised learning. Nevertheless, the sensor locations were compared to achieve an optimal agreement with PSG. In other words, the sensor that was very close to the thorax achieved more favorable results than the other sensor. In our study, we considered an MFOS for detecting vital signs. Fiber optic sensors (FOSs) are usually used as transducers to detect various environmental changes, such as pressure, temperature, and acceleration [38]. Owing to their electromechanical field immunity and high sensitivity to variations in environmental properties such as the strain, FOSs have been adopted to monitor important physiological parameters, for example, pulse rate and RR, which in turn can help detect cardiovascular diseases and respiratory anomalies [39,40]. Among other sensors, MFOSs have proven to be efficient in detecting ballistic forces correlated with heart movements. They are also moderately small, lightweight, and economical. Hence, they become popular in contactless monitoring of vital signs [41].

The contribution of our study is two-fold. First, we analyzed the robustness of the MFOSs for the simultaneous detection of HR and breathing rate (BR). Second, we examined the capacities of an MFOS for contactless detection of sleep apneic events versus the gold standard overnight in-laboratory PSG.

Methods

Recruitment

This study is approved by the National Healthcare Group Domain-Specific Review Board (NHG DSRB Ref: 2017/01117). A written informed consent form was obtained from all patients before data collection. We completed all the processes, as stated in the guidelines and regulations of the NHG DSRB. We recruited 10 patients diagnosed with OSA and scheduled to

undergo a full night PSG in the sleep laboratory at Khoo Teck Puat Hospital. Patient demographics and related medical history are presented in Table 1. The MFOS was placed under the patient's mattress, and the sensor mat's data were collected in parallel with the overnight PSG data. We imposed no restriction on the exact location of the sensor mat. However, we notified the nurses to locate the mat in the upper part of the bed so that we could acquire cardiac signals as well as respiratory effort signals. The sensor mat does not add any complications to standard PSG protocols because the mat has its own data storage unit. In addition, it does not add any complexity to the patient being monitored. In addition, data analysis was executed offline to align with the ethics approval for the study. To preserve the anonymity of the patients, acquired data were registered with a unique identifier linked to each patient.

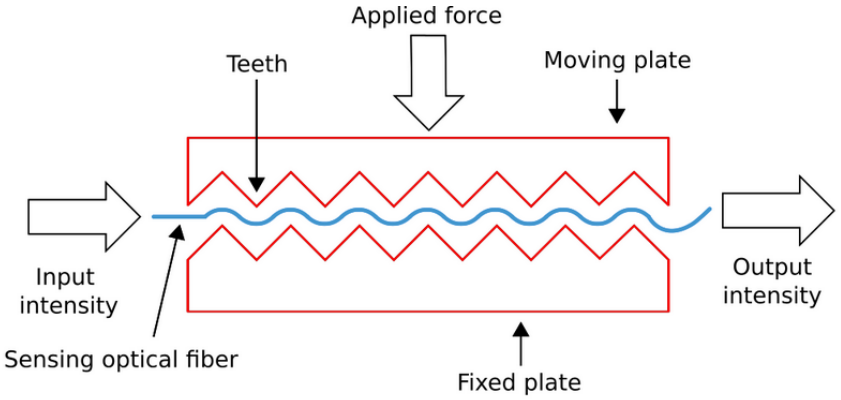
Table 1. Demographics and past medical history of recruited patients.

Patients	Gender	BMI (kg/m^2)	Past medical history	Age (years)	Apnea-hypopnea index	Smoking
1	Female	32.8	Nil	51	36.8	No
2	Male	34	Nil	28	33.7	Yes
3	Male	25.5	Nil	23	32.8	No
4	Male	23	Nil	27	58.3	Unknown
5	Male	25.5	Nil	42	26	Unknown
6	Male	24.8	Nil	33	29	No
7	Male	27.5	Hypertension	49	76.6	Yes
8	Male	33.3	Hypertension and dyslipidemia	43	78.2	No
9	Male	34	Dyslipidemia	61	54.8	No
10	Male	31.2	Nil	29	93.2	No

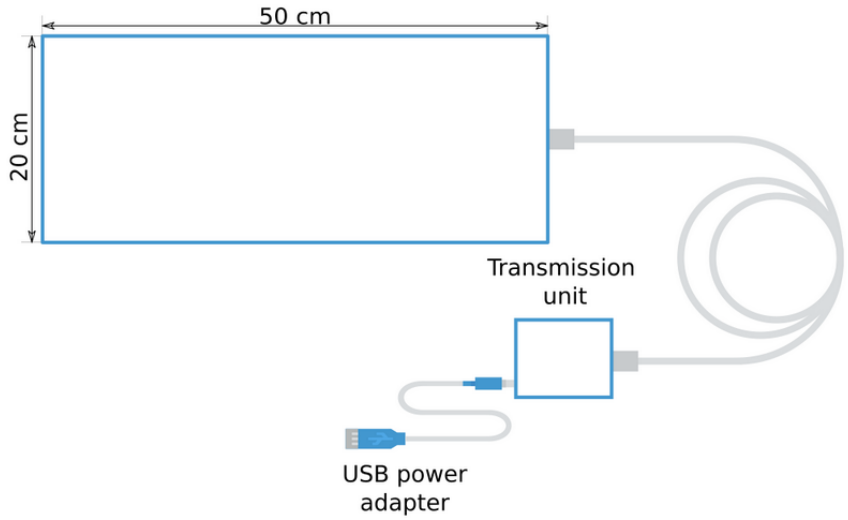
Microbend Fiber Optic Sensor

The proposed monitoring system incorporates a sensor mat and transmission unit. The sensor mat is assembled to a dimension of 20 cm×50 cm×0.5 cm, which promotes its portability and inclusion into cushions, pillows, chairs, and beds. The transmission unit has a built-in microstorage device card for data storage, digital electronics for signal handling, and a Wi-Fi signal transmission module for sending the data to a cloud-based platform. The deployed sensor employs light-intensity modulation caused by the microbending effect in multimode optical fibers, which can be used as a transduction mechanism for detecting pressure. Further information about the sensor's working principle can be found in Multimedia Appendix 1. The system was set to collect data at a sampling frequency of 50 Hz. Figure 1 (top) shows a longitudinal cross-section of the deployed MFOS, and Figure 1 (bottom) presents a schematic diagram of the deployed sensor mat.

Figure 1. Longitudinal cross-section of the microbend fiber optic sensor along with a schematic diagram of the deployed sensor mat.



(a) Longitudinal section of the microbend fiber optic sensor



(b) The deployed sensor mat along with the transmission unit

Data Analysis

For real-time applications, acquired data are deposited in 5-min chunks on a microstorage device card (4 GB internal storage) consolidated with the transmission unit, and then the chunks are dispatched to a cloud server to extract correlated vital signs. Data chunks are encrypted binary files (BIN), and each file consumes 206 KB. Against this background, data are indecipherable without a proper interpreter. In our application, data chunks were gathered directly from the card. Further information about the structure of the data can be found in Multimedia Appendix 1. The data analysis consisted of 2 stages: (1) vital sign detection and (2) apneic event detection. Ensuing, we describe each stage separately.

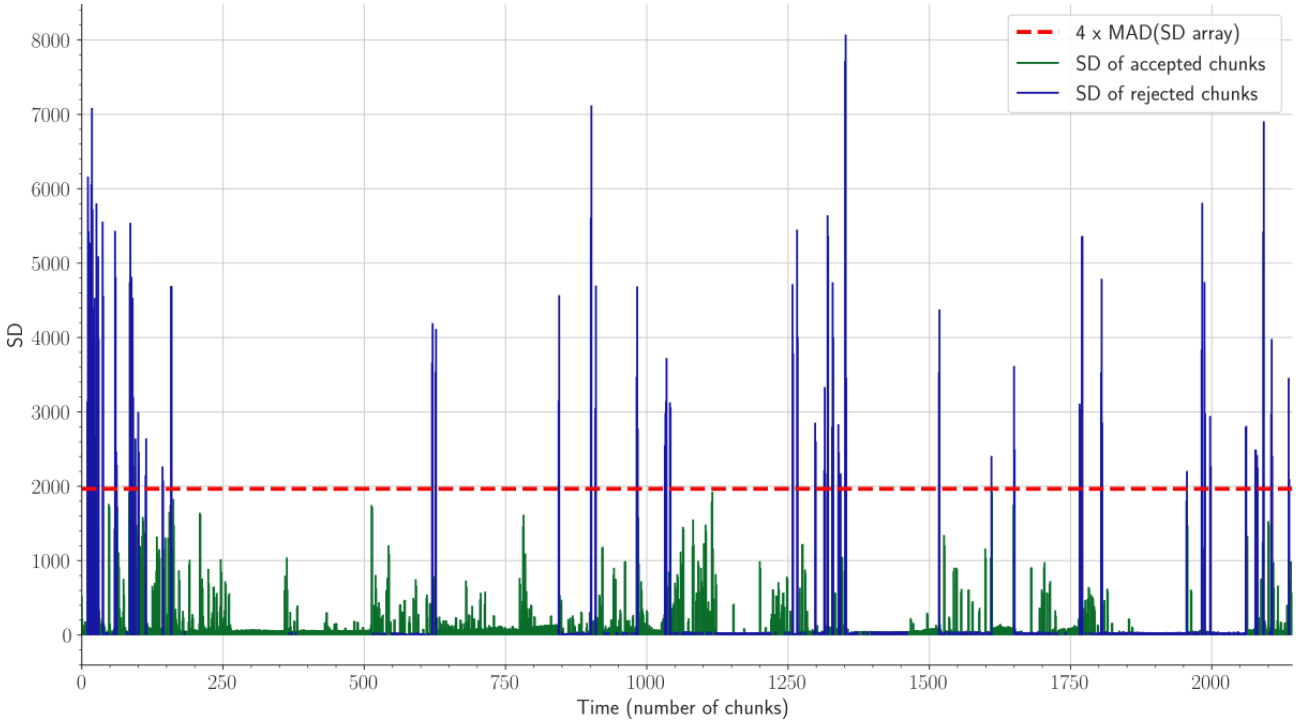
Vital Sign Detection

The force applied to the sensor mat is the summation of the 3 sources. This force is caused by gross body movements and chest wall movement because of the respiration and cardioballistic effect (BCG) [42]. The BCG signal delivers information about HR and HR variability. Similarly, respiratory signals can report on the RR. Extracting both signals can be completed in different ways, for example, via band-pass filtering, wavelet analysis, or other decomposition methods, namely, EMD. To obtain a successful decomposition, motion artifacts must be suppressed from the raw data. Although they are important indicators of sleep quality, information about vital signs cannot be extracted, as the shape of a typical physiological signal is demolished.

Motion artifacts in our approach were suppressed by applying an adaptive threshold method that employed the SD of the raw data [41]. We defined 2 thresholds to remove motion artifacts, that is, *out-of-bed artifact* and *motion artifact*. We divided the raw data stream into equal chunks of 30 seconds, with an overlap of 15 seconds. For each 30-second chunk, we computed the SD and stored all SD values in a single array. Following that, we computed the median absolute deviation (MAD) for the SD array. If the SD of a 30-second chunk was 4 times greater than the MAD, we considered this chunk as a *motion artifact*. In this step, we can control the extent to which the motion artifacts need to be suppressed. When we suppress data chunks with an SD value that is 4 times greater than the MAD, we allow the algorithm to preserve portions of the data with moderately high variation in the signal amplitude (Figure 2). A further increase in this value will allow the algorithm to retain portions of the data with an extremely high variation in the signal amplitude. By selecting this threshold value, we were able to achieve a signal coverage of 79.79%, 81.33%, 78.58%, 84.83%, 86.36%, 87.51%, 81.82%, 51.24%, 75.58%, and 70.59% for all patients, respectively. The coverage is the ratio between the duration of artifact-free signal and raw signal. Ultimately, there should be

a balance between the number of recovered signals and the algorithm performance to measure vital signs of interest. If the SD was lower than a predetermined threshold (5 mV), we considered this chunk as an *out-of-bed* activity. This implies that there were no variations in the amplitude of the acquired data. We only calculated HR and RR for data chunks with SD values between these 2 thresholds. HR and RR were detected according to Sadek et al [41], and further information can be found in Multimedia Appendix 1.

Figure 2. Illustration of isolated motion artifacts. Data chunks were suppressed if they were 4 times greater than the median absolute deviation of the SD array. MAD: median absolute deviation.



HRs were measured using a sliding time window of 30 seconds, with an overlap of 15 seconds. The electrocardiogram (ECG) signal was used as a reference to detect interbeat intervals (Figure 3). To achieve this objective, we selected the well-known Pan and Tompkins algorithm because of its reasonable results [43]. RRs were calculated using a sliding time window of 30 seconds, with an overlap of 15 seconds (Figure 4). The effort signal obtained from the thoracic belt was used as a reference to detect respiratory cycles. Compared with abdominal effort and airflow (ie, pressure and thermistor) signals, the effort thoracic signal was highly correlated with the signal acquired from the MFOS.

Figure 3. The first row displays a 30-second ballistocardiogram signal and the fourth-level smooth coefficient. The second row displays the equivalent electrocardiogram signal. ECG: electrocardiogram.

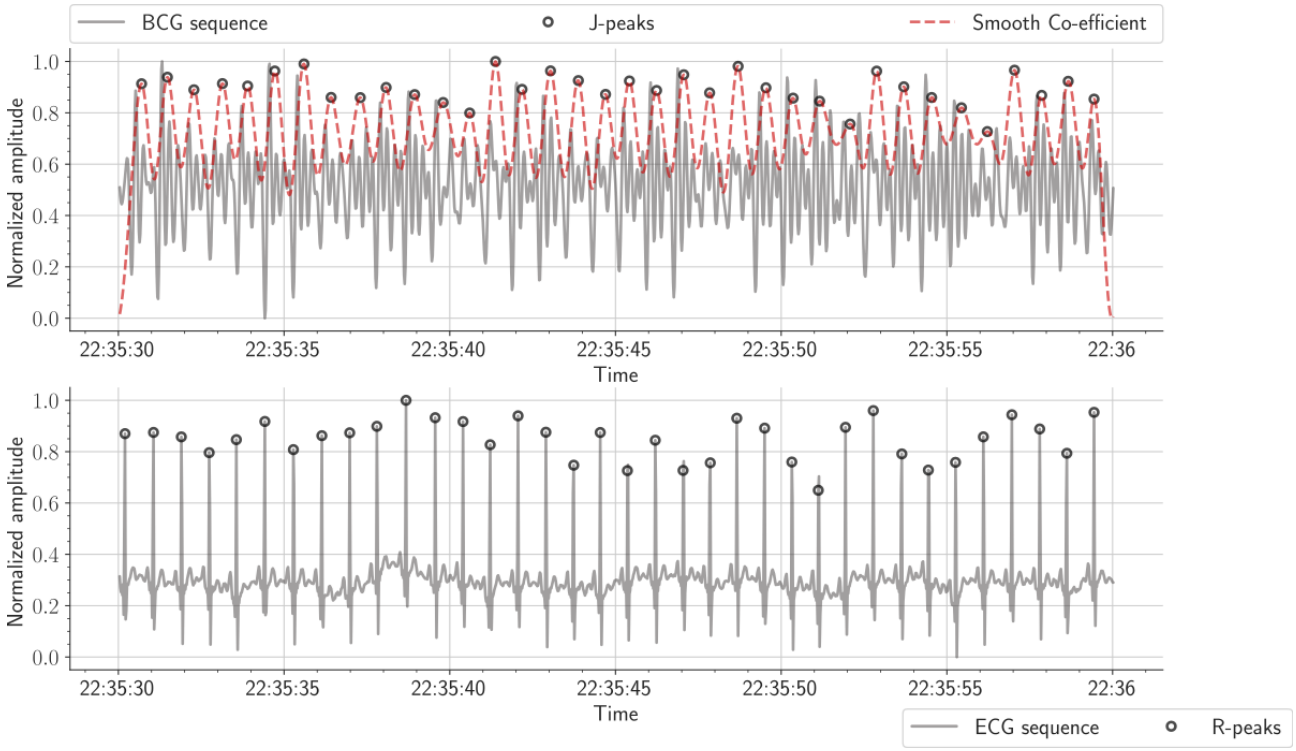
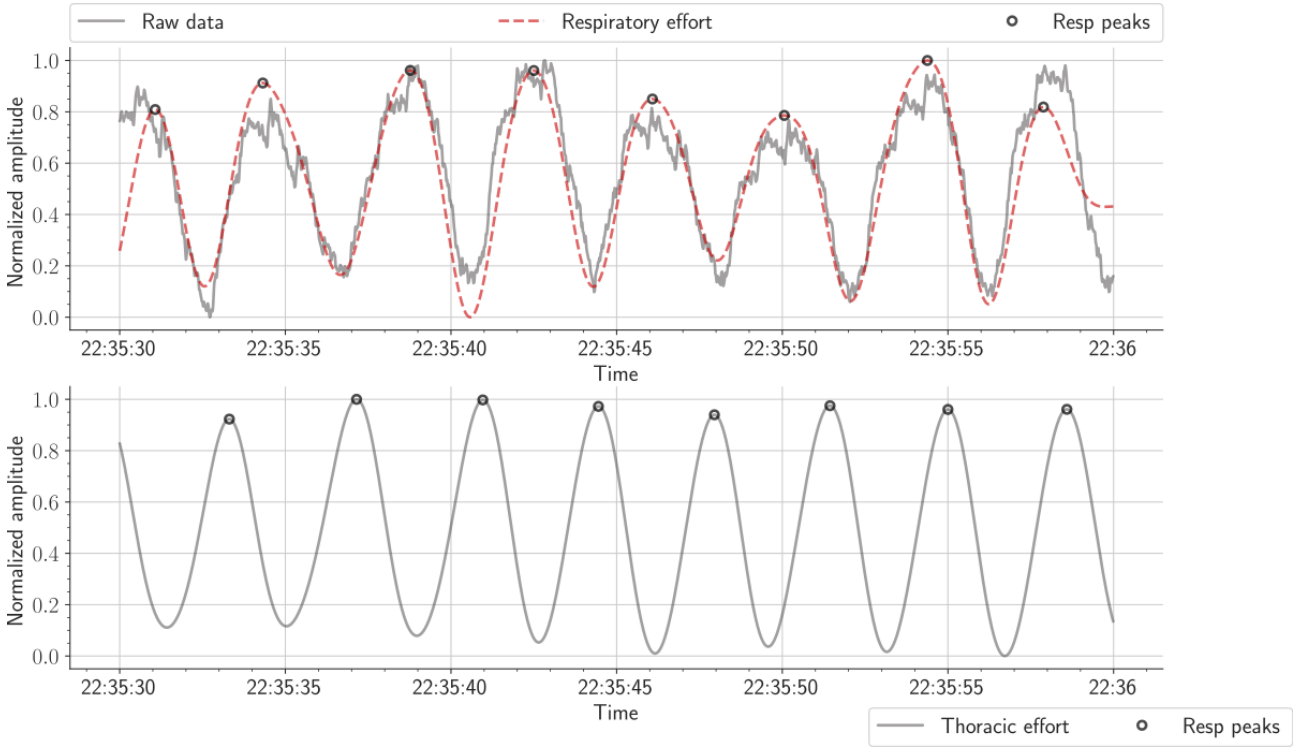


Figure 4. The first row displays a 30-second raw signal and the respiratory effort signal. The second row displays the equivalent thoracic belt signal.



Apneic Event Detection

As we quoted earlier, most of the existing methods use supervised learning algorithms to identify apneic events from BCG signals. Although such methods yield favorable results, they impose restrictions because of the morphological variations of the acquired signals. As such, we implemented an adaptive histogram-based thresholding approach for apnea detection. Pauses in breathing must last at least 10 seconds to be counted as apneic events and can last longer depending on the severity of the disease.

These pauses in breathing are accompanied by an increase in the body and breathing movements and snoring. After matching the scored apneic events (ie, PSG manual scoring) with derived breathing signals, we found that most of the apneic events fell

during *motion artifact*–labeled slices. Thus, motion artifacts were not removed during apnea detection. In our approach, we aimed to differentiate between apneic and nonapneic events via derived breathing signals. To meet this target for each patient, we constructed a histogram from the average absolute deviation (AAD) of the extracted respiratory signal time windows. The time windows were obtained by slicing the signal into equal slices of 30 seconds, with 50% overlapping; afterward, the histogram (ie, the gray bars in Figure 5) was sorted in descending order. In other words, the first histogram value represented the location of the mode of the AAD values (ie, the AAD value that occurred the most often). The hypothesis was that the most frequent histogram values would correspond to normal breathing events. In this regard, we designated the AAD value equivalent to the 6th histogram value as a threshold to detect apneas, that is, AAD values greater than the selected threshold were assumed to represent apneic events (Figure 5). This histogram value was selected (see the Parameters Selection section) based on the proposed method's ability to discriminate between normal and apneic events. This value shows consistent results across all patients. After detecting the threshold, we split the breathing signal into equal slices of 60 seconds, with 50% overlapping. Then, every 60-second slice was further split into three 20-second slices. Next, for each 60-second slice, we computed the AAD of the three 20-second slices and stored them in ascending order. If the difference between the third and second elements was greater than 45% (see the Parameters Selection section) of the histogram threshold, we marked the 60-second slice as an apneic event; otherwise, we labeled it as a nonapneic event (Figure 6).

Figure 5. Histogram of the average absolute deviation values for a breathing signal; the gray bars count the average absolute deviation values that fall into each bin. The selected threshold is represented as a red dashed line. The values between 0 and 400 are only displayed to visualize the histogram bins better. AAD: average absolute deviation.

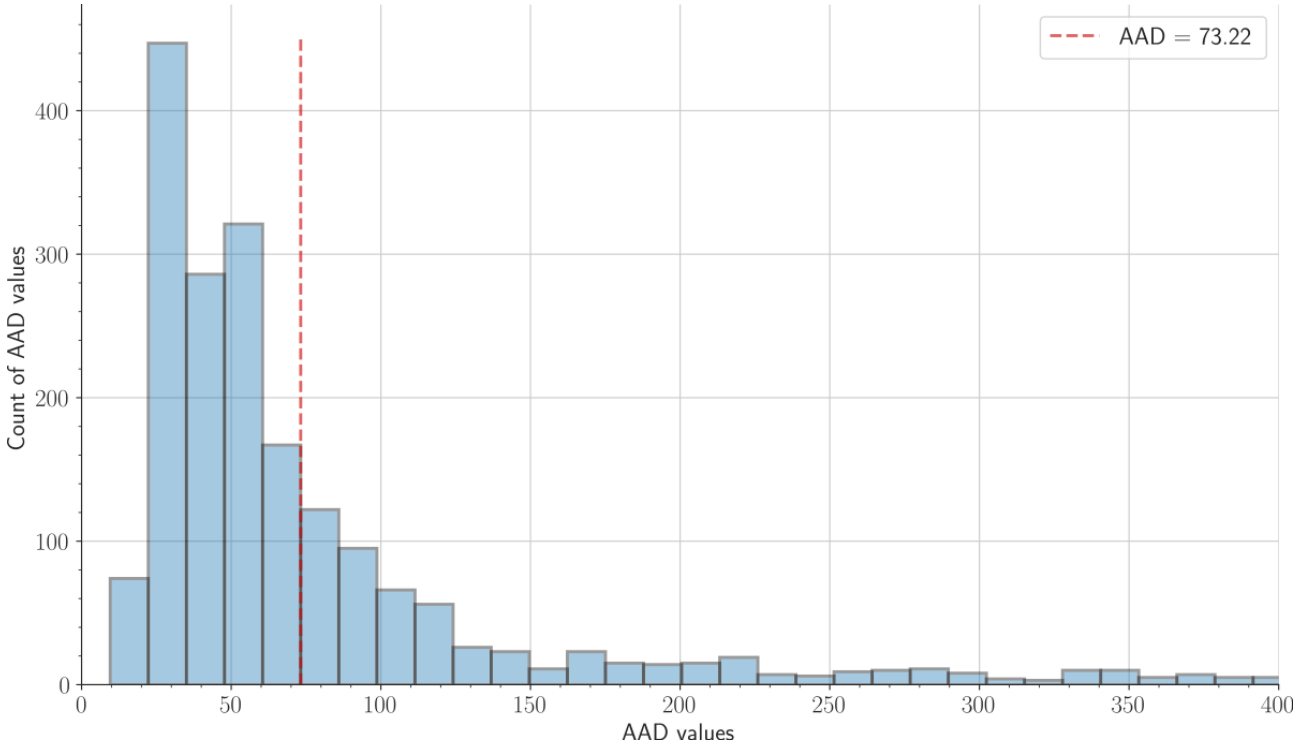
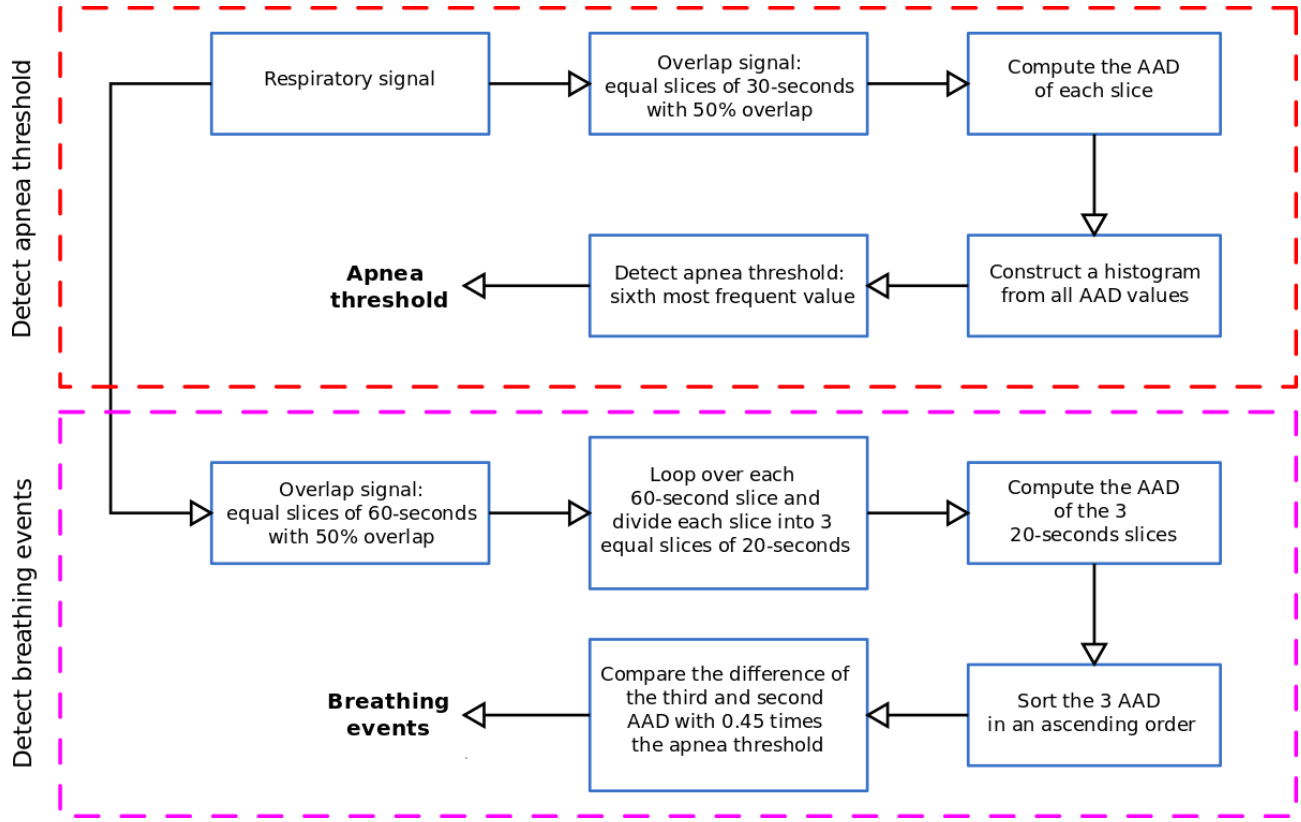


Figure 6. Flowchart of apneic event detection. AAD: average absolute deviation.



Parameter Selection

The proposed method requires the optimization of 2 parameters: a histogram value and a threshold value. A leave-one-out-cross-validation (LOOCV) was implemented to complete this task. The effect of the 2 parameters for apneic event detection was examined in each iteration (ie, 10 iterations in our case) using 9 distinct patient data rather than recording the system's performance for the held-out patient. A system's performance via the LOOCV is usually carried out based on the outcome of a held-out point (ie, a patient in our case). Then, the overall performance is computed by taking the average of the evaluation metrics across all iterations. For our study, we capitalized on this approach to choose the optimal values for the 2 specified parameters. First, we aimed to determine the optimal kth histogram value for an arbitrary threshold value. Various threshold values were exploited (range 0.2-0.95, with a step size of 0.05). Similarly, several histograms (2nd histogram to 16th histogram) were tested against each individual threshold. In other words, for a single threshold value, 3 evaluation metrics (ie, sensitivity, specificity, and accuracy) were measured in accordance with 15 histogram values. This process was repeated 10 times using 9 distinct patient data, and in each iteration, the mean of each metric was recorded. The objective of this process was to find an optimal histogram value applicable to any arbitrary threshold.

For any threshold value, the sensitivity was inversely proportional to the histogram values. However, the specificity and accuracy were directly proportional to the histogram values. Thus, the 6th histogram value can be considered as a critical point. As shown in Figure 7, there was a rapid change in the sensitivity and specificity between the 5th and 7th histograms. The sensitivity keeps decreasing with small fluctuations beyond the 6th histogram value, whereas the opposite occurred for specificity and accuracy. The same behavior occurred for all arbitrary threshold values. As a result, the 6th histogram value was selected as the optimal value for apneic event detection. Second, we aimed to determine the optimal threshold value compatible with the 6th histogram value. To achieve this task, we computed the overall mean area under the curve (AUC), also called balanced accuracy across the 10 iterations for each individual threshold value. The threshold yielding the highest AUC was selected as the optimal value. As shown in Figure 8, the highest AUC occurred at a 45% threshold value, that is, 51.67%. As a result, the 45% threshold value was selected as the optimal value for apneic event detection.

Figure 7. Optimal histogram selection at a 45% threshold value. The 1st row represents the mean sensitivity in each iteration versus different histogram values. The second and third rows represent the specificity and accuracy, respectively.

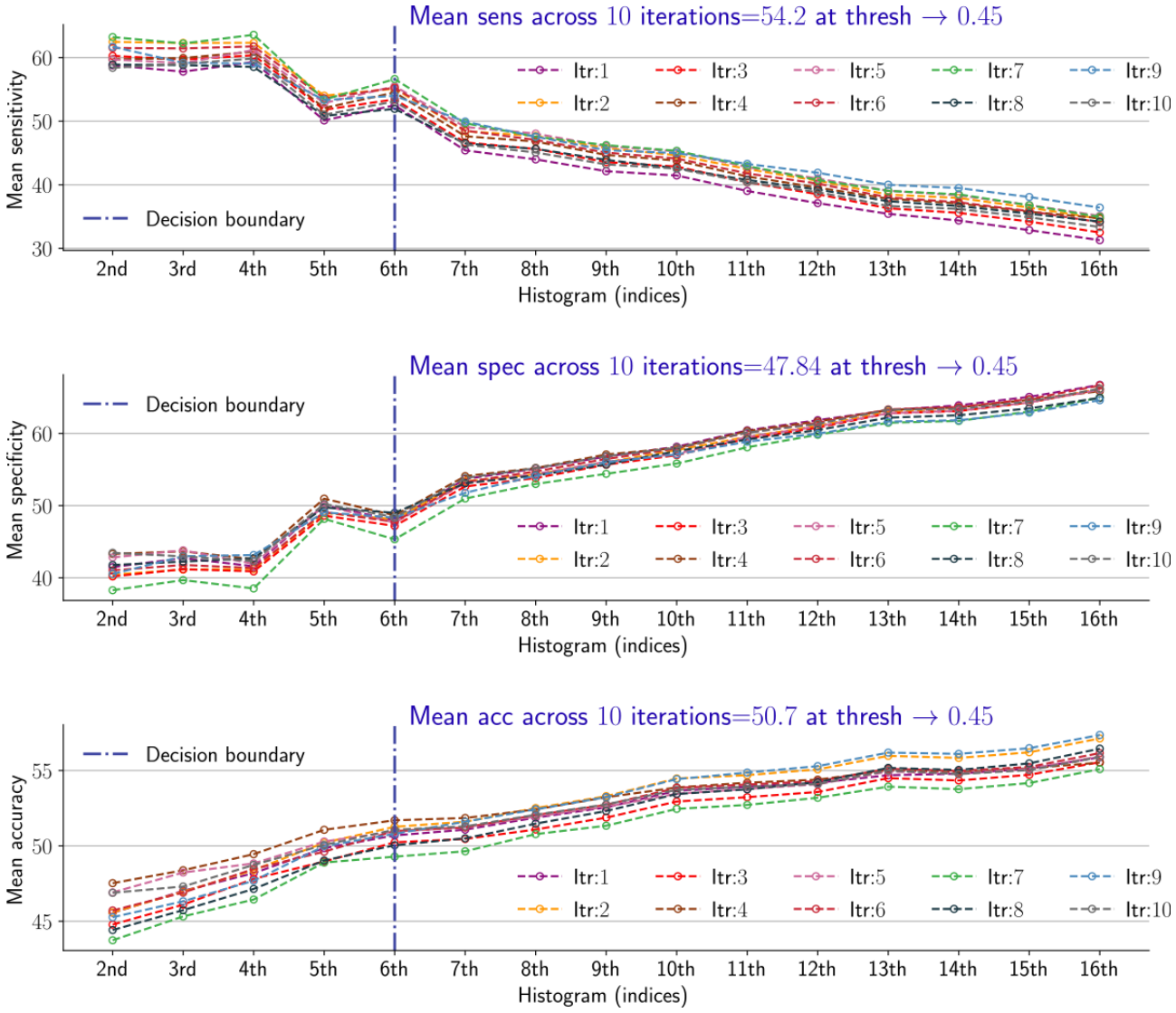
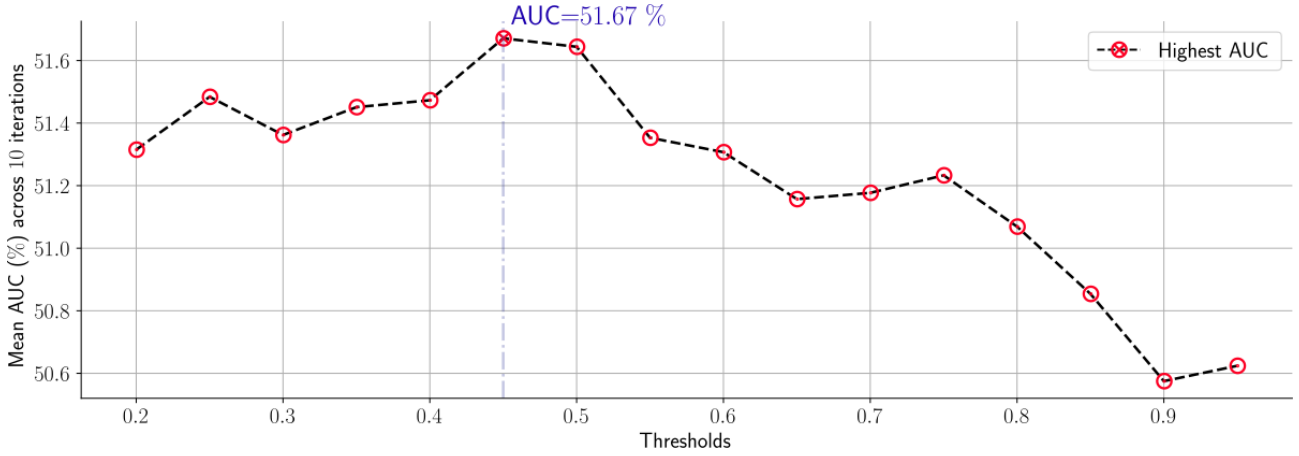


Figure 8. Optimal threshold selection corresponding to the 6th histogram value. The overall mean area under the curve was computed across the 10 iterations and plotted against different threshold values. AUC: area under the curve.



Statistical Analysis

For apneic event detection, we compared the apneic events provided by the PSG with those recovered from the advised sensor mat. Different metrics were adopted in performing the appraisal, that is, sensitivity, specificity, and accuracy. On the other hand, HR and RR were assessed in beats per minute and breaths per minute, respectively. To quantify the performance of the proposed sensor mat for HR and BR estimates compared with the reference ECG signal and effort belt signal, the

Bland-Altman plot, Pearson correlation coefficient, normalized root mean square error (NRMSE), normalized mean absolute error (NMAE), and mean absolute percentage error (MAPE) were adopted. These metrics are commonly employed to determine the difference between medical instruments [44-48]. Further information about these error metrics can be found in Multimedia Appendix 1. High-quality PNG images of all figures presented in the study can be found in Multimedia Appendix 2.

Results

Heart and Respiratory Measurements

On average, the NMAE, NRMSE, and MAPE were 5.42% (SD 0.57), 6.54% (SD 0.56), and 5.41% (SD 0.58) for HR estimation, respectively (Table 2). In addition, the NMAE, NRMSE, and MAPE were 11.42% (SD 2.62), 13.85% (SD 2.78), and 11.60% (SD 2.84), for RR estimation, respectively (Table 3).

Table 2. Normalized mean absolute error, normalized root mean square error, and mean absolute percentage error for heart rate estimation.

Patients	Normalized mean absolute error (%)	Normalized root mean square error (%)	Mean absolute percentage error (%)
1	4.20	5.30	4.16
2	5.74	6.91	5.69
3	5.96	7.15	6.02
4	5.28	6.45	5.27
5	5.08	6.19	5.03
6	5.22	6.33	5.20
7	5.66	6.84	5.73
8	6.26	7.19	6.20
9	5.32	6.39	5.29
10	5.45	6.57	5.40
Mean (SD)	5.42 (0.57)	6.54 (0.56)	5.41 (0.58)

Table 3. Normalized mean absolute error, normalized root mean square error, and mean absolute percentage error for respiratory rate estimation.

Patients	Normalized mean absolute error (%)	Normalized root mean square error (%)	Mean absolute percentage error (%)
1	8.84	11.34	8.80
2	8.43	10.91	8.54
3	12.76	15.04	12.66
4	14.44	16.93	15.00
5	14.33	16.88	14.86
6	14.69	17.64	15.22
7	8.20	10.28	8.14
8	10.22	12.09	10.28
9	9.74	12.20	9.67
10	12.51	15.21	12.84
Mean (SD)	11.42 (2.62)	13.85 (2.78)	11.60 (2.84)

Tables 4 and 5 summarize the limits of agreement (LoA) of the Bland-Altman plot, *r* value, and *P* value for HRs and RRs, respectively. To provide some examples, we provided the Bland-Altman plots and Pearson correlation coefficient plots of the HRs for patients 1, 2, 6, and 10 in Figure 9 (top left, top right, bottom left, and bottom right, respectively) and Figure 10 (top left, top right, bottom left, and bottom right, respectively). In addition, we presented the Bland-Altman plots and the Pearson correlation coefficient plots of the BRs for patients 3, 4, 5, and 9 in Figure 11 (top left, top right, bottom left, and bottom right, respectively) and Figure 12 (top left, top right, bottom left, and bottom right, respectively).

Table 4. Limits of agreement of the Bland-Altman plots, Pearson correlation coefficient, and *P* value for heart rate detection.

Patients	Heart rate				
	Limits of agreement		SD of difference	*r*	*P* value
	Lower	Upper			
1	−3.68	7.89	2.95	0.62	<.001
2	−5.52	9.94	3.94	0.74	<.001
3	−9.46	7.41	4.30	0.68	<.001
4	−7.38	8.06	3.94	0.77	<.001
5	−5.80	8.59	3.67	0.31	<.001
6	−6.85	7.92	3.77	0.63	<.001
7	−8.68	6.44	3.86	0.70	<.001
8	−3.77	11.10	3.80	0.63	<.001
9	−7.13	8.28	3.93	0.64	<.001
10	−5.22	9.53	3.76	0.45	<.001

Table 5. Limits of agreement of the Bland-Altman plots, Pearson correlation coefficient, and *P* value for respiratory rate detection.

Patients	Respiratory rate				
	Limits of agreement		SD of difference	*r*	*P* value
	Lower	Upper			
1	−3.77	3.71	1.91	0.43	<.001
2	−3.87	3.43	1.86	0.58	<.001
3	−3.71	6.40	2.58	0.38	<.001
4	−5.34	5.50	2.77	0.46	<.001
5	−5.79	3.95	2.48	0.46	<.001
6	−5.28	4.95	2.61	0.43	<.001
7	−3.32	4.56	2.01	0.39	<.001
8	−4.22	5.26	2.42	0.42	<.001
9	−4.10	4.32	2.15	0.23	<.001
10	−5.57	4.21	2.49	0.36	<.001

Figure 9. Bland-Altman plots of the heart rates for patients 1, 2, 6, and 10 (ie, top left, top right, bottom left, and bottom right, respectively). ECG: electrocardiogram; HR: heart rate; MFOS: microbend fiber optic sensor; RR: respiratory rate.

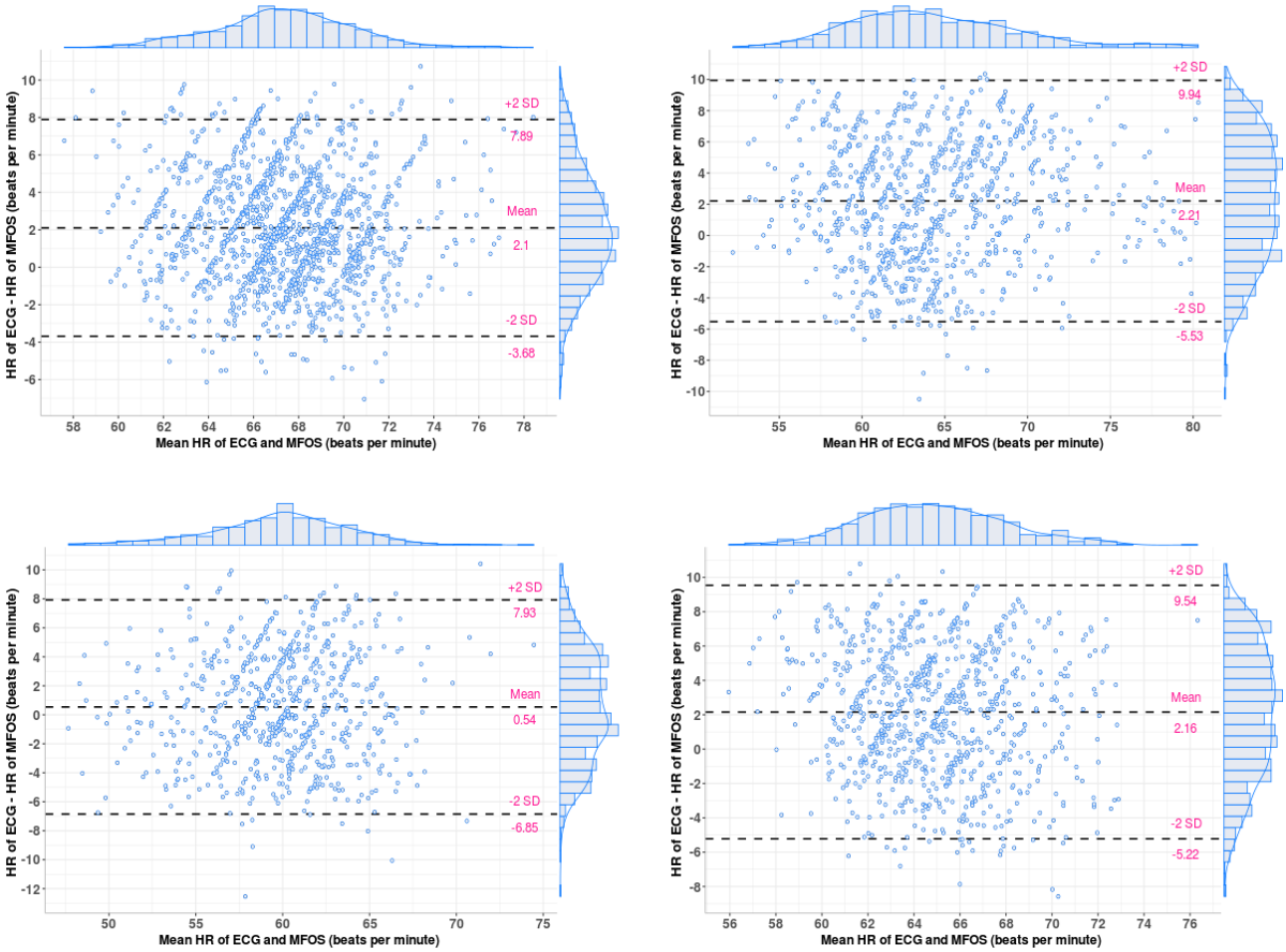


Figure 10. Pearson correlation plots of the heart rates for patients 1, 2, 6, and 10 (ie, top left, top right, bottom left, and bottom right, respectively). The blue circles represent reference heart rate against the estimated heart rate, and the blue line represents the fitted line. ECG: electrocardiogram; MFOS: microbend fiber optic sensor.

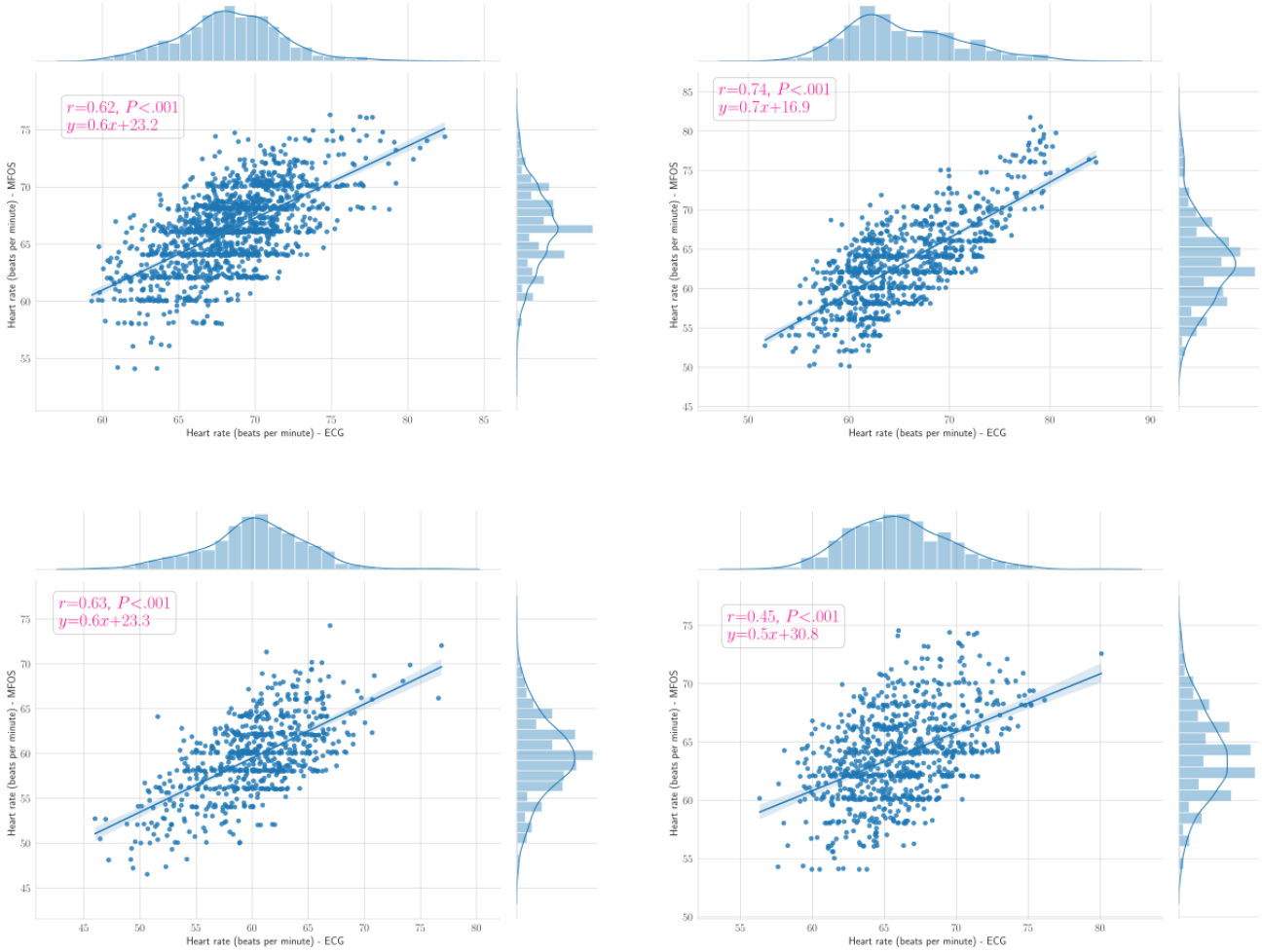


Figure 11. Bland-Altman plots of the respiratory rates for patients 3, 4, 5, and 9 (ie, top left, top right, bottom left, and bottom right, respectively). ECG: electrocardiogram; HR: heart rate; MFOS: microbend fiber optic sensor; RR: respiratory rate; THO: thoracic belt.

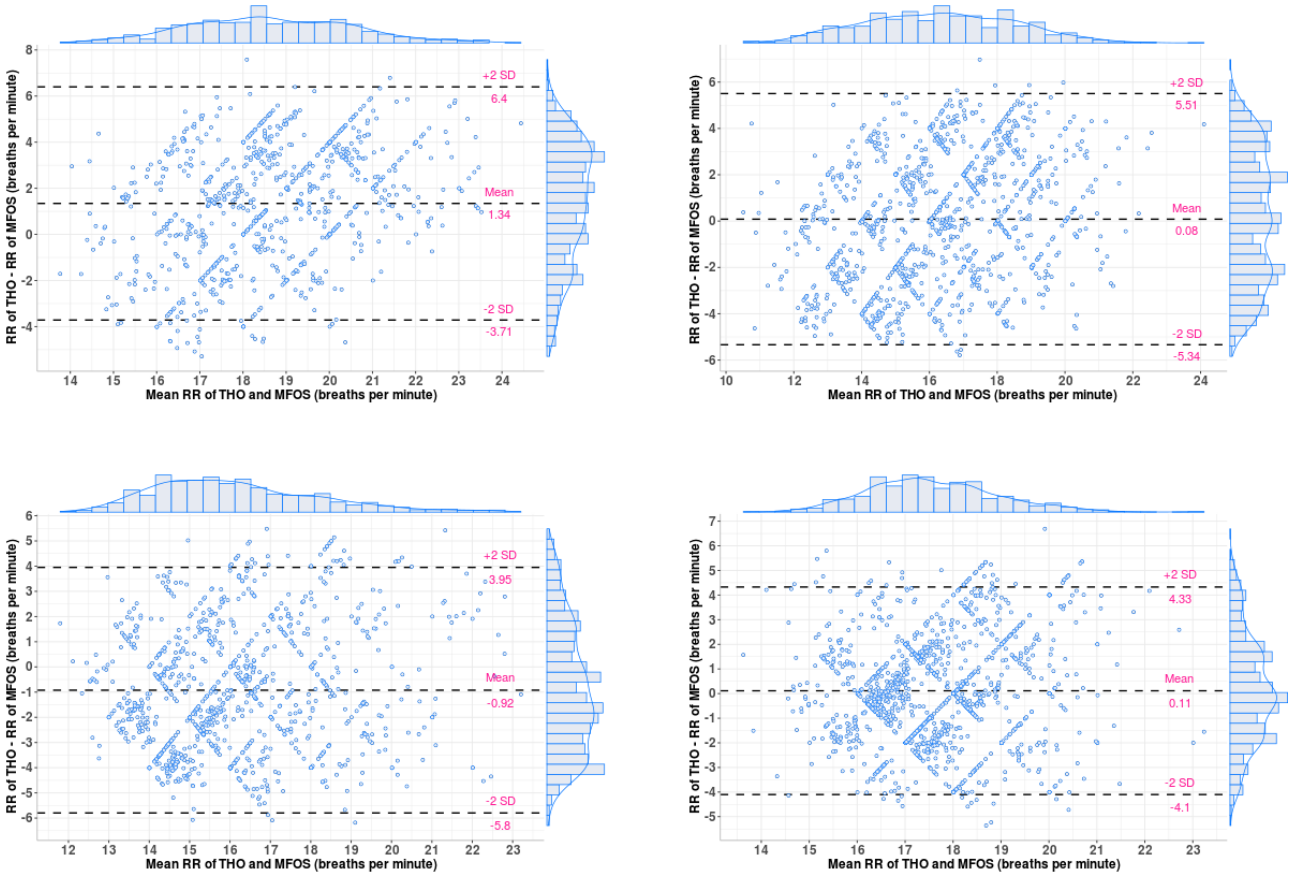
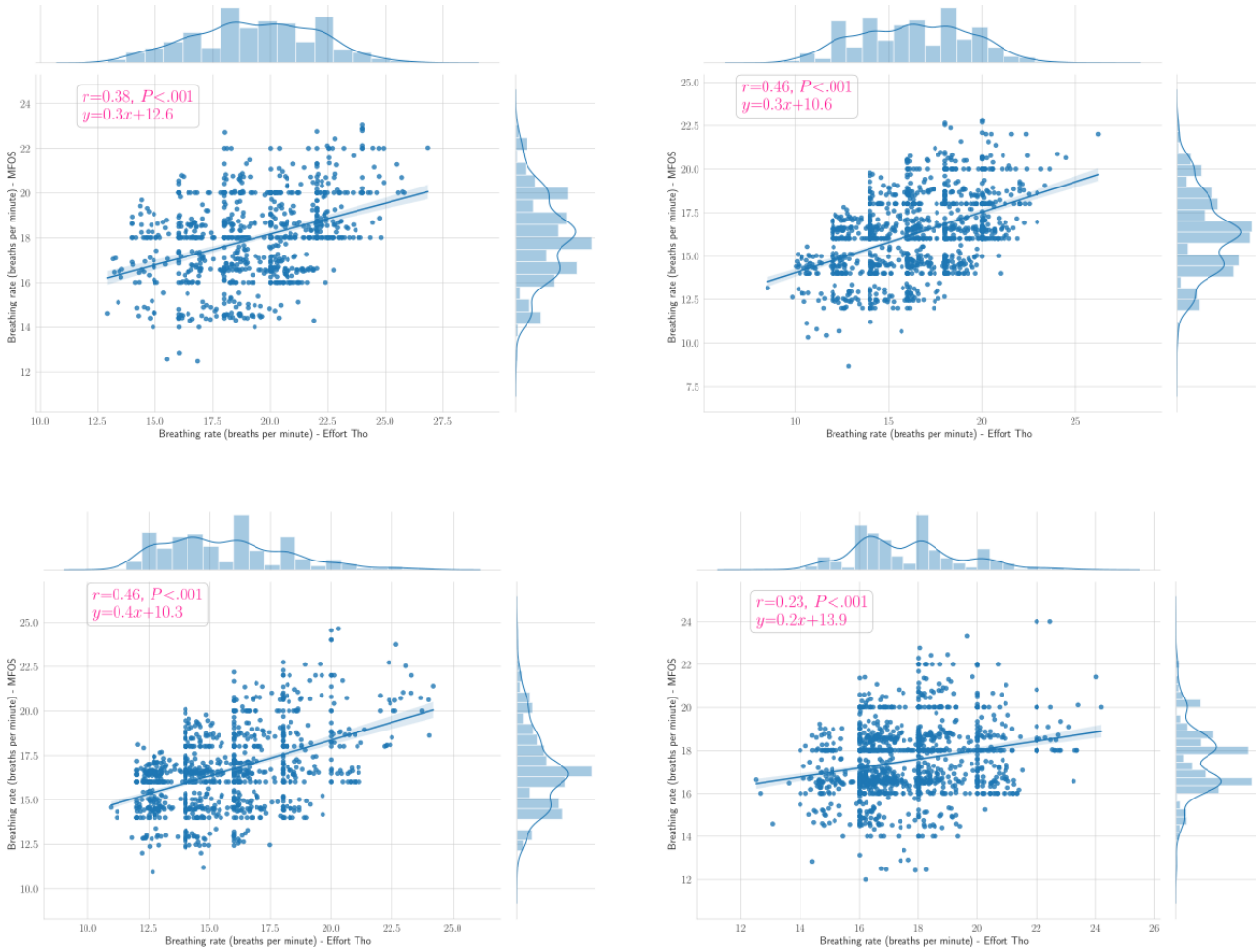


Figure 12. Pearson correlation plots of the respiratory rates for patients 3, 4, 5, and 9 (ie, top left, top right, bottom left, and bottom right, respectively). The blue circles represent reference RR against the estimated respiratory rate, and the blue line represents the fitted line. MFOS: microbend fiber optic sensor.



Apneic Event Detection

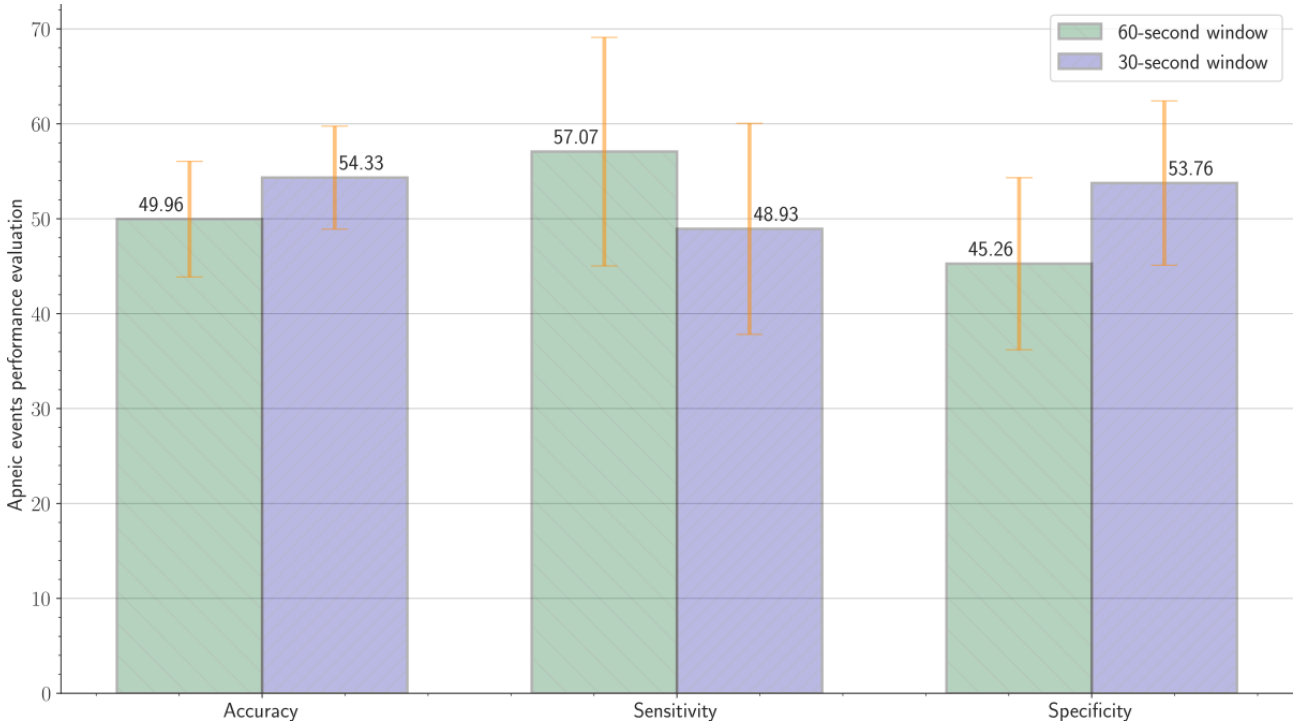
The apneic events were detected from the derived respiratory signals via windowing with overlapping. The assessment was made against the manually scored apneic events. Each sliding window was classified as either an apneic breathing event or a normal breathing event. Apneic events consisted of obstructive apneas, hypopnea, central apneas, and mixed apneas. For detection, we tested a sliding time window of 2 different sizes, that is, a sliding time window of 60 seconds with an overlap of 30 seconds as well as a sliding time window of 30 seconds with an overlap of 15 seconds. For the former, if any 20-second slice satisfied the apneic threshold condition, we considered the entire 60-second window as an apneic event. Similarly, for the latter, if any 10-second slice satisfied the apneic threshold condition, we considered the complete 30-second window as an apneic condition.

As presented in Table 6, for the 60-second time window, on average, the proposed system achieved an accuracy of 49.96% (SD 6.39), a sensitivity of 57.07% (SD 12.63), and specificity of 45.26% (SD 9.51). In addition, for the 30-second time window, on average, the proposed system achieved an accuracy of 54.33% (SD 5.72), a sensitivity of 48.93% (SD 11.72), and a specificity of 53.76% (SD 9.12). Figure 13 displays bar charts with error bars for the reported accuracy, sensitivity, and specificity related to apneic events' detection of the 60-second and 30-second time windows.

Table 6. Accuracy, sensitivity, specificity, and *P* value of apneic event detection; the two-tailed test was used to determine the *P* value.

Characteristics	Patients									
	1	2	3	4	5	6	7	8	9	10
Sliding time window of 60 seconds										
Accuracy (%)	49.95	53.93	39.26	46.29	46.53	63.2	57.01	50.8	46.56	46.02
Sensitivity (%)	71.57	62.02	51.94	45.6	46.55	34.72	78.48	62.4	65.82	51.61
Specificity (%)	45.24	52.26	37.56	46.42	46.52	69.54	33.9	38.86	41.28	41
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.03	<.001	<.001
Sliding time window of 30 seconds										
Accuracy (%)	55.18	58.05	47.19	51.05	56.35	66.82	56.88	53.13	49.24	49.44
Sensitivity (%)	56.64	49.99	46.3	40.98	32.19	31.39	69.44	56.77	54.39	51.24
Specificity (%)	54.86	59.7	47.31	52.94	59.62	74.71	43.38	49.39	47.83	47.83
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.12

Figure 13. Bar charts with error bars for accuracy, sensitivity, and specificity regarding apneic events detection (60-second and 30-second time windows). The bars represent the mean of each measure, and the cap-tipped lines represent the uncertainty (SD) in each measure.



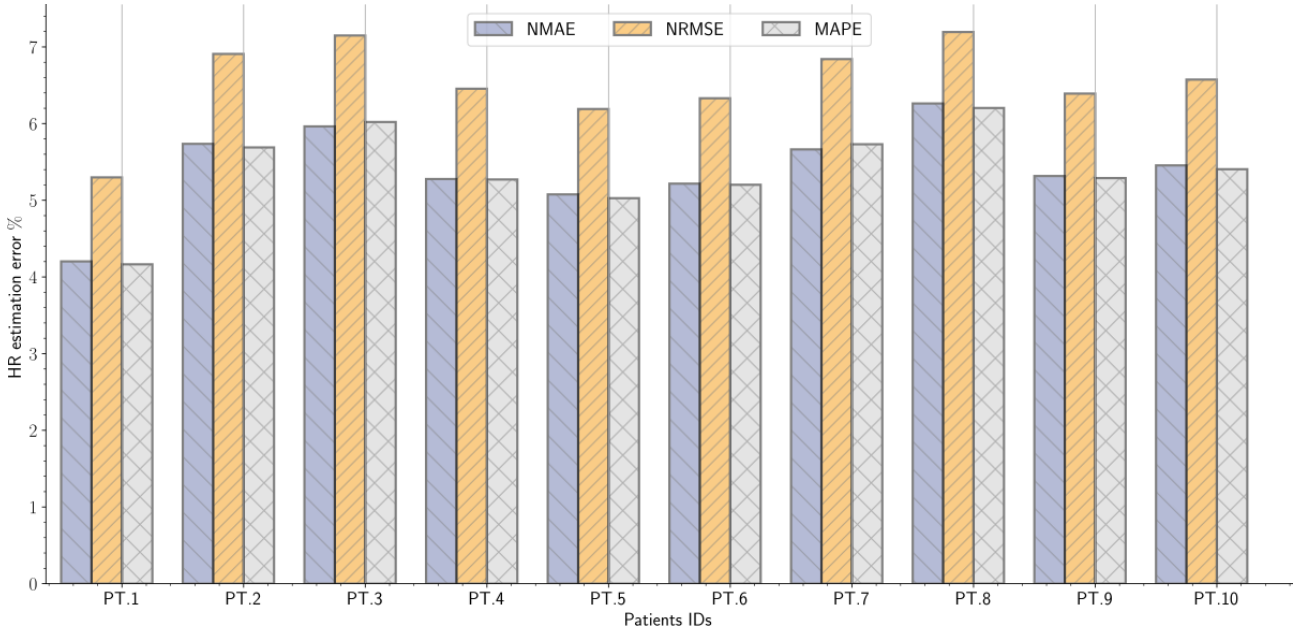
Discussion

Principal Findings

In this study, we aimed to estimate the potential of using a single-channel monitoring device (ie, a bed-embedded FOS) for contactless monitoring of vital signs (ie, HRs and RRs) and apneic breathing events during an overnight sleep study. For HR estimation, the devised method achieved reasonably accurate results compared with the reference ECG signals. For the first patient, the system achieved the lowest NMAE, NRMSE, and MAPE, such as 4.20%, 5.30%, and 4.16%, respectively, whereas the highest NMAE, NRMSE, and MAPE were 6.26%, 7.19%, and 6.20%, respectively, for the eighth patient (Figure 14). The signal coverage for patients with severe OSA (eg, patients 8 and 10) was small compared with patients with moderate OSA (eg, patients 5 and 6). The signal coverage was the lowest for the eighth patient (ie, 51.24%), and the error in beats per minute was the highest among other patients. It is not necessarily true that patients with higher signal coverage will have the lowest error; however, the signal quality is the main factor affecting the outcomes. For instance, the first patient did not have the highest signal coverage (79.79%) but had the lowest error. This situation occurred because this patient had a small number of apneas (ie, 14) and a large number of hypopneas (ie, 191). The fifth patient had the highest coverage (87.51%); however, the error was slightly larger compared with the first patient. This situation occurred because this patient had a large number of apneas (ie, 203) and a small number of hypopneas (ie, 20). Overall, it may be said that the error in beats per minute is likely to increase for patients with a large number of apneas. This is

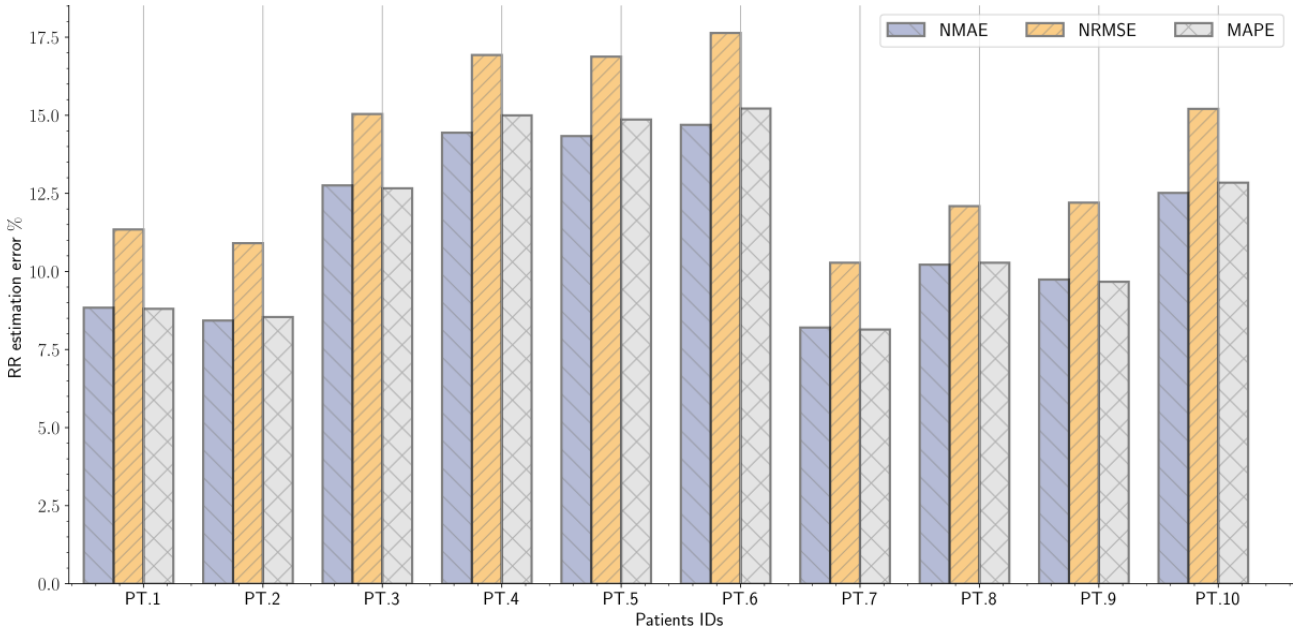
because the amount of motion artifacts progresses for patients with severe apnea. In addition, the morphology of the BCG is significantly affected by cardiovascular complications of sleep apnea. It should also be recalled that the designated threshold value to eliminate motion artifacts had an effect on the estimation process. To explain, in our method, we rejected time windows that had an SD value 4 times greater than the MAD. Decreasing the threshold value would have allowed us to reject a large number of motion artifacts and consequently obtain lower errors. Nonetheless, the signal coverage could have been much lower. As a result, we balanced between achieving reasonable errors and retaining reasonable signal coverage. HR results were also supported by the Pearson correlation coefficients and LoA of the Bland-Altman plot. The system achieved the highest correlation coefficient for the fourth patient (r=0.77; P<.001) and the lowest correlation for the fifth patient (r=0.31; P<.001).

Figure 14. Bar plots of the normalized mean absolute error, normalized root mean square error, and mean absolute percentage error between reference device (electrocardiogram) and the microbend fiber optic sensor for heart rate estimation. HR: heart rate; MAPE: mean absolute percentage error; NMAE: normalized mean absolute error; NRMSE: normalized root mean square error.



RR findings, on the other hand, were slightly inferior to HR results. By way of illustration, the lowest NMAE, NRMSE, and MAPE were 8.20%, 10.28%, and 8.14%, respectively, for the seventh patient, whereas the highest NMAE, NRMSE, and MAPE were 14.69%, 17.64%, and 15.22%, respectively, for the sixth patient (Figure 15). In general, detecting RRs in healthy subjects is simpler than detecting HRs. This is because respiratory cycles, that is, inhalation and exhalation, can be located through a peak detector. However, the situation is more challenging for patients with sleep apnea for different reasons. To illustrate, in our approach, RRs represent the movement of the chest and abdominal wall; however, because of the recurrent decrease and increase in breathing effort, detecting respiratory cycles has become a challenging task. These variations in breathing efforts affected the accuracy of the peak detector and consequently contributed to increasing the error between the devised sensor and the reference thoracic belt. Compared with HR detection, the lowest correlation coefficient was (r=0.23; P<.001) for the ninth patient, whereas the highest correlation was (r=0.58; P<.001) for the second patient. In general, patients with very severe OSA (ie, patients 7, 8, and 10) presented slightly worse correlation coefficients than patients with less severe OSA (ie, patients 1, 2, 4, 5, and 6). The value of the correlation coefficient depended to no small extent on the types of apneas and also the duration of apneic events presented in each patient. These parameters significantly influenced the respiratory signal's typical shape, and thus, the respiratory cycles were more difficult to detect.

Figure 15. Bar plots of the normalized mean absolute error, normalized root mean square error, and mean absolute percentage error between reference device (thoracic belt) and the microbend fiber optic sensor for respiratory rate estimation. MAPE: mean absolute percentage error; NMAE: normalized mean absolute error; NRMSE: normalized root mean square error; RR: respiratory rate.



This highest error occurred in the sixth patient, most likely because of poor contact between the sensor mat and patient. The structure of both BCG and respiratory signals was highly different compared with that of other patients. Furthermore, the amplitude of the acquired raw data was very low. These issues contributed to a large discrepancy between the true peaks and detected peaks, as presented in Figure 16. Sleep apnea detection, from a different angle, demands multiple sensors and wires fixed to the patient's body throughout one night, including, for example, airflow, respiratory effort, and oximetry; notwithstanding, in this study, we only employed a single-channel BCG sensor. The conceived sensor delivered reasonably good results considering the fact that we were using a single-channel BCG sensor. The detection evaluation metrics (ie, accuracy, sensitivity, and specificity) were measured according to the overlapping between manually scored events and events obtained by the deployed sensor mat. Across all recruited patients, the 60-second sliding moving window slightly outperformed the 30-second moving window. As shown in Table 6, the average sensitivity of the former window was 57.07% (SD 12.63) compared with 48.93% (SD 11.72) for the second window. Although the accuracy and specificity of the 30-second window were slightly better than those of the 60-second window, the *P* value of the last patient was .12. In contrast, the 60-second window reached a *P* value of <.001 for all patients but the eighth patient (*P*=.03). In addition, the lowest sensitivity for the 30-second window was 31.39%; however, it was 34.72% for the 60-second window. Both time windows achieved the highest sensitivity for the seventh patient, such as 78.48% and 69.44% in a row. Figure 17 presents an example of an apneic event annotation related to the first patient throughout the night.

Figure 16. The first row displays a 30-second raw signal plus the respiratory effort signal for patient 6. The second row shows the equivalent thoracic belt signal.

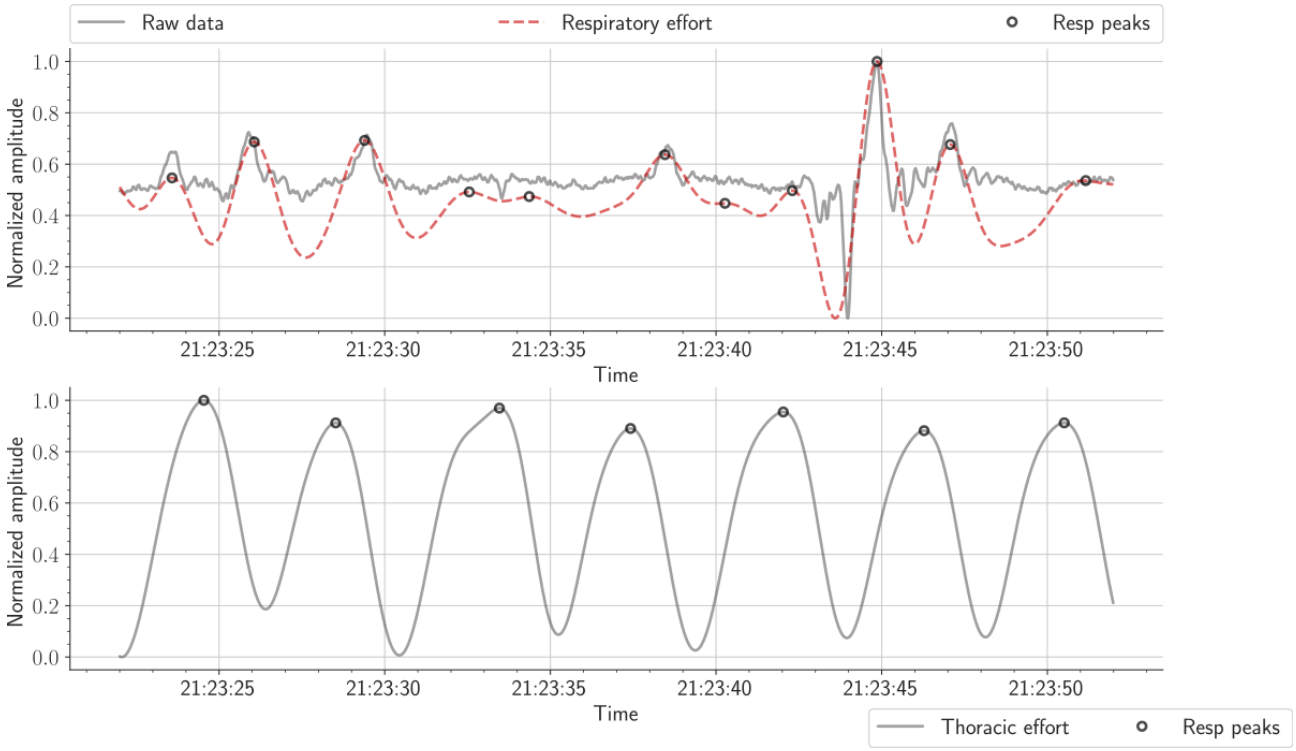
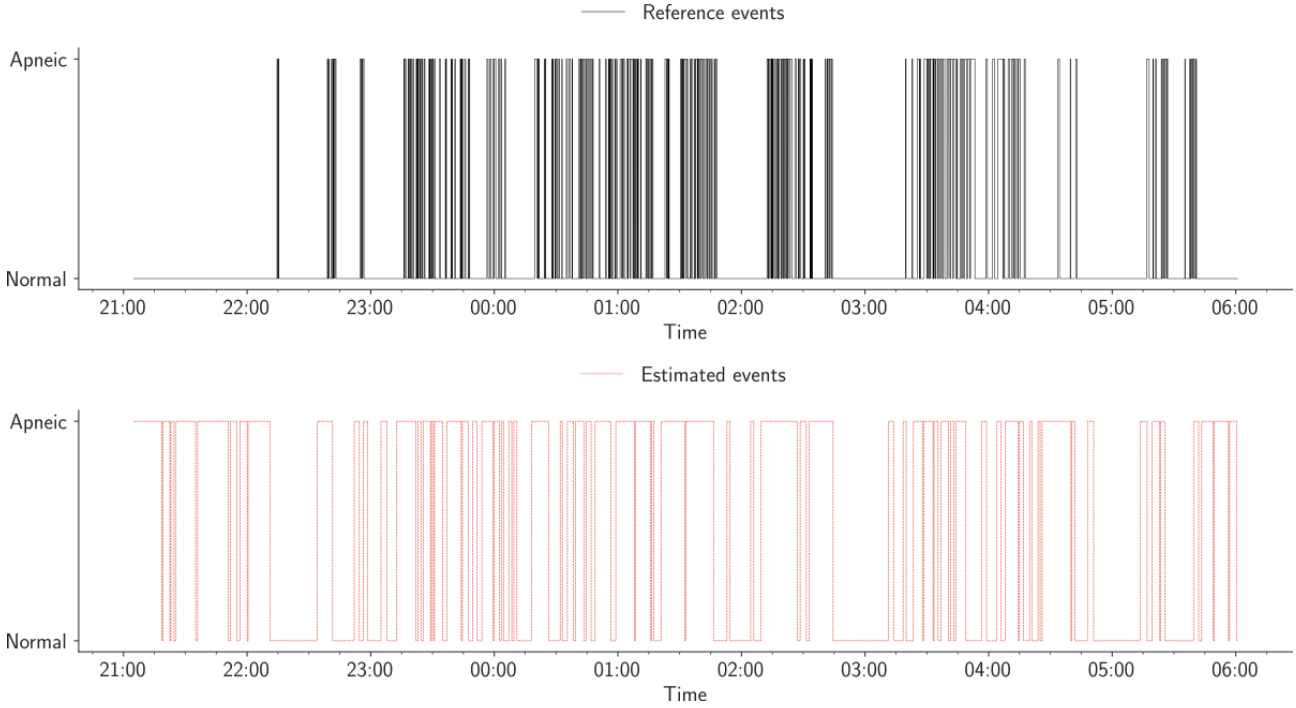


Figure 17. Annotation of apneic events for the first patient. The first row describes the reference events (apnea-hypopnea index is 36.8 events per hour), whereas the second row describes the estimated events (apnea-hypopnea index is 23 events per hour).



Owing to the reasons mentioned earlier, the 60-second window was selected for apneic event detection. As discussed in the Parameter Selection section, a threshold of 0.45 was selected because it contributed to a balanced result between sensitivity and specificity. Selecting a smaller threshold value (eg, 0.2) would have resulted in a very high sensitivity across all patients; however, the opposite would have happened for both specificity and accuracy. Regarding the 0.45 threshold, the sensitivity tended to increase exponentially for patients with less severe OSA (ie, patient 5: AHI=26; patient 3: AHI=32.8; patient 2: AHI=33.7; and patient 1: AHI=36.8) in the order of 46.55%, 51.94%, 62.02%, and 71.57%, respectively. Nonetheless, the sixth patient (AHI=29 and sensitivity=34.72%) did not follow this order because of the presence of central apnea events (ie, 13 events). These events were more challenging to detect. On the other hand, the sensitivity tended to decrease exponentially

in patients with very severe OSA (ie, patient 7: AHI=76.6%; patient 8: AHI=78.2; and patient 10: AHI=93.2) in the order of 78.48%, 62.40%, and 51.61%, respectively. This particular behavior was because of the apnea/hypopnea ratio. In other words, patients with a large number of hypopneas tended to have higher sensitivity when compared with other patients (Table 7). The designated threshold undoubtedly contributed to this outcome, and we could say that the proposed system was more appropriate for patients with less severe OSA. Such a configuration can be helpful in detecting OSA early and avoiding further complications. Table 7 presents the number of apneic events (obstructive apneas, hypopneas, central apneas, and mixed apneas) for each patient versus the proposed system's sensitivity and the manually scored AHI.

Table 7. The counts of the different apneic events for each patient versus achieved sensitivity and the manually scored apnea-hypopnea index.

Patients	Obstructive apneas (n)	Hypopneas (n)	Central apneas	Mixed apneas	Sensitivity	Apnea-hypopnea index
1	12	189	0	0	71.57	36.8
2	23	168	13	0	62.02	33.7
3	11	92	0	0	51.94	32.8
4	201	18	0	0	45.6	58.3
5	47	50	0	0	46.55	26
6	90	64	13	0	34.72	29
7	499	47	1	9	78.48	76.6
8	394	21	3	34	62.4	78.2
9	132	127	1	0	65.82	54.8
10	577	24	2	0	51.61	93.2

Limitations

The limitation of this study is the small sample size; despite that, our ultimate goal was to quantify the predictive outcomes of the fiber optic mat for vital signs and sleep apnea detection in a real-life sleep study. For HR and RR, the findings of the study have shown that the proposed system can provide results close to those of reference devices used in the PSG study. For sleep apnea detection, the designed system provided favorable results for patients with less severe OSA compared with patients with very severe OSA. This issue can be investigated in the future by adding another sensor, for example, an accelerometer, as a noise reference to eliminate body movements [49]. It should be pointed out that the suggested method for apnea detection did not follow the supervised learning models, and hence, we avoided labeling sensor data. The manually scored apneic events could have been used as a guide to label sensor data; however, the labeling process will be a restricted property, given large-scale deployment at users' homes. Another issue to consider is data availability; BCG signals are not benchmarked; as a result, a training model can only be limited to specific sensor data. This problem occurs because the outcome of BCG sensors is not necessarily similar, which restricts testing across different data sets. As stated by Inan et al [18] in a recent review article, there should be a comprehensive and open database of BCG signals. Such databases will allow researchers to employ them in their environments and improve the field into an accepted technique appropriate for clinical studies [18].

Comparison With Prior Work

We attempted to detect apneic events in a previous study [41], where the trial was performed during a drug-induced sleep endoscopy, and the optical fiber mat was compared with the ApneaLink device (ResMed). In a previous clinical study, the system delivered very low sensitivity because of the short evaluation period, that is, around 120 min per study. In addition, the ApneaLink device is not as accurate as the gold-standard PSG. Moreover, the employed algorithm did not consider the fact that there will always be a significant variation in the signal amplitude because of the chest movement. Therefore, a smaller threshold was selected to achieve realistic results. In this study, to mitigate these issues, the analysis was completed during a realistic overnight sleep study with the PSG as a gold standard for comparison. In addition, we improved the apneic detection algorithm to cope with real-life scenarios.

Conclusions

In this study, we evaluated a single-channel monitoring device for detecting vital signs, namely, HR and RR, as well as sleep apnea events. The monitoring device consisted of a mat embedded with a microbending multimode fiber. We consolidated data from 10 patients diagnosed with OSA, in which the devised sensor mat was placed underneath the patient's mattress, and raw data were collected without altering any typical configuration for the overnight sleep study. A wavelet-based analysis method was implemented for HR estimation, and satisfactory results were obtained in comparison with the reference ECG. RRs were detected from the derived effort respiratory signal after removing the nonlinear trend. Furthermore, the proposed method delivered results close to those of the reference piezoelectric thoracic belt. Both HR and RR were computed via a sliding time window of 30 seconds with an overlap of 15 seconds. The apneic events were detected on a minute-by-minute basis through an adaptive histogram-based thresholding approach. The suggested method provided average results for the distinction between normal breathing and apneic breathing events. Nevertheless, the results are encouraging considering the relative complexity of

diagnosing sleep apnea via PSG. Indeed, the proposed sensor is not designed to substitute the gold-standard method. However, it can be seen as an assistive tool capable of providing longitudinal data without interfering with the subject's everyday activities. Longitudinal data enable monitoring trends in vital signs that, in turn, can help to predict clinical deterioration in patients diagnosed with sleep-disordered breathing or cardiovascular diseases. In future work, we plan to integrate pulse oximetry with the proposed sensor mat to investigate the impact of adding another sensing modality for apneic event detection.

Acknowledgments

The authors would like to thank all participants who were involved in the study. The authors would also like to acknowledge the funding received from the Research Center on Aging—University of Sherbrooke for supporting the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information about the MFOS, data analysis, vital signs detection, as well as statistical analysis.
[PDF File (Adobe PDF File), 291 KB - jmir_v22i9e18297_app1.pdf]

Multimedia Appendix 2
High-quality PNG images.
[ZIP File (Zip Archive), 47297 KB - jmir_v22i9e18297_app2.zip]

References

1. Gandarillas MA, Goswami N. Merging current health care trends: innovative perspective in aging care. Clin Interv Aging 2018;13:2083-2095 [FREE Full text] [doi: 10.2147/CIA.S177286] [Medline: 30425463]
2. Winter V, Schreyögg J, Thiel A. Hospital staff shortages: environmental and organizational determinants and implications for patient satisfaction. Health Policy 2020 Apr;124(4):380-388. [doi: 10.1016/j.healthpol.2020.01.001] [Medline: 31973906]
3. Morgan P. Predicted shortages of physicians might even disappear if we fully account for PAs and NPs. J Am Acad Physician Assist 2019 Oct;32(10):51-53. [doi: 10.1097/01.JAA.0000580580.89002.f4] [Medline: 31513036]
4. Bradley TD, Floras JS. Sleep apnea and heart failure: part I: obstructive sleep apnea. Circulation 2003 Apr 1;107(12):1671-1678. [doi: 10.1161/01.CIR.0000061757.12581.15] [Medline: 12668504]
5. Cook DJ, Krishnan N. Mining the home environment. J Intell Inf Syst 2014 Dec;43(3):503-519 [FREE Full text] [doi: 10.1007/s10844-014-0341-4] [Medline: 25506128]
6. Sadek I, Biswas J, Abdulrazak B. Ballistocardiogram signal processing: a review. Health Inf Sci Syst 2019 Dec;7(1):10. [doi: 10.1007/s13755-019-0071-7] [Medline: 31114676]
7. Malasinghe L, Ramzan N, Dahal K. Remote patient monitoring: a comprehensive study. J Ambient Intell Human Comput 2017 Oct 26;10(1):57-76. [doi: 10.1007/s12652-017-0598-x]
8. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors (Basel) 2019 May 9;19(9):1-45 [FREE Full text] [doi: 10.3390/s19092164] [Medline: 31075985]
9. Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. Crit Care 2019 May 30;23(1):194 [FREE Full text] [doi: 10.1186/s13054-019-2485-7] [Medline: 31146792]
10. Gaisl T, Baumgartner P, Rejmer P, Osswald M, Roeder M, Thiel S, et al. Prevalence of obstructive sleep apnea in patients with thoracic aortic aneurysm: a prospective, parallel cohort study. Respiration 2020;99(1):19-27. [doi: 10.1159/000502892] [Medline: 31553996]
11. Tietjens JR, Claman D, Kezirian EJ, de Marco T, Mirzayan A, Sadroonri B, et al. Obstructive sleep apnea in cardiovascular disease: a review of the literature and proposed multidisciplinary clinical management strategy. J Am Heart Assoc 2019 Jan 8;8(1):e010440 [FREE Full text] [doi: 10.1161/JAHA.118.010440] [Medline: 30590966]
12. Rundo JV. Obstructive sleep apnea basics. Cleve Clin J Med 2019 Sep 1;86(9 Suppl 1):2-9 [FREE Full text] [doi: 10.3949/ccjm.86.s1.02] [Medline: 31509498]
13. Kimoff R. Obstructive Sleep Apnea. Semantic Scholar 2016. [doi: 10.1016/B978-1-4557-3383-5.00088-9]
14. Kasai T, Floras JS, Bradley TD. Sleep apnea and cardiovascular disease: a bidirectional relationship. Circulation 2012 Sep 18;126(12):1495-1510. [doi: 10.1161/CIRCULATIONAHA.111.070813] [Medline: 22988046]
15. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph beddit sleep tracker: a validation study. J Clin Sleep Med 2019 Mar 15;15(3):483-487 [FREE Full text] [doi: 10.5664/jcsm.7682] [Medline: 30853052]

16. Starr I, Rawson A, Schroeder H. Apparatus for recording the heart's recoil and the blood's impacts in man (ballistocardiograph), experiments on the principles involved, records in normal and abnormal conditions. Am J Physiol 1938;123:195.
17. Etemadi M, Inan OT. Wearable ballistocardiogram and seismocardiogram systems for health and performance. J Appl Physiol (1985) 2018 Feb 01;124(2):452-461 [FREE Full text] [doi: 10.1152/japplphysiol.00298.2017] [Medline: 28798198]
18. Inan OT, Migeotte P, Park K, Etemadi M, Tavakolian K, Casanella R, et al. Ballistocardiography and seismocardiography: a review of recent advances. IEEE J Biomed Health Inform 2015 Jul;19(4):1414-1427. [doi: 10.1109/JBHI.2014.2361732] [Medline: 25312966]
19. He DD, Winokur ES, Sodini CG. An ear-worn vital signs monitor. IEEE Trans Biomed Eng 2015 Nov;62(11):2547-2552 [FREE Full text] [doi: 10.1109/TBME.2015.2459061] [Medline: 26208264]
20. Will C, Shi K, Schellenberger S, Steigleder T, Michler F, Fuchs J, et al. Radar-based heart sound detection. Sci Rep 2018 Jul 26;8(1):11551 [FREE Full text] [doi: 10.1038/s41598-018-29984-5] [Medline: 30068983]
21. Tran VP, Al-Jumaily AA, Islam SM. Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: a comprehensive review. Big Data Cogn Comput 2019 Jan 1;3(1):3. [doi: 10.3390/bdcc3010003]
22. Laurino M, Arcarisi L, Carbonaro N, Gemignani A, Menicucci D, Tognetti A. A smart bed for non-obtrusive sleep analysis in real world context. IEEE Access 2020;8:45664-45673. [doi: 10.1109/access.2020.2976194]
23. Conn NJ, Schwarz KQ, Borkholder DA. In-home cardiovascular monitoring system for heart failure: comparative study. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e12419 [FREE Full text] [doi: 10.2196/12419] [Medline: 30664492]
24. Shao D, Tsow F, Liu C, Yang Y, Tao N. Simultaneous monitoring of ballistocardiogram and photoplethysmogram using a camera. IEEE Trans Biomed Eng 2017 May;64(5):1003-1010 [FREE Full text] [doi: 10.1109/TBME.2016.2585109] [Medline: 27362754]
25. Li KH, White FA, Tipoe T, Liu T, Wong MC, Jesuthasan A, et al. The current state of mobile phone apps for monitoring heart rate, heart rate variability, and atrial fibrillation: narrative review. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e11606 [FREE Full text] [doi: 10.2196/11606] [Medline: 30767904]
26. Guidoboni G, Sala L, Enayati M, Sacco R, Szopos M, Keller JM, et al. Cardiovascular function and ballistocardiogram: a relationship interpreted via mathematical modeling. IEEE Trans Biomed Eng 2019 Oct;66(10):2906-2917 [FREE Full text] [doi: 10.1109/TBME.2019.2897952] [Medline: 30735985]
27. Paalasmaa J, Waris M, Toivonen H, Leppäkorpi L, Partinen M. Unobtrusive online monitoring of sleep at home. Conf Proc IEEE Eng Med Biol Soc 2012;2012:3784-3788. [doi: 10.1109/EMBC.2012.6346791] [Medline: 23366752]
28. Chen Z, Lin M, Chen F, Lane N, Cardone G, Wang R. Unobtrusive Sleep Monitoring using Smartphones. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth'17; May 5-8, 2013; Venice, Italy. [doi: 10.4108/icst.pervasivehealth.2013.252148]
29. Tenhunen M, Elomaa E, Sistonen H, Rauhala E, Himanen S. Emfit movement sensor in evaluating nocturnal breathing. Respir Physiol Neurobiol 2013 Jun 15;187(2):183-189. [doi: 10.1016/j.resp.2013.03.009] [Medline: 23583829]
30. Hwang SH, Lee HJ, Yoon HN, Jung DW, Lee YG, Lee YJ, et al. Unconstrained sleep apnea monitoring using polyvinylidene fluoride film-based sensor. IEEE Trans Biomed Eng 2014 Jul;61(7):2125-2134. [doi: 10.1109/TBME.2014.2314452] [Medline: 24718565]
31. Beattie ZT, Hayes TL, Guilleminault C, Hagen CC. Accurate scoring of the apnea-hypopnea index using a simple non-contact breathing sensor. J Sleep Res 2013 Jun;22(3):356-362 [FREE Full text] [doi: 10.1111/jsr.12023] [Medline: 23363404]
32. Waltisberg D, Amft O, Brunner DP, Troster G. Detecting disordered breathing and limb movement using in-bed force sensors. IEEE J Biomed Health Inform 2017 Jul;21(4):930-938. [doi: 10.1109/JBHI.2016.2549938] [Medline: 27076472]
33. Wang Z, Zhou X, Zhao W, Liu F, Ni H, Yu Z. Assessing the severity of sleep apnea syndrome based on ballistocardiogram. PLoS One 2017;12(4):e0175351 [FREE Full text] [doi: 10.1371/journal.pone.0175351] [Medline: 28445548]
34. Hsu M, Fang S, Wang F, Chan H, Huang H, Yang S. Sleep apnea assessment using declination duration-based global metrics from unobtrusive fiber optic sensors. Physiol Meas 2019 Jul 30;40(7):075005. [doi: 10.1088/1361-6579/ab21b5] [Medline: 31361598]
35. Torres M, Colominas MA, Schlotthauer G, Flandrin P. A Complete Ensemble Empirical Mode Decomposition With Adaptive Noise. In: International Conference on Acoustics, Speech and Signal Processing. 2011 Presented at: ICASSP'11; May 22-27, 2011; Prague, Czech Republic. [doi: 10.1109/icassp.2011.5947265]
36. Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020 Apr 9;11(1):1760 [FREE Full text] [doi: 10.1038/s41467-020-15432-4] [Medline: 32273514]
37. Huysmans D, Borzée P, Testelmans D, Buyse B, Willemen T, Huffel SV, et al. Evaluation of a commercial ballistocardiography sensor for sleep apnea screening and sleep monitoring. Sensors (Basel) 2019 May 8;19(9):1-17 [FREE Full text] [doi: 10.3390/s19092133] [Medline: 31072036]
38. Joe H, Yun H, Jo S, Jun MB, Min B. A review on optical fiber sensors for environmental monitoring. Int J Precis Eng Manuf-Green Tech 2018 Feb 14;5(1):173-191. [doi: 10.1007/s40684-018-0017-6]
39. McKnight M, Agcayazi T, Ghosh T, Bozkurt A. Fiber-based sensors: enabling next-generation ubiquitous textile systems. Wearable Technol Med Health Care 2018:171. [doi: 10.1016/b978-0-12-811810-8.00008-7]

40. Leal-Junior AG, Diaz CA, Avellar LM, Pontes MJ, Marques C, Frizera A. Polymer optical fiber sensors in healthcare applications: a comprehensive review. Sensors (Basel) 2019 Jul 18;19(14):1-30 [FREE Full text] [doi: 10.3390/s19143156] [Medline: 31323734]
41. Sadek I, Seet E, Biswas J, Abdulrazak B, Mokhtari M. Nonintrusive vital signs monitoring for sleep apnea patients: a preliminary study. IEEE Access 2018;6:2506-2514. [doi: 10.1109/access.2017.2783939]
42. Klap T, Shinar Z. Using piezoelectric sensor for continuous-contact-free monitoring of heart and respiration rates in real-life hospital settings. Comput Cardiol 2013 Apr 7;2(1):54-90 [FREE Full text] [doi: 10.26522/ti.v2i1.745]
43. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985 Mar;32(3):230-236. [doi: 10.1109/TBME.1985.325532] [Medline: 3997178]
44. Hunt KJ, Fankhauser SE, Saengsuwan J. Identification of heart rate dynamics during moderate-to-vigorous treadmill exercise. Biomed Eng Online 2015 Dec 21;14:117 [FREE Full text] [doi: 10.1186/s12938-015-0112-7] [Medline: 26689430]
45. Taebi A, Mansy HA. Time-frequency distribution of seismocardiographic signals: a comparative study. Bioengineering (Basel) 2017 Apr 7;4(2):1-21 [FREE Full text] [doi: 10.3390/bioengineering4020032] [Medline: 28952511]
46. Vandenberk T, Stans J, Mortelmans C, van Haelst R, van Schelvergem G, Pelckmans C, et al. Clinical validation of heart rate apps: mixed-methods evaluation study. JMIR Mhealth Uhealth 2017 Aug 25;5(8):e129 [FREE Full text] [doi: 10.2196/mhealth.7254] [Medline: 28842392]
47. Severo M, Gama J. Change Detection with Kalman Filter and CUSUM. In: Todorovski L, Lavra N, Jantke KP, editors. The Discovery Science Encyclopedia. Berlin, Heidelberg: Springer; 2006.
48. Antink CH, Mai Y, Aalto R, Bruser C, Leonhardt S, Oksala N, et al. Ballistocardiography can estimate beat-to-beat heart rate accurately at night in patients after vascular intervention. IEEE J Biomed Health Inform 2020 Aug;24(8):2230-2237. [doi: 10.1109/JBHI.2020.2970298] [Medline: 32011272]
49. Inan OT, Etemadi M, Widrow B, Kovacs GT. Adaptive cancellation of floor vibrations in standing ballistocardiogram measurements using a seismic sensor as a noise reference. IEEE Trans Biomed Eng 2010 Mar;57(3):722-727. [doi: 10.1109/TBME.2009.2018831] [Medline: 19362900]

Abbreviations

AAD: average absolute deviation
AHI: apnea-hypopnea index
BCG: ballistocardiogram
BR: breathing rate
ECG: electrocardiogram
EMD: empirical mode decomposition
HR: heart rate
HSAT: home sleep apnea test
LoA: limits of agreement
LOOCV: leave-one-out-cross-validation
MAD: median absolute deviation
MAPE: mean absolute percentage error
MFOS: microbend fiber optic sensor
MODWT: maximal overlap discrete wavelet transform
NHG: National Healthcare Group
NMAE: normalized mean absolute error
NRMSE: normalized root mean square error
OSA: obstructive sleep apnea
PSG: polysomnography
RR: respiratory rate

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by A Suliman, S VanHuffel; comments to author 06.04.20; revised version received 10.04.20; accepted 26.07.20; published 18.09.20.

Please cite as:
Sadek I, Heng TTS, Seet E, Abdulrazak B
A New Approach for Detecting Sleep Apnea Using a Contactless Bed Sensor: Comparison Study
J Med Internet Res 2020;22(9):e18297
URL: http://www.jmir.org/2020/9/e18297/
doi:10.2196/18297
PMID:32945773

Original Paper

Automated Fall Detection Algorithm With Global Trigger Tool, Incident Reports, Manual Chart Review, and Patient-Reported Falls: Algorithm Development and Validation With a Retrospective Diagnostic Accuracy Study

Elisa Dolci[1*], MScN, RN; Barbara Schärer[2*], MScN, RN; Nicole Grossmann[3,4], MScN, RN; Sarah Naima Musy[2,4], PhD; Franziska Zúñiga[4], PhD, RN; Stefanie Bachnick[4], PhD, RN; Michael Simon[2,4], PhD, RN

[1]MediZentrum Täuffelen, Täuffelen, Switzerland

[2]Nursing & Midwifery Research Unit, Inselspital Bern University Hospital, Bern, Switzerland

[3]Department of General Internal Medicine, Inselspital Bern University Hospital, Bern, Switzerland

[4]Institute of Nursing Science, Department of Public Health, Faculty of Medicine, University of Basel, Basel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Michael Simon, PhD, RN
Institute of Nursing Science, Department of Public Health
Faculty of Medicine
University of Basel
Bernoullistrasse 28
Basel, 4056
Switzerland
Phone: 1 41 612670912
Email: m.simon@unibas.ch

Abstract

Background: Falls are common adverse events in hospitals, frequently leading to additional health costs due to prolonged stays and extra care. Therefore, reliable fall detection is vital to develop and test fall prevention strategies. However, conventional methods—voluntary incident reports and manual chart reviews—are error-prone and time consuming, respectively. Using a search algorithm to examine patients' electronic health record data and flag fall indicators offers an inexpensive, sensitive, cost-effective alternative.

Objective: This study's purpose was to develop a fall detection algorithm for use with electronic health record data, then to evaluate it alongside the Global Trigger Tool, incident reports, a manual chart review, and patient-reported falls.

Methods: Conducted on 2 campuses of a large hospital system in Switzerland, this retrospective diagnostic accuracy study consisted of 2 substudies: the first, targeting 240 patients, for algorithm development and the second, targeting 298 patients, for validation. In the development study, we compared the new algorithm's in-hospital fall rates with those indicated by the Global Trigger Tool and incident reports; in the validation study, we compared the algorithm's in-hospital fall rates with those from patient-reported falls and manual chart review. We compared the various methods by calculating sensitivity, specificity, and predictive values.

Results: Twenty in-hospital falls were discovered in the development study sample. Of these, the algorithm detected 19 (sensitivity 95%), the Global Trigger Tool detected 18 (90%), and incident reports detected 14 (67%). Of the 15 falls found in the validation sample, the algorithm identified all 15 (100%), the manual chart review identified 14 (93%), and the patient-reported fall measure identified 5 (33%). Owing to relatively high numbers of false positives based on falls present on admission, the algorithm's positive predictive values were 50% (development sample) and 47% (validation sample). Instead of requiring 10 minutes per case for a full manual review or 20 minutes to apply the Global Trigger Tool, the algorithm requires only a few seconds, after which only the positive results (roughly 11% of the full case number) require review.

Conclusions: The newly developed electronic health record algorithm demonstrated very high sensitivity for fall detection. Applied in near real time, the algorithm can record in-hospital falls events effectively and help to develop and test fall prevention measures.

(J Med Internet Res 2020;22(9):e19516) doi:10.2196/19516

KEYWORDS

falls; adverse event; harm; algorithm; natural language processing

Introduction

Falls are among the most common adverse events in hospitals [1]. For example, US hospitals report fall rates per 1000 patient days ranging from 3.3 to 11.5 [1], while Swiss studies have reported rates between 2.2 and 8.9 [2,3]. A fall is defined as "an unexpected event in which the person comes to rest on the ground, floor or other lower level [4]." Approximately 25% of in-hospital falls lead to injuries, the most serious of which are fractures and intracranial hemorrhages [1]. Increasing disability-related dependence, length of stay, and care costs make falls a major burden, not only for the affected patients, but for the entire health care system [5].

Therefore, the development, evaluation, and improvement of interventions to prevent falls are a high-priority for health researchers. However, quick, accurate, and cost-effective fall detection methods are needed to provide reliable and robust fall data; currently, no such method is available.

The 3 most common fall detection methods are voluntary incident reporting, chart reviews, and patient self-reports [6]. Voluntary incident reports are provided by frontline staff directly, often nurses, involved in falls or in the action leading up to it [6-8]. Traditional chart reviews consist of reading the full patient records. The Global Trigger Tool is a retrospective chart review method developed by the Institute for Healthcare Improvement. It is widely used internationally for detecting adverse events and uses so-called triggers (ie, key elements that help reviewers to identify potential adverse events including falls) [9-12]. Finally, in Switzerland, prevalence data of patient-reported in-hospital falls are recorded based on the LPZ method (*Landelijke Prevalentiemeting Zorgproblemen*, National Prevalence Measurement of Quality of Care) [13]. Unlike incident reports, where staff fill forms when a fall occurs, this measure is based on a self-reported questionnaire or retrospective interview by hospital staff (30-day period). Since 2009, this measurement has been conducted annually on a single day by the ANQ (Swiss National Association for Quality Development in Hospitals and Clinics) in almost all Swiss acute care hospitals [14].

Each of these methods is limited in important ways. Nurse voluntary incident reports are prone to underreporting or nonreporting [8]. Chart review is time consuming and costly. And the LPZ/ANQ patient reports are affected both by underreporting and by the lack of flexibility regarding their timing. These limitations make a quick, accurate, and timely fall detection system highly desirable.

One very promising target for research is hospitals' electronic health records. As digital databases, these offer the opportunity to develop automated detection algorithms. In addition to being inexpensive to use, such methods would potentially be both highly sensitive and fast enough to deliver real-time or near real-time data on adverse events [6,12,15]. Setting the technical advantages of electronic health record–based adverse event detection aside, research in this area is still relatively new and not systematically under study. In their review, Musy et al [16] found a broad interstudy variation in reported adverse event prevalence and positive predictive value, which led to difficulties regarding interpretation. To improve quality, they see the need for adequate reporting of future adverse event detection studies [16].

Because of these and other potential advantages, algorithms for adverse event detection are being developed more and more [17-20]. To our knowledge, only one—in Japan—has used electronic health record data for fall detection—with mixed results [21]. The algorithm, which used natural language processing to read medical professionals' chart notes for a sample of 1204 patients, was highly sensitive regarding fall detection (100%); however, its positive predictive value was very low (6%) [21]. Therefore, this study's goal was to develop and validate an electronic health record–based fall detection algorithm (using the given German-language electronic health record systems), then to test its diagnostic accuracy against manual chart review and patients' reports of falls.

Methods

Design (Study 1 and Study 2)

This retrospective diagnostic accuracy study consisted of 2 parts: the first, for algorithm development and the second, for validation. For the development of the electronic health record fall detection algorithm, we used falls identified through the Global Trigger Tool in a previous study [22] along with incident reports for comparison. To validate the algorithm, we collected additional data to compare the algorithm against falls identified through the manual chart review of electronic health records ("Global Trigger Tool for falls only") and patient-reported falls based on the LPZ/ANQ measure.

Setting (Study 1 and Study 2)

This study was conducted in one large Swiss university hospital and in one rural hospital belonging to the same hospital system in the German part of Switzerland. From the university hospital, 2 departments participated: Internal Medicine (110 beds, approximately 4600 admissions per year, average length of stay 6.5 days) and Orthopedics and Plastic Surgery (59 beds, approximately 2400 admissions per year, average length of stay 7.4 days). The rural hospital's general medicine, general surgery, visceral surgery, traumatology, and orthopedics units participated (totaling 72 beds, approximately 5200 admissions per year, average length of stay 5.4 days). Because internal medicine and orthopedics departments treat older people with chronic diseases, which are risk factors for falls, these departments have relatively high fall rates. The university hospital introduced electronic health records in 2011, while the rural hospital introduced electronic health records in 2010. Until September 2017, the 2 facilities had separate electronic health

record systems but with similar internal databases. The algorithm development study occurred only in the university hospital's Internal Medicine department; the validation was performed in the 2 university hospital departments and in all the participating departments of the rural hospital.

Study 1: Algorithm Development

Sample and Sampling

The algorithm was developed by one of the first authors (BS), using data from a previous Global Trigger Tool study [22]. That study's [22] sample consisted of patients admitted to the Internal Medicine department between September 1, 2016 and August 31, 2017. Further inclusion criteria were (1) adult patients (aged ≥18 years), (2) closed and completed patient record, and (3) inpatients with a length of stay of at least 24 hours. From the eligible patients' data sets, we randomly selected 240. The first 120 (the development data set) were used to develop the algorithm; the remaining 120 (the testing data set) were used to validate the algorithm. Because this was a diagnostic accuracy study, no formal power analysis for sample size was conducted. However, for the Global Trigger Tool study [22] and an expected overall adverse event rate of 12.3% as detected by Soop et al [23], a sample size of 240 gives a 95% confidence interval of 8.9%-16.7%. Electronic patient records (n=30) from hospitalized patients were randomly selected each month and checked for eligibility, including their general consent, by one reviewer. Of the 30 records, the first 20 that were eligible were used for chart review each month.

Data Collection and Management

Three health care professionals completed the Global Trigger Tool review: 2 nurses as primary reviewers (with 5 years of clinical experience and knowledge of the electronic health record) and a physician (with 10 years of clinical experience). As preparation, reviewers read the Health Care Improvement handbook for the Global Trigger Tool and underwent training provided on the website [9]. Furthermore, the primary reviewers practiced on 15 patient charts, 5 of which were discussed with the physician. The interrater reliability (Cohen κ) on the number of adverse events between the primary reviewers was 0.96 and between the primary reviewers and the physician was 0.98.

Variables and Measurement

In order to describe the sample, we also extracted basic patient characteristics, such as age, gender, length of stay, and primary diagnosis, from the electronic health records. All 4 variables were also considered risk factors for falls [3]. We focused on in-hospital fall rates recorded by our algorithm, the Global Trigger Tool, and voluntary incident reports (Table 1). For the in-hospital falls variable we used Hauer et al's definition [4], which includes 3 categories: assisted falls (eg, when the patient begins to fall and is assisted to the ground by another person); unassisted falls; and falls resulting from syncopes, epileptic seizures, strokes, and hypoglycemia. All types of accidents (eg, sporting, road traffic, work-related) leading to falls as the cause for hospitalization were excluded. Electronic health record reviews using the Global Trigger Tool were limited to 20 minutes. Multimedia Appendix 1 presents details of these variables.

Table 1. Variables of the algorithm development and validation study.

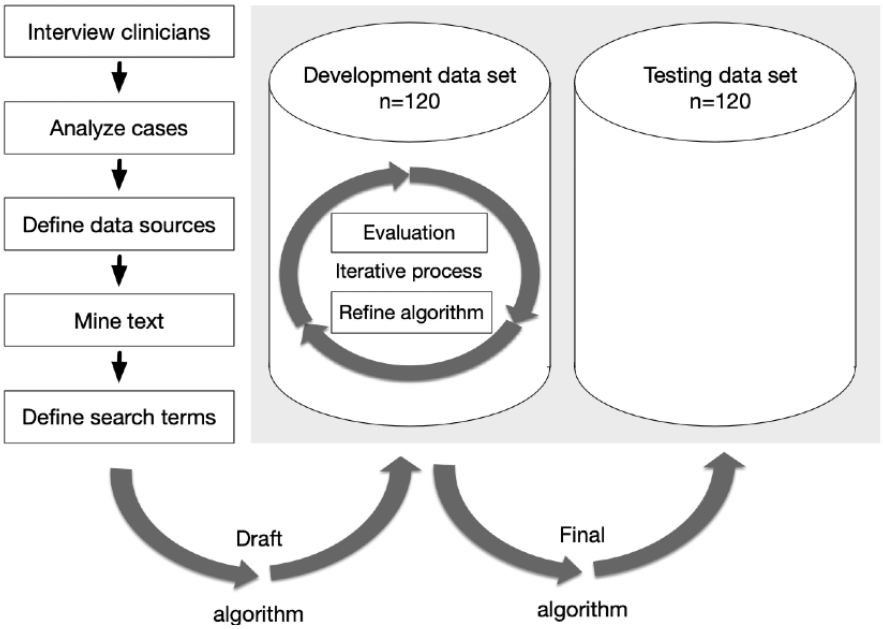
Variable	Description	Development: Method as data source	Validation: Method as data source
Age	Years at the time of admission	Global Trigger Tool	LPZ/ANQ[a] measure
Gender	Female or male sex	Global Trigger Tool	LPZ/ANQ measure
Length of stay	Number of days in the hospital	Global Trigger Tool	Manual chart review
Primary diagnose	Cardiac, musculoskeletal, endocrinologic, gastrointestinal, pulmonary, infectious, neurological, psychiatric, cancer, dementia	Global Trigger Tool	LPZ/ANQ measure
Presence of fall	Yes or no	Algorithm, Global Trigger Tool, voluntary incident reports	Algorithm, manual chart review, LPZ/ANQ-measure
Fall rates	Number of falls	Algorithm, Global Trigger Tool, voluntary incident reports	Algorithm, manual chart review, LPZ/ANQ measure
Time for data collection	Time for data collection in hours	Global Trigger Tool	Manual chart review

[a]LPZ/ANQ: Landelijke Prevalentiemeting Zorgproblemen/Swiss National Association For Quality Development in Hospitals and Clinics.

Algorithm Development

For the algorithm development, a positive test case was comprehensively analyzed to identify appropriate data sources within the electronic health record system (Figure 1). Nurses' and physicians' narrative progress notes proved the most promising data source. It was important to take both sets of progress notes into consideration, as fall events were not always mentioned by both physicians and nurses. We compiled a list of fall-related terms—*fall*, *fell*, *slip*, *floor*, etc, which would be used to describe an event. Common terms used in the record were *am Boden* (*on the floor*), *ausgerutscht* (*slipped*), *Sturz/Stürze* (*fall/falls*), *Synkope/synkopiert* (*collapse/collapsed*). After identifying the most fall-specific terms, we transformed the words into search strings to build the algorithm. Extraction was performed using the widely used structured query language (SQL) for Oracle Databases.

Figure 1. Algorithm development process.



To distinguish true positives from false positives, algorithm results were compared to those of the manual Global Trigger Tool study in the development set. For false positives, a comprehensive investigation was performed to identify misleading terms. For instance, the German term *Boden* (*floor*) was used to report that a patient has been found on the floor. However, the term was also used in other contexts, notably *Bodenbett* (low-level bed used to reduce the risk of bed fall injuries). As this resulted in a large number of false positive cases, *Bodenbett* was added to the exclusion criterion in the query.

If an event identified by the manual Global Trigger Tool study was not found by the algorithm, progress notes were analyzed comprehensively to identify further search terms. Since first iterations revealed difficulties distinguishing inpatient falls from fall injuries present on admission, we used a text-mining approach. This identified terms related to emergency situations, for example, *emergency* or *ambulance*, or the term *at home* as indicators of preadmission events. Fall events could also result from critical events (eg, loss of consciousness) or accidents, either of which can lead to emergency hospital admissions on their own.

Through the process described above, selection criteria were defined for patient records (Multimedia Appendix 2). These were used in the algorithm. One of our first steps was to query records indicating the presence of fall events; in subsequent steps we excluded events present on admission. The algorithm's accuracy was compared to the manual Global Trigger Tool results, then optimized by iteratively testing it with the development set of electronic health record medical charts (Figure 1).

Study 2: Algorithm Validation Study

Sample and Sampling

From each of the 2 university hospital departments and for the entire rural hospital, 100 patients were randomly selected, for a total sample of 300 patients. In order to have enough patients and the same number from every site, patients were selected based on participation in LPZ/ANQ data collection in 2015, 2016, or 2017. Of a total of 942 patients invited to participate, 705 accepted (75%) for the 3 years. Further inclusion criteria were (1) age ≥18 years, (2) closed and completed patient record, and (3) length of stay ≥24 hours (to avoid outpatients). No formal sample size estimation was conducted. Based on the paper of Schwendimann and colleagues [3], a sample size of 300 yielding a 95% confidence interval ranging of 4.65%-10.89% would be expected.

Variables and Measurement

Besides the demographic (age, gender) and diagnostic variables (length of stay, main diagnosis), we focused on the in-hospital fall rates recorded by the 3 methods (Table 1). We also registered the time each method required for data collection. For the fall occurrence variable, we used Hauer et al's [4] definition and the same inclusion and exclusion criteria as those of our algorithm development study.

Data Collection and Management

We used data of patients who participated in the LPZ/ANQ survey, which looks retrospectively at the 30 days before the LPZ/ANQ measurement day. ("Did you fall in the previous 30 days?") A manual chart review of electronic health records of the 300 patients was carried out using the various electronic health record systems. The electronic health records included patient demographics, diagnoses, clinical data, laboratory results, order entry, reports, and narrative notes.

The manual chart review of the 300 electronic health records was performed by a researcher (ED, one of the first authors) with 5 years of clinical experience in internal medicine as well as knowledge of the hospital's records. Physicians' and nurses' progress notes, physicians' discharge summaries and nurses' anamneses of every chart were reviewed. To explore the reliability of ED's manual chart review, the electronic health records of 20 patients for each of the 3 samples from the 2

departments and the rural hospitals site (20% of the overall sample) were double-reviewed by a clinical nurse specialist with at least 5 years of experience in each respective setting. We obtained a Cohen κ of 0.87, indicating good interrater reliability. Finally, the electronic health record fall detection algorithm was applied by the other first author (BS, informatics nurse) to the full 300-patient sample.

Algorithm Development and Validation Studies (Study 1 and Study 2)

Ethical Considerations

Ethical approval for this study was obtained from the regional Ethics Committee of Bern (development study: 2016-01720; validation study: 2018-01250). Participants of both studies gave informed consent. For data management in both studies, SharePoint (Microsoft Inc) was used. After the merging of the study sample, the patients' identification numbers were removed and the patients were coded from 1 to 240 and from 1 to 300 for the development and validation studies, respectively. To minimize bias, ED and BS conducted their analyses independently.

Data Analysis

R statistical software was used in Windows for all analyses [24]. Version 3.2.4 of the tm package [25] was used for the text-mining part of the algorithm development; version 3.2.5 was used for all other analyses.

For both the development and validation studies, descriptive analyses with means and percentages were conducted for 5 variables: age, gender, length of stay, main diagnosis of the patients, and time for data collection. To gauge diagnostic accuracy, true positive, false positive, false negative, and true negative fall rates were determined. Furthermore, sensitivity, specificity, positive predictive value, and negative predictive value were calculated for each detection method.

Initially the Global Trigger Tool manual method was considered the gold standard for fall detection [10]. However, we recognized that our algorithm detected valid cases that the manual Global Trigger Tool method did not. For example, where patient records are more extensive due to longer hospitalization, reviewer fatigue can lead to inpatient fall events going unnoticed. Therefore, to test the accuracy in the first study, we created a pseudo gold standard by combining the results of the manual Global Trigger Tool study with those of incident reporting and the electronic health record algorithm in the first study and manual chart review, LPZ/ANQ patient reports, and the algorithm in the second study. For both studies, cases with differences between measures (ie, fall in one method versus nonfall in another) were discussed by ED and BS until an agreement was reached.

Results

Study 1: Algorithm Development

Descriptive Analysis

The mean patient age was 69.3 years (range 18-103). The mean length of stay for patients with fall events of 24.1 (SD 17.6 days) was longer that of the overall study population 13.8 (SD 11.6) days. The study population's main diagnoses were neurological diseases (48/240, 20.0%), sepsis (37/240, 15.0%), infectious diseases (32/240, 13.3%), and neoplasms (28/240, 11.7%).

Diagnostic Accuracy

We report the overall results of the development and validation data sets together (n=240). Twenty fall events were identified by our first composite gold standard and 19 by the development algorithm (sensitivity 95%). The manual Global Trigger Tool method resulted in 18 true positives (90%), whereas incident reporting produced 14 (67%). The manual Global Trigger Tool method and incident reporting produced no false positives, whereas the algorithm resulted in 19 (negative predictive value 99%; positive predictive value 50%); however, most of these related to preadmission fall events: only 2 had no relation to fall events. As noted above, though, while 20 inpatient falls were detected by at least one of the commonly employed methods, the algorithm identified one more legitimate event than the manual Global Trigger Tool method; incident reporting missed 6. For more detailed information see Table 2.

Table 2. Diagnostic accuracy results of the comparison between algorithm and all other detection methods in the development and validation studies.

Method	True positive, n	False positive, n	True negative, n	False negative, n	Sensitivity, %	Specificity, %	Positive predictive value, %	Negative predictive value, %
Development study, development data set (n=120)								
Algorithm	11	10	99	0	100	91	52	100
Manual GTT	9	0	109	2	82	100	100	98
Incident reporting	7	0	109	4	64	100	100	96
Development study, testing data set (n=120)								
Algorithm	8	9	101	2	80	92	47	98
Manual GTT	9	0	110	0	100	100	100	100
Incident reporting	7	0	110	3	70	100	100	97
Validation study (n=298)								
Algorithm	15	17	266	0	100	94	47	100
Manual chart review	14	0	283	1	93	100	100	99
ANQ[b] measure	5	0	283	10	33	100	100	97

[a]GTT: Global Trigger Tool.

[b]ANQ: Swiss National Association For Quality Development in Hospitals and Clinics.

Study 2: Algorithm Validation

Descriptive Analysis

Two patients were excluded because they were minors (<18 years), reducing the total sample to 298 adult inpatients (age: mean 65.3, SD 18.0 years; length of stay: mean 12.1, SD 13.2 days), of which 152 (51.0%) were female (153/298). The most common diagnoses were cardiac (170/298, 57.0%), musculoskeletal (165/298, 55.4%), and endocrine diseases (88/298, 29.5%). The demographics of patients with in-hospital falls versus those without falls did not show any significant differences; however, patients with falls stayed longer in hospital (mean 22.6, SD 19.0 days versus mean 11.5, SD 12.6; *P*=.03). For the manual chart review, ED spent roughly 54 hours (time per record: mean 10.8 minutes).

Diagnostic Accuracy

The pseudo gold standard detected 15 falls over the 3606 patient-days (4.16 falls per 1000 patient days) covered by the data period for our study sample (2015-2017). The algorithm recognized all 15 fall events (sensitivity 100%), and the manual chart review identified 14 fall events (93%), whereas the ANQ measure identified only 5 (33%). The algorithm produced no false negatives but 17 false positives, leading to a negative predictive value of 100% and a positive predictive value of 47%. For more detailed information see Table 2.

Discussion

Principal Findings

For this study, we first developed an electronic health record algorithm in a single-site sample of 240 patients (development study). We then validated the electronic health record algorithm in a 298-patient sample in 3 departments on 2 sites (validation study). From an epidemiological point of view, the fall rates of 8.3 (development study) and 4.2 (validation study) per 1000 patient-days fit within the range of 2.2-8.9 per 1000 patient-days for Switzerland reported elsewhere [2,3]. For both of our samples, the electronic health record algorithm showed very high sensitivity (95% and 100%), as confirmed by a pseudo gold standard combining the Global Trigger Tool, chart review, voluntary incident reporting, and patient reports of falls. Incident reporting achieved a sensitivity of 67%, the Global Trigger Tool achieved a sensitivity of 90%, manual chart review achieved a sensitivity of 93%, and the patient-reported method of the LPZ/ANQ achieved a sensitivity of only 33%. In the validation study, we found the algorithm's specificity decreased to 94%, reflecting 17 false positives.

Although the Global Trigger Tool in the development study and manual chart review for falls in the validation study are viewed as the most sensitive methods to identify adverse events [6], our electronic health record algorithm performed slightly better than both, identifying one additional fall in our sample. Unlike manual chart review, the electronic health record algorithm automatically retrieves and evaluates fall cases and is not prone to the subjective weaknesses of manual review, including shortfalls of time, training, or stamina [17-21].

The algorithm's main disadvantage is its tendency to flag false positive cases, which reduced its positive predictive value to 50% in the development study and 47% in the validation study. All other methods shared a positive predictive value of 100%, as they produced no false positives.

Looking more closely at our algorithm's false positives, we found that all indicated actual falls, but that the falls had occurred before admission; in several cases, falls were even the reason for admission. While the presence of false positives necessitates further manual screening, the time and effort that this requires is far less than that required for full manual chart review or application of the Global Trigger Tool. For example,

while our validation study required 3218.4 minutes (298 patient records × 10.8 minutes) for full manual chart review, based on the mean time spent to review each case, identifying the 17 false positives took only 345.6 minutes ((17 patient records + 15 patient records) × 10.8 minutes)—roughly an 89% reduction.

Additionally, while we will continue to adjust the algorithm to distinguish between preadmission and inpatient falls, a history of falls is an extremely important fall risk indicator [26,27]: identifying any falls will contribute to fall prevention [3,28]. Nevertheless, the argument in favor of using our algorithm for inpatient fall detection is a matter of efficiency: instead of requiring 10 minutes per case for a full manual review or 20 to apply the Global Trigger Tool, the algorithm requires only a few seconds, after which only the positive results (roughly 11% of the full case number) require review. The algorithm can detect falls near real time and can be used on a daily or weekly basis while the patient is still in hospital. Detection during the patient stay is probably less relevant for the clinical management of individual patients but could provide a management tool to identify areas with unusually high fall incidence, which could be supported by additional resources.

Another vitally important value this study provides is the transferability of the electronic health record algorithm to other departments and institutions with a broad range of electronic health record systems. As data sources, electronic health records are rich but often somewhat chaotic, adding to the complexity of adverse event detection. Terms used to report fall events vary between settings, which could limit an algorithm's performance [29]. However, in our validation study, the same unmodified version of our algorithm returned excellent results in 3 clinical departments on 2 sites (using 2 electronic health record systems) [30-32].

In contrast, the LPZ/ANQ measure identified only 5 of 15 confirmed fall events. Underreporting and nonreporting are possible and frequent with this method, as it depends on each patient's capacity to remember and report fall events. It is well-established that retrospective reports depend on the cognitive, mental, and physical condition of the patient at the moment of the interview [33]. In addition, a patient might not know what qualifies as a fall (such as when the patient begins to fall and is assisted to the ground by another person). The low count of the LPZ/ANQ-measure is also explained by their 1-point prevalence measurement, which only captures falls from admission until the LPZ/ANQ measurement date. Because the prevalence measure can occur on any day of the hospital stay of the patients only half of the length of stay will be taken into account. If falls occur evenly distributed throughout the hospital stay the number of falls detected is also cut in half. As the LPZ/ANQ detected only about one-third or less of our sample's in-hospital falls (5 versus 15), we can only conclude that it cannot provide robust prevalence rates. Although it is not unreasonable to assume that the described biases will be similar across hospitals [34], the extent of this method's underreporting and the high cost of each primary data collection on a national scale raise doubts about its overall value.

The use of highly sensitive electronic health record algorithms to detect adverse events and small-scale validation studies such as ours opens up at least 2 productive pathways for future research. First, the current algorithm allows expanding data sets by manual screening of the cases identified by the algorithm. With the same resources (validation study 300 records), we are now able to screen 3000 records. Such a data set could then be used for refining the algorithm to improve the specificity but also allows for conducting substantive analysis (eg, on risk factors of falls). Second, the study design could serve as a template for developing additional electronic health record adverse event detection algorithms. This is particularly interesting when exploring the association of structural and process measures with quality of care outcomes in a causal inference data-fusion framework [35]. Data fusion in this context would allow using data from a validation study to overcome measurement error in routine electronic health record data.

Limitations

This study is subject to several notable limitations. First, the quality of any algorithm's results cannot surpass that of the documentation upon which it is based, that is, the quality and the completeness of the documentation define the limits of the algorithm's performance [17,29]. Therefore, heavy workloads, which influence documentation quality, also influence our algorithm's capacity to detect falls. In case of acute situations, documentation is often done on paper and later transcribed to the electronic health record, which can lead to missing information [17]. While these limitations also apply to other fall detection methods [29], both the Global Trigger Tool and manual review draw their data from broader sources, which may increase the chances of detecting traces of an event. The small sample size for both the development and the validation studies, as well as the lack of a true gold standard represent other limitations.

Finally, we based our selection of patients on the LPZ/ANQ measure, which suffers from selection bias: patients who did not speak one of the Swiss national languages, had cognitive limitations (eg, dementia or delirium), were dying or in unstable states were excluded. For example, in our validation study, only 75% of the patients participated in the LPZ/ANQ data collection.

Conclusions

For this study, we successfully developed and evaluated a newly developed algorithm for fall detection, which we tested in the electronic health records of 3 different departments situated on 2 sites. Weighing the advantages and disadvantages of the different methods used in this study, our algorithm is extremely attractive: of all the methods employed in the tests, our fall detection algorithm offered the highest sensitivity with by far the smallest time investment. And although it produced false positives, thereby necessitating a manual chart review of all identified cases, the overall time investment and sensitivity were roughly 90% better than those for the other methods with comparable sensitivity. Applied in near real time, the algorithm can record in-hospital fall events at least as effectively as manual chart review or the Global Trigger Tool but requires a small fraction of the time or human resources demanded by either. Not only will this algorithm contribute to a better understanding of inpatient falls, it will also highlight fall-influencing factors, thereby helping identify the patients with the highest risk of

falls, all of which will promote development and targeting of preventive interventions. Each implementation of this algorithm will offer an opportunity to fine-tune it, particularly to distinguish between inpatient and preadmission falls (false positives). Further research on this algorithm using a larger data sample or using the algorithm on a weekly basis can generate further data and feedback in order to improve it.

Acknowledgments

The authors are grateful to Franziska Gratwohl RN, MScN, and Prof. Dr. med. Jacques Donzé from the participating clinic for their many contributions to the algorithm development study. We also thank Elisabeth Lanz, RN, MScN, of the University Hospital Bern/Inselspital Department of Orthopaedic Surgery and Traumatology, and Regula Pfäffli, RN, MScN, of Aarberg Hospital, for their contributions to testing the reliability of manual chart review in the algorithm validation study. Finally, we thank Chris Shultis for the language corrections. No funding was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variable details.
[DOCX File , 13 KB - jmir_v22i9e19516_app1.docx]

Multimedia Appendix 2
Patient record selection criteria.
[DOCX File , 13 KB - jmir_v22i9e19516_app2.docx]

References

1. Bouldin ELD, Andresen EM, Dunton NE, Simon M, Waters TM, Liu M, et al. Falls among adult patients hospitalized in the United States: prevalence and trends. J Patient Saf 2013 Mar;9(1):13-17 [FREE Full text] [doi: 10.1097/PTS.0b013e3182699b64] [Medline: 23143749]
2. Halfon P, Eggli Y, Van Melle G, Vagnair A. Risk of falls for hospitalized patients: a predictive model based on routinely available data. J Clin Epidemiol 2001 Dec;54(12):1258-1266. [doi: 10.1016/s0895-4356(01)00406-1] [Medline: 11750195]
3. Schwendimann R, Bühler H, De Geest S, Milisen K. Falls and consequent injuries in hospitalized patients: effects of an interdisciplinary falls prevention program. BMC Health Serv Res 2006 Jun 07;6:69 [FREE Full text] [doi: 10.1186/1472-6963-6-69] [Medline: 16759386]
4. Hauer K, Lamb SE, Jorstad EC, Todd C, Becker C, PROFANE-Group. Systematic review of definitions and methods of measuring falls in randomised controlled fall prevention trials. Age Ageing 2006 Jan;35(1):5-10. [doi: 10.1093/ageing/afi218] [Medline: 16364930]
5. Cina-Tschumi B, Schubert M, Kressig RW, De Geest S, Schwendimann R. Frequencies of falls in Swiss hospitals: concordance between nurses' estimates and fall incident reports. Int J Nurs Stud 2009 Feb;46(2):164-171. [doi: 10.1016/j.ijnurstu.2008.09.008] [Medline: 18986651]
6. Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. J Biomed Inform 2003;36(1-2):131-143 [FREE Full text] [Medline: 14552854]
7. Blegen MA, Vaughn T, Pepper G, Vojir C, Stratton K, Boyd M, et al. Patient and staff safety: voluntary reporting. Am J Med Qual 2004;19(2):67-74. [doi: 10.1177/106286060401900204] [Medline: 15115277]
8. Evans SM, Berry JG, Smith BJ, Esterman A, Selim P, O'Shaughnessy J, et al. Attitudes and barriers to incident reporting: a collaborative hospital study. Qual Saf Health Care 2006 Feb;15(1):39-43 [FREE Full text] [doi: 10.1136/qshc.2004.012559] [Medline: 16456208]
9. Griffin F, Resar RK. IHI Global Trigger Tool for measuring adverse events. IHI Innovation Series White Paper. 2009. URL: www.IHI.org [accessed 2019-06-01]
10. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011 Apr;30(4):581-589 [FREE Full text] [doi: 10.1377/hlthaff.2011.0190] [Medline: 21471476]
11. Doupi P, Svaar H, Bjørn B, Deilkås E, Nylén U, Rutberg H. Use of the Global Trigger Tool in patient safety improvement efforts: Nordic experiences. Cogn Tech Work 2015:45-54. [doi: 10.1007/s10111-014-0302-2]
12. Govindan M, Van CAD, Nelson EC, Kelly-Cummings J, Suresh G. Automated detection of harm in healthcare with information technology: a systematic review. Qual Saf Health Care 2010 Oct;19(5):e11. [doi: 10.1136/qshc.2009.033027] [Medline: 20671081]
13. Landelijke Prevalentiemeting Zorgproblemen (LPZ). URL: https://ch.lpz-um.eu/de [accessed 2019-06-01]

14. Swiss National Association For Quality Development in Hospitals and Clinics (ANQ). URL: http://www.anq.ch [accessed 2019-06-01]
15. Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. J Biomed Inform 2003;36(1-2):131-143 [FREE Full text] [Medline: 14552854]
16. Musy SN, Ausserhofer D, Schwendimann R, Rothen HU, Jeitziner M, Rutjes AW, et al. Trigger Tool–Based Automated Adverse Event Detection in Electronic Health Records: Systematic Review. J Med Internet Res 2018 May 30;20(5):e198. [doi: 10.2196/jmir.9901]
17. Li Q, Melton K, Lingren T, Kirkendall ES, Hall E, Zhai H, et al. Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. J Am Med Inform Assoc 2014;21(5):776-784 [FREE Full text] [doi: 10.1136/amiajnl-2013-001914] [Medline: 24401171]
18. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc 2005;12(4):448-457 [FREE Full text] [doi: 10.1197/jamia.M1794] [Medline: 15802475]
19. Penz JFE, Wilcox AB, Hurdle JF. Automated identification of adverse events related to central venous catheters. J Biomed Inform 2007 Apr;40(2):174-182 [FREE Full text] [doi: 10.1016/j.jbi.2006.06.003] [Medline: 16901760]
20. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]
21. Toyabe S. Detecting inpatient falls by using natural language processing of electronic medical records. BMC Health Serv Res 2012 Dec 05;12:448 [FREE Full text] [doi: 10.1186/1472-6963-12-448] [Medline: 23217016]
22. Grossmann N, Gratwohl F, Musy SN, Nielen NM, Simon M, Donz J. Describing adverse events in medical inpatients using the Global Trigger Tool. Swiss Med Wkly 2019 Nov 10:45-46. [doi: 10.4414/smw.2019.20149]
23. Soop M, Fryksmark U, Köster M, Haglund B. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. Int J Qual Health Care 2009 Aug;21(4):285-291 [FREE Full text] [doi: 10.1093/intqhc/mzp025] [Medline: 19556405]
24. The R Project for Statistical Computing. The R Foundation, Vienna, Austria; 2020. URL: https://www.r-project.org/ [accessed 2019-06-01]
25. Feinerer I, Hornik K, Meyer D. Text Mining Infrastructure in. J. Stat. Soft 2008;25(5):1-54. [doi: 10.18637/jss.v025.i05]
26. Müller R, Halfens R, Schwendimann R, Müller M, Imoberdorf R, Ballmer PE. Risikofaktoren für Stürze und sturzbedingte Verletzungen im Akutspital – Eine retrospektive Fall-Kontroll-Studie. Pflege 2009 Dec 01;22(6):431-441. [doi: 10.1024/1012-5302.22.6.431]
27. Tzeng H, Yin C. Frequently observed risk factors for fall-related injuries and effective preventive interventions: a multihospital survey of nurses' perceptions. J Nurs Care Qual 2013;28(2):130-138. [doi: 10.1097/NCQ.0b013e3182780037] [Medline: 23117794]
28. Schwendimann R. [Prevention of falls in acute hospital care. Review of the literature]. Pflege 2000 Jun;13(3):169-179. [doi: 10.1024/1012-5302.13.3.169] [Medline: 10969583]
29. Halfon P, Staines A, Burnand B. Adverse events related to hospital care: a retrospective medical records review in a Swiss hospital. Int J Qual Health Care 2017 Aug 01;29(4):527-533. [doi: 10.1093/intqhc/mzx061] [Medline: 28586414]
30. Hripcsak G, Bakken S, Stetson PD, Patel VL. Mining complex clinical data for patient safety research: a framework for event discovery. J Biomed Inform 2003;36(1-2):120-130 [FREE Full text] [doi: 10.1016/j.jbi.2003.08.001] [Medline: 14552853]
31. Murphy DR, Thomas EJ, Meyer AND, Singh H. Development and Validation of Electronic Health Record-based Triggers to Detect Delays in Follow-up of Abnormal Lung Imaging Findings. Radiology 2015 Oct;277(1):81-87 [FREE Full text] [doi: 10.1148/radiol.2015142530] [Medline: 25961634]
32. Liao KP, Ananthakrishnan AN, Kumar V, Xia Z, Cagan A, Gainer VS, et al. Methods to Develop an Electronic Medical Record Phenotype Algorithm to Compare the Risk of Coronary Artery Disease across 3 Chronic Disease Cohorts. PLoS One 2015;10(8):e0136651 [FREE Full text] [doi: 10.1371/journal.pone.0136651] [Medline: 26301417]
33. ANQ. Überprüfung ANQ-Messplan auf Vollständigkeit und Relevanz: Kurzfassung zum 2. Teil des Forschungsberichts des ISGF inkl. Identifikation von Handlungsoptionen 2014:1-40. [doi: 10.7767/boehlau.9783205793199.intro]
34. ANQ. Auswertungskonzept ANQ: Nationale Prävalenzmessung Sturz & Dekubitus Erwachsene und Dekubitus Kinder. www.anq.ch 2018:1-31. [doi: 10.1007/978-3-7091-4423-7_12]
35. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. Proc Natl Acad Sci USA 2016 Jul 05;113(27):7345-7352. [doi: 10.1073/pnas.1510507113]

Abbreviations

ANQ: Swiss National Association For Quality Development in Hospitals and Clinics
LPZ: Landelijke Prevalentiemeting Zorgproblemen

Edited by G Eysenbach; submitted 24.04.20; peer-reviewed by J Mayoh, H Singh, A Snoswell; comments to author 12.06.20; revised version received 26.06.20; accepted 26.07.20; published 21.09.20.

Please cite as:
Dolci E, Schärer B, Grossmann N, Musy SN, Zúñiga F, Bachnick S, Simon M
Automated Fall Detection Algorithm With Global Trigger Tool, Incident Reports, Manual Chart Review, and Patient-Reported Falls: Algorithm Development and Validation With a Retrospective Diagnostic Accuracy Study
J Med Internet Res 2020;22(9):e19516
URL: http://www.jmir.org/2020/9/e19516/
doi:10.2196/19516
PMID:32955445

Original Paper

Comparability of Emotion Dynamics Derived From Ecological Momentary Assessments, Daily Diaries, and the Day Reconstruction Method: Observational Study

Stefan Schneider[1], PhD; Doerte U Junghaenel[1], PhD; Tania Gutsche[1], BA; Hio Wa Mak[1], PhD; Arthur A Stone[1], PhD

University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Stefan Schneider, PhD
University of Southern California
Dornsife Center for Self-Report Science
635 Downey Way
Los Angeles, CA, 90089-3332
United States
Phone: 1 2138211774
Email: schneids@usc.edu

Abstract

Background: Interest in the measurement of the temporal dynamics of people's emotional lives has risen substantially in psychological and medical research. Emotions fluctuate and change over time, and measuring the ebb and flow of people's affective experiences promises enhanced insights into people's health and functioning. Researchers have used a variety of intensive longitudinal assessment (ILA) methods to create measures of emotion dynamics, including ecological momentary assessments (EMAs), end-of-day (EOD) diaries, and the day reconstruction method (DRM). To date, it is unclear whether they can be used interchangeably or whether ostensibly similar emotion dynamics captured by the methods differ in meaningful ways.

Objective: This study aims to examine the extent to which different ILA methods yield comparable measures of intraindividual emotion dynamics.

Methods: Data from 90 participants aged 50 years or older were collected in a probability-based internet panel, the Understanding America Study, and analyzed. Participants provided positive and negative affect ratings using 3 ILA methods: (1) smartphone-based EMA, administered 6 times per day over 1 week, (2) web-based EOD diaries, administered daily over the same week, and (3) web-based DRM, administered once during that week. We calculated 11 measures of emotion dynamics (addressing mean levels, variability, instability, and inertia separately for positive and negative affect, as well as emotion network density, mixed emotions, and emotional dialecticism) from each ILA method. The analyses examined mean differences and correlations of scores addressing the same emotion dynamic across the ILA methods. We also compared the patterns of intercorrelations among the emotion dynamics and their relationships with health outcomes (general health, pain, and fatigue) across ILA methods.

Results: Emotion dynamics derived from EMAs and EOD diaries demonstrated moderate-to-high correspondence for measures of mean emotion levels ($\rho \geq 0.95$), variability ($\rho \geq 0.68$), instability ($\rho \geq 0.51$), mixed emotions ($\rho = 0.92$), and emotional dialecticism ($\rho = 0.57$), and low correspondence for measures of inertia ($\rho \geq 0.17$) and emotion network density ($\rho = 0.36$). DRM-derived measures showed correlations with EMAs and EOD diaries that were high for mean emotion levels and mixed emotions ($\rho \geq 0.74$), moderate for variability ($\rho = 0.38$-$.054$), and low to moderate for other measures ($\rho = 0.03$-0.41). Intercorrelations among the emotion dynamics showed high convergence across EMAs and EOD diaries, and moderate convergence between the DRM and EMAs as well as EOD diaries. Emotion dynamics from all 3 ILA methods produced very similar patterns of relationships with health outcomes.

Conclusions: EMAs and EOD diaries provide corresponding information about individual differences in various emotion dynamics, whereas the DRM provides corresponding information about emotion levels and (to a lesser extent) variability, but not about more complex emotion dynamics. Our results caution researchers against viewing these ILA methods as universally interchangeable.

(J Med Internet Res 2020;22(9):e19201) doi:10.2196/19201

KEYWORDS

ecological momentary assessment; daily diaries; day reconstruction method; emotion dynamics; emotion variability; mobile phone

Introduction

The use of intensive longitudinal assessments (ILAs) in medical research has risen dramatically over the last few decades. In ILA studies, participants rate their experiences (eg, positive and negative affective states) repeatedly over time, often using electronic data collection via the internet or smartphones. The family of ILA methods encompasses real-time data collection as realized in ecological momentary assessments (EMAs) [1] and day-recall methods, including end-of-day (EOD) diaries [2] and the day reconstruction method (DRM) [3]. ILA methods offer several advantages compared with conventional types of assessment (eg, traditional questionnaires or clinical interviews). By inquiring about experiences that occurred over brief periods (eg, the past few minutes or the last day) in people's natural environments, ILA methods reduce recall bias and reliance on memory heuristics and can provide self-reports with improved ecological validity. In addition, the fine-grained data resulting from densely repeated assessments can be used to examine short-term, within-person processes that cannot be captured with traditional cross-sectional study designs [4,5].

ILA methods offer many novel insights into people's health and emotional functioning. For a long time, mental health research has predominantly focused on individuals' average levels of emotions. However, many aspects of people's emotional lives are not captured by how they feel on average. ILA adds a needed time dimension that allows assessment of the ebb and flow of subjective experiences and emotions. In many ways, ILA methods encourage a paradigm shift, changing the focus from emotions as static entities to studying them as dynamic processes [6]. For example, many mental health problems, including borderline personality disorder and bipolar disorder, are characterized by emotion dynamics, including increased intrapersonal *variability* or *instability* in affect [7-9]. Relatedly, the degree to which people's feelings are self-predictive or *linger* over time (described as emotional *inertia* or emotion *network density*) has been viewed as an important indicator of problems with successful emotion regulation [9-11] and as an important feature of mental health. Emotion dynamics may also be associated with physical health. The dynamic interplay of positive and negative affective states (the ability to experience positive and negative affective states in concert, *emotional dialecticism*, or *mixed emotions*) has been proposed as an indicator of emotional complexity and has been shown to benefit health outcomes [12,13].

An important advantage of ILAs is the ability to capture individual differences in emotion dynamics directly from repeated assessments. Researchers have enthusiastically embraced this possibility and have constructed measures of emotion dynamics from a range of ILA data sources, including EMAs, EOD diaries, and the DRM. EMAs have often been regarded as a *gold standard* among ILA methods in that respondents describe their momentary experiences as they are happening in real time (or close to real time) [1]. However, EMAs are relatively expensive to implement and burdensome for participants. EOD diaries, wherein respondents complete a single rating at the end of each day (often using a 24-hour recall period), represent a less costly and more practical alternative that is frequently used for the study of emotion dynamics [9]. The DRM offers yet another alternative that affords a granular assessment of emotional experiences over the course of a day [3,14]. This method has respondents first revive memories of (ie, reinstantiate) the previous day by asking them to divide the day into episodes; for each episode of the day, respondents then provide information about what they were doing and rate their emotional experiences. Compared with EOD diaries and EMAs, the DRM has found less widespread attention in research on emotion dynamics [13]. However, the fact that respondents complete the DRM in a single session (eg, administered over the internet) makes it an attractive method for this purpose because it can be readily implemented with large samples to address population-level research questions [3].

It is often implicitly assumed that data from different ILA methods can be used interchangeably to construct measures of emotion dynamics, even though the specific emotion processes captured by the methods differ in possibly meaningful ways (Table 1). For instance, whereas EMA and the DRM collect information based on relatively brief time intervals between assessments (over periods of hours), EOD diaries provide 1 assessment per day. It is not at all clear that emotion regulation processes occur in similar ways across these different timescales [6]. In addition, whereas EMA protocols typically capture emotion fluctuations that occur both within and across days, EOD diaries target between-day variation, and the DRM is limited to within-day variation (given that the DRM is typically administered only for a single day). The reasons for day-to-day fluctuations in emotions may be conceptually quite different from those generating intraday fluctuations. Furthermore, compared with EMAs, both EOD diary and DRM ratings require retrospection, which can introduce recall bias. Memory heuristics in EOD diary recall (eg, reporting the most salient or peak experiences) have been shown to distort estimates of people's average experience levels to some extent [15,16], and the assessment of dynamic aspects of emotions may be similarly (or even more) distorted by these memory processes. Another consideration is that a lower number of data points per person negatively impacts the reliability of measures of emotion dynamics [17]; administering the DRM for a single day (ie, the previous day) puts an upper limit on the number of episodes obtained per person, whereas the number of assessments in EMAs and EOD diaries is determined by the research design. In addition, in contrast to EMAs and EOD diaries, the DRM leaves the selection of the number and temporal spacing of episodes to the respondents. This self-selection of episodes may introduce potential selection biases if the experiences respondents choose to report systematically differ from those that are not reported.

Table 1. Comparison of ecological momentary assessments, end-of-day diaries, and day reconstruction method measurement features.

Measurement characteristics	Intensive longitudinal assessment methods		
	EMA[a]	EOD[b] diary	DRM[c]
Environmental or ecological context			
Temporal dynamics covered	Within-day and across multiple days	Across multiple days	Within a single day
Spacing between targeted experiences	Several minutes to several hours	1 day	Several minutes to several hours
Timing of completion of assessments	Completed in succession over the course of each day	Completed in succession at the end of each day	All assessments completed in a single session
Reliance on recall			
Targeted reporting period	Moments or very short time periods	Overall summary of all experiences of the day	Summary of experiences during episodes of the day
Participant recall required	Moment (working memory)	Same day (episodic memory)	Previous day (episodic memory), with reinstantiation
Potential for selection bias			
Selection of targeted experiences	Number and spacing of moments determined by prompting schedule	Fixed (daily)	Number and spacing of episodes determined by respondent
Reliability of measures of emotion dynamics			
Targeted number of measurement occasions	Dependent on prompting schedule (total of 42 in this study)	7 per week (total of 7 in this study)	Approximately 5 to 20 in total, determined by the respondent

[a]EMA: ecological momentary assessment.

[b]EOD: end-of-day.

[c]DRM: day reconstruction method.

Given that research on emotion dynamics derived from ILA is relatively new, researchers are using these measures without solid empirical knowledge about the impact that different ILA methods may have on their findings. To address this problem, this study aims to directly examine the extent to which common ILA methods produce comparable or dissimilar measures of intraindividual emotion dynamics. We used affect ratings collected using EMAs, EOD diaries, and DRM from the same individuals to derive many of the most commonly used measures of emotion dynamics from each method (including mean affect levels; affect variability, stability, and inertia; and emotion network density, emotional dialecticism, and mixed emotions) and examined the convergent validity of these measures across ILA methods. To address the question of whether the use of different ILA methods hampers the reproducibility of findings in health research, we further examined the extent to which measures of emotion dynamics constructed from each of the 3 methods demonstrate corresponding relationships with physical health outcomes (general health, pain, and fatigue). As the relationships between emotion dynamics and physical health are of particular interest in research on later points of the adult life span, this study focused on adults aged 50 years and older.

Methods

Participants and Procedures

The data for this study were collected as part of a larger project conducted in the Understanding America Study (UAS). The UAS is a probability-based internet panel maintained at the University of Southern California (USC) Center for Economic and Social Research [18]. It comprises about 8000 panel members, including about 3500 respondents aged 50 years or older. In contrast to convenience (*opt-in*) panels, where people self-select to participate, UAS panel members are recruited through nation-wide, address-based sampling. UAS respondents without previous internet access are equipped with a tablet and broadband internet. In 2017, the UAS started building a data collection environment to enable ILA research with a nationally representative sample [19]. As is typical for large-scale internet panels, UAS panelists answer various surveys on a regular basis, but their level of participation typically does not rise to the magnitude of involvement required for ILAs. The goal of the original study was to demonstrate the feasibility of implementing ILA methods in older participants in this national internet panel. The data for this study were collected between July 2017 and September 2018 during the pilot waves of this project.

The study was approved by the USC Institutional Review Board. UAS panelists aged 50 years or older who were using iOS or Android mobile devices were eligible to participate in the pilot waves. Respondents were screened for current smartphone usage for the purposes of EMA data collection (about 30% were deemed ineligible because they did not use iOS or Android mobile devices). Eligible respondents were provided with information about the project and asked if they were willing to participate. Participants were selected randomly among eligible respondents who consented to participate, with a consent rate

of 86% (112/130). As part of the study, participants were asked to complete EMAs and EOD diary questions for 7 consecutive days and to complete the DRM for one of these days (randomly selected). The questions included 1 item addressing positive affect (PA; happiness) and 1 item addressing negative affect (NA; sadness or dejection) that were used for the analyses (additional emotion questions, such as cheerful, frustrated, angry, lonely, or relaxed, were only administered to a small subset of respondents in the DRM and are therefore not used here).

EMAs

EMA data were collected on participants' own mobile phones with an app programmed using NubiS software. NubiS is an open-source, secure data collection, storage, and dissemination system developed by the Center of Economic and Social Research at USC. On each study day, respondents were prompted through cell phone beeps to complete EMA questions 6 times per day. Participants could specify the first and last possible prompt time for each day in the app before data collection, where the first prompt could be selected to occur between 6 AM and 11 AM and the last prompt between 7 PM and 11 PM of each day. Prompts were delivered using a stratified random sampling scheme that generated consecutive random prompts within the user-specified time window with a uniform probability between 0.75×(time window/6) and 1.10×(time window/6) hours (approximately between 1 and 3 hours) after the previous prompt. Each time the respondents received the prompt, they had 8 min to start answering questions. A reminder prompt was sent halfway during the 8-min time window if a participant did not immediately respond to an incoming prompt. Respondents were instructed not to respond to any prompt when they were driving and answer questions only if they were in a safe, secure, and private place. Each EMA survey included questions about location, social environment, and physical symptoms that were not analyzed here. The questions examined in this study were "before the prompt, how happy were you feeling" (PA), and "before the prompt, how dejected/blue/downhearted were you feeling" (NA). Responses were given on a 0 to 100 horizontal visual analog scale with anchors that ranged from *not at all* to *extreme*.

EOD Diaries

Respondents were asked to complete a web-based EOD diary survey in the evening (after 6 PM) of each of the 7 study days. Respondents completed the daily diaries using their laptops or tablets in their UAS account. The EOD diary survey questions paralleled those used in the EMA reports. The PA question was, "please move the slider to represent how happy you felt today," and the NA question was, "please move the slider to represent how dejected/blue/downhearted you felt today." Responses were given on a 0 to 100 horizontal visual analog scale with anchors that ranged from *not at all* to *extremely*.

DRM

The DRM was administered on the web once between days 2 and 7 of the study after participants had completed the daily diary questions. Participants were first asked when they woke up on the previous day and for how many hours they were awake. Next, they were asked to "think of yesterday as a series of scenes in a movie" and to divide the day into episodes. Starting at the time they woke up, participants entered a label (using open-ended text entry) for the first episode that best described what they did during that time, specified an ending time for the episode, and clicked *next* to move to the next episode. This process was repeated for subsequent episodes until participants reached the end of the day. Participants were told that many people define episodes that last between 15 min and 2 hours, but they were encouraged to define as many episodes as made sense to them and could specify episodes of any duration. If participants were awake for more than 24 hours, they were asked to enter episodes for the first 24 hours they were awake. After the complete day had been reconstructed, participants completed questions about where they were, whom they interacted with, and how they felt during each episode. Consistent with the rating scale format used in the original DRM [3], emotions were rated on a 7-point scale from 0=*not at all* to 6=*very much*. Participants were asked to "please rate each feeling on the scale given," and the emotions presented included *happy* (PA) and *sad* (NA). As the number of scale points used in the DRM differed from the number of scale points used in EMAs and EOD diaries, the DRM scores were transformed into a 0 to 100 scale using the following formula: transformed score=100×(original score+0.5)/7 [20].

Construction of Measures of Emotion Dynamics

The selection of measures of emotion dynamics was guided by previous reviews of the most commonly employed measures in applied research settings [9,11,21]. All measures were calculated in the same way for each ILA method based on each respondent's PA and NA ratings across moments (EMAs), days (EOD diaries), or episodes (DRM). To be included in the analyses, we required that a participant provide at least 4 observations for each of the ILA methods, which was deemed the minimum number of observations required to reasonably compute the various measures.

Mean PA and NA Levels

Measures of an individual's mean PA and NA levels represent the most well-known and prominent indicators of psychological well-being. For each ILA method, they were calculated by taking the average of each respondent's ratings across assessment time points, separately for PA and NA.

Variability of PA and NA

Emotion variability captures the magnitude (or amplitude) of fluctuations of a person's emotional states around the person's average emotion level [22]. Measures of variability were created by calculating the within-person SD of each respondent's PA and NA ratings across assessment time points.

Instability of PA and NA

Emotional instability measures differ from variability measures in that they explicitly consider the temporal ordering of affective states. Specifically, emotional instability refers to the magnitude of shifts in PA and NA levels across consecutive assessments [7]. The root mean square successive difference measure was used for this purpose, which was calculated by taking the

average of the squared differences between successive ratings and taking the square root of this average for each respondent.

Inertia of PA and NA

The concept of emotional inertia refers to the degree to which an individual's emotions are resistant to change, such that levels of PA or NA persistently carry on over time. Inertia is commonly operationalized using the first-order autocorrelation of consecutive measurements in a time series to capture the temporal dependency of a person's PA and NA ratings [10]. Correspondingly, we obtained measures of inertia in PA and NA as the person-specific autoregressive slope in regression models in which an affect rating at one time point predicts the rating of the same affect item at the subsequent measurement time point.

Emotion Network Density

The concept of emotion network density is an extension of the inertia concept to multiple affect items. It evaluates the degree to which multiple emotions predict each other over time, reflecting the extent to which a person's overall system of PA and NA states is resistant to change. Whereas measures of inertia are calculated separately for each affect item, network density combines the temporal dependencies of multiple affect items in a single measure [23,24]. For 2 emotion items, it comprises the sum of the absolute value of 4 lagged parameters in a vector autoregressive model involving 2 autoregressive and 2 cross-lagged parameters of PA and NA. Specifically, we obtained the person-specific (autoregressive and cross-lagged) parameters of PA and NA (entered simultaneously as within-person centered predictors) in regression models in which either PA or NA served as the outcome variable. The network density measure was then created by taking the absolute value of the 4 regression parameters and calculating the sum of these absolute values separately for each person.

Mixed Emotions

The experience of mixed (or bittersweet) emotions has been conceptualized as the extent to which both PA and NA are felt together at the same point in time. The construct of mixed emotions involves the simultaneous activation (or co-occurrence) of both positive and negative experiences [13]. Simply knowing whether an individual has high average levels of PA and NA over time may tell us nothing about whether they experienced a blending of PA and NA at any given point in time [25]. Following previous research, a measure of mixed emotions was calculated using the MIN index [13,25,26]. The index is based on the ambivalence metric proposed by Kaplan [27], which is defined as *total affect* (the sum of ratings for PA and NA) minus *polarity* (the absolute difference between ratings for PA and NA) for a given time point. Arithmetically, this formula is equivalent to taking (2 times) the smaller value of the PA and NA ratings (ie, MIN [PA, NA]) at a given time point [28], such that the MIN index is high only if both emotions co-occur at high levels at that time point. The MIN index was calculated for each time point and averaged across the assessment time points of each respondent.

Emotional Dialecticism

The concept of emotional dialecticism (the reverse of the concept of affective bipolarity) refers to the degree to which individuals tend to experience PA and NA independently from each other rather than as bipolar opposites [29,30]. Although conceptually similar to mixed emotions, the measure is calculated as the within-subject correlation between PA and NA ratings of each respondent [29]. As the correlation between PA and NA is expected to be negative on average, more strongly negative values (ie, values approaching a correlation of −1.0) on the measure represent less dialecticism, and less strongly negative values (ie, values approaching or exceeding a correlation of 0) represent more dialecticism.

Several additional considerations regarding the calculation of the measures of emotion dynamics are noteworthy. First, measures that take the temporal ordering into account (ie, measures of emotional instability, inertia, and emotion network density) require that the time intervals between 2 consecutive measurements are approximately equal. For this reason, when calculating these measures from EMA data, time periods from the evening of one day to the morning of the next day were omitted, as were momentary ratings that occurred after a time gap of more than 10 hours. Similarly, consecutive DRM ratings that were more than 10 hours apart from each other (measured from the midpoint of one episode to the midpoint of the next episode) were omitted when calculating these measures. Second, measures of emotion dynamics that involve lagged within-person relationships (ie, measures of inertia and emotion network density) are prone to high imprecision because of sampling error unless the number of measurement occasions is very large [17]. As the number of occasions in this study was relatively modest (especially for daily diaries), we used multilevel models to obtain empirical Bayes estimates of the person-specific regression parameters for calculating the inertia and network density measures. Compared with parameters from regression models that are estimated separately per respondent, the multilevel approach yields more precise parameter estimates for individual respondents [31].

Health Variables

Self-reported general health, pain, and fatigue were measured at the end of each of the 7 study days as part of the EOD diary assessments. For each health variable, the 7 scores were averaged into a summary measure for each person. Self-reported health was measured with the question "How was your health today?" with response options from the Short Form 36 general health item (excellent, very good, good, fair, or poor) [32]. To assess pain severity and fatigue levels, participants were asked to rate the following statements, *your average level of bodily pain* (no pain at all to extreme pain) and *how fatigued (ie, weary or tired) you were* (no fatigue at all to extreme fatigue), using a 0 to 100 horizontal visual analog scale.

Statistical Analysis

Data analyses were performed in 3 broad steps. The first step of the analysis examined the extent to which the 3 ILA methods (ie, EMAs, EOD diaries, and DRM) yielded comparable scores for each emotion dynamic. To accomplish this, we compared

the means and examined the correlations of measures addressing the same emotion dynamic across the ILA methods. Differences in mean scores were tested using analysis of variance (ANOVA) procedures with an ILA method as a within-person (ie, repeated measures) factor. An omnibus test of overall mean differences between the 3 methods was conducted first, followed by Bonferroni-adjusted pairwise tests of mean differences between the methods. For correlations of emotion dynamic measures between ILA methods, we roughly considered correlations of 0 to 0.35 as low, 0.36 to 0.67 as moderate, 0.68 to 0.89 as high, and 0.90 to 1.00 as indicating very high correspondence, following conventions [33]. To test whether the correlations differed between pairs of ILA methods (such that some ILA methods more highly correspond with each other than with other methods), we used Fisher z-transformed correlation coefficients and conducted Wald tests for differences in dependent correlations as implemented in Mplus version 8.4 (Muthén & Muthén) [34].

Whereas the first analysis step considered each measure of emotion dynamics separately, the second set of analyses examined how the measures are interconnected within each ILA method. Different measures of emotion dynamics may positively or negatively relate with each other in complex ways, and these relationships can be viewed to form a system or network of interdependencies among different features of emotional experience. Thus, evidence that the pattern of associations among the various measures corresponds across ILA methods would support the convergent validity of the methods. To examine this, we first inspected the structure of interconnections graphically using network visualization in the R package qgraph [35]. This technique represents a correlation matrix as a network in which each measure of emotion dynamics is represented as a node, and their interconnections are shown as edges between the nodes, allowing for a visual comparison of the correlation networks between ILA methods. To quantify the similarity of the correlation networks across ILA methods, we capitalized on centrality indices used in network analysis. In centrality analysis, the emotion dynamics are ordered in terms of the degree to which they occupy a central place in the overall network and exhibit many strong associations. We focused on the *simplest centrality metric*, *node strength*, which was calculated for each emotion dynamic as the sum of its absolute correlations with all other emotion dynamics in the network [36]. The ordering of the emotion dynamics' centralities was then descriptively compared across the ILA methods.

The third step examined the extent to which measures of emotion dynamics constructed from the different ILA methods demonstrate corresponding correlations with the health outcomes (general health, pain, and fatigue). To statistically compare the correlations of each emotion dynamic across the different ILA methods, we conducted Wald tests for differences in dependent Fisher r- to z-transformed correlation coefficients. In addition, effect sizes (Cohen *q*, where values of 0.1, 0.3, and 0.5 can be interpreted as small, medium, and large effects, respectively) were calculated to quantify the magnitude of differences in the correlations between ILA methods. *P* values <.05 were considered significant for all analyses.

Results

Descriptive Characteristics

A total of 100 participants completed assessments using all 3 ILA methods, with 10 participants being excluded from the analyses because they provided less than 4 observations for at least one of the ILA methods (1 participant had fewer than 4 observations for all ILA methods, 3 for EMA, and 6 for the DRM), resulting in an analysis sample of 90 participants. The mean age of the analyzed sample was 62.4 years (SD 7.7; range 51-87 years), 57% (51/90) were female, and 63% (57/90) were married. The sample was predominantly White (72/90, 80%) and non-Hispanic (83/90, 92%). The median household income was in the category between US $60,000 and US $74,999, and 57% (51/90) had a college degree. About half (42/90, 47%) were currently working, and about one-third (32/90, 36%) indicated that they were retired. Participants who were excluded from the analyses were less likely to hold a college degree (*P*=.04) but otherwise did not differ from the analysis sample on these demographics.

In terms of participants' compliance with the ILA protocol, the mean number of completed EMA prompts per person in the analyzed sample was 29.9 (SD 11.2; median 34) out of 42 possible ratings (6 per day across 7 days), yielding an average EMA completion rate of 71.2% (29.9/42). EOD diaries were, on average, completed on 6.9 (SD 0.5; median 7) out of the 7 days, yielding an average completion rate of 99.0% (6.9/7). For the DRM, the mean number of episodes provided per person was 11.8 (SD 5.7; median 11.0), which is roughly comparable with previous research using the DRM [3].

As the expected values of emotion dynamics that involve lagged within-person associations depend on the length of the lag (ie, autocorrelations are expected to be, on average, lower, the longer the lag time is), descriptive statistics for the time distances between ratings for each ILA method were also examined. For EMA prompts, the average distance between assessments (after elimination of overnight gaps and gaps >10 hours) was 2.13 hours (SD 0.89; median 1.94), with an IQR of 1.70 to 2.18 hours. The average time distance between consecutive EOD diaries was 24.01 hours (SD 2.16; median 24.05; IQR 23.32-24.76 hours). For the DRM, the average distance between the midpoints of consecutive episodes was 1.68 hours (SD 1.25; median 1.25; IQR 0.58-2.25 hours).

To evaluate the proportion of the total variance in affect that was accounted for by stable, trait-like differences as opposed to within-person fluctuations, intraclass correlations of PA and NA ratings were calculated for each ILA method, computed as the ratio of between-person variance to total (sum of within- and between-person) variance in ratings. The intraclass correlations were 0.63 (PA) and 0.54 (NA) for EMA reports; 0.57 (PA) and 0.53 (NA) for EOD reports; and 0.60 (PA) and 0.51 (NA) for DRM reports. Thus, between 37% (151.4/403.8) and 43% (175.6/410.9) of the variance in PA was within-person, and between 46% (142.6/311.2) and 49% (121.2/249.2) of the variance in NA was within-person, with consistency across the 3 ILA methods.

Correspondence of Measures of Emotion Dynamics Across the ILA Methods

ANOVA models yielded no significant differences between the ILA methods for measures capturing mean levels in PA and NA, PA variability, and PA instability, whereas significant method differences were evident for the remaining measures of emotion dynamics (Table 2). In pairwise comparisons between the ILA methods, the averages of measures from EMAs and EOD diaries did not significantly differ from each other, with one exception: the average inertia for NA was smaller for EOD diaries than for EMAs. This difference is consistent with expectations, given a longer lag time between EOD diary assessments compared with the lag times between EMAs. For DRM-derived measures, the averages were less comparable with those from the other ILA methods. The DRM yielded higher averages of inertia for PA and NA, and a higher emotion network density, compared with both EMAs and EOD diaries (consistent with the shorter time lags of the DRM). In addition, the DRM yielded lower average scores for NA variability and NA instability and higher scores for mixed emotions and emotional dialecticism compared with both EMAs and EOD diaries (Table 2).

The correlations between measures derived from EMAs and EOD diaries suggested very high correspondence (ρ>0.90) between these 2 methods for individuals' mean PA levels, mean NA levels, and mixed emotions, and high correspondence (ρ=0.68 and 0.80) for measures of PA and NA variability (Table 3). EMA- and EOD diary–derived measures further showed moderate-to-high correspondence (ρ ranging between 0.51 and 0.70) for PA and NA instability and emotional dialecticism measures, and low-to-moderate correspondence (ρ ranging between 0.17 and 0.57) for measures of PA and NA inertia and emotion network density. For most of the measures derived from the DRM, the correlations with both EMAs and EOD diaries were significantly weaker by comparison (Table 3). Specifically, the DRM showed high correspondence with EMAs and EOD diaries (ρ ranging between 0.74 and 0.80) for individuals' mean PA levels, mean NA levels, and mixed emotions, moderate correspondence (ρ between 0.38 and 0.54) for PA and NA variability measures, low-to-moderate correspondence (ρ between 0.19 and 0.41) for PA and NA instability measures, and low correspondence (ρ between 0.03 and 0.32) for the remaining measures (PA and NA inertia, emotion network density, and emotional dialecticism).

Table 2. Mean (SD) of measures of emotion dynamics for each assessment method.

Emotion dynamic	Assessment method, mean (SD)			ANOVA[a] of mean differences	
	EMA[b]	EOD[c] diary	DRM[d]	*F* value (*df*=2178)	*P* value
PA[e] mean level	66.24 (16.1)	68.40 (16.1)	66.79 (17.6)	2.61	.08
NA[f] mean level	12.60 (13.3)	12.12 (14.7)	14.00 (11.6)	2.73	.07
PA variability	11.29 (4.6)	11.65 (6.7)	11.93 (7.8)	0.42	.66
NA variability	9.05[g] (7.3)	9.02[g] (9.3)	5.79[h] (8.5)	10.14	<.001
PA instability	12.92 (6.3)	14.51 (9.1)	13.53 (9.6)	1.16	.32
NA instability	10.41[g] (8.8)	11.08[g] (11.6)	6.58[h] (9.8)	10.13	<.001
PA inertia	0.41[g] (0.1)	0.38[g] (0.1)	0.46[h] (0.2)	8.72	<.001
NA inertia	0.34[g] (0.1)	0.26[h] (0.1)	0.42[i](0.2)	32.49	<.001
Network density	0.86[g] (0.2)	0.91[g] (0.2)	0.99[h] (0.3)	10.23	<.001
Mixed emotions	10.90 (10.6)	10.03[g] (11.6)	12.42[h](9.5)	6.24	.002
Dialecticism	−0.34[g] (0.3)	−0.36[g] (0.2)	−0.21[h] (0.2)	13.14	<.001

[a]ANOVA: analysis of variance.

[b]EMA: ecological momentary assessment.

[c]EOD: end-of-day.

[d]DRM: day reconstruction method.

[e]PA: positive affect.

[f]NA: negative affect.

[g,h,i]Means in the same row with different superscripts differ significantly from each other using Bonferroni-adjusted pairwise comparisons and means in the same row without superscripts or that share the same superscript do not differ.

Table 3. Correlations of emotion dynamics across assessment methods.

Emotion dynamic	Correlation coefficients			Difference between correlations	
	EMA[a]-EOD[b] diary	EMA-DRM[c]	EOD diary-DRM	χ^2 (*df*=2)	*P* value
PA[d] mean level	0.95[e]	0.80[f]	0.79[f]	50.52	<.001
NA[g] mean level	0.95[e]	0.78[f]	0.74[f]	70.68	<.001
PA variability	0.68[e]	0.54	0.38[f]	15.46	<.001
NA variability	0.80[e]	0.47[f]	0.45[f]	25.68	<.001
PA instability	0.51[e]	0.33	.019[f]	9.05	.01
NA instability	0.70[e]	0.37[f]	0.41[f]	14.99	<.001
PA inertia	0.46[e]	0.18	.017[f]	6.43	.04
NA inertia	0.17	0.26	0.03	3.33	.19
Network density	0.36	0.32	0.26	0.77	.68
Mixed emotions	0.92[e]	0.76[f]	0.76[f]	30.35	<.001
Dialecticism	0.57[e]	0.17[f]	0.31	14.25	<.001

[a]EMA: ecological momentary assessment.

[b]EOD: end-of-day.

[c]DRM: day reconstruction method.

[d]PA: positive affect.

[e,f]Correlations in the same row with different superscripts differ significantly from each other using Bonferroni-adjusted pairwise comparisons and means in the same row without superscripts or that share the same superscript do not differ. Correlations >0.21 are significant at *P*<.05, correlations >0.28 are significant at *P*<01, and correlations >0.35 are significant at *P*<.001.

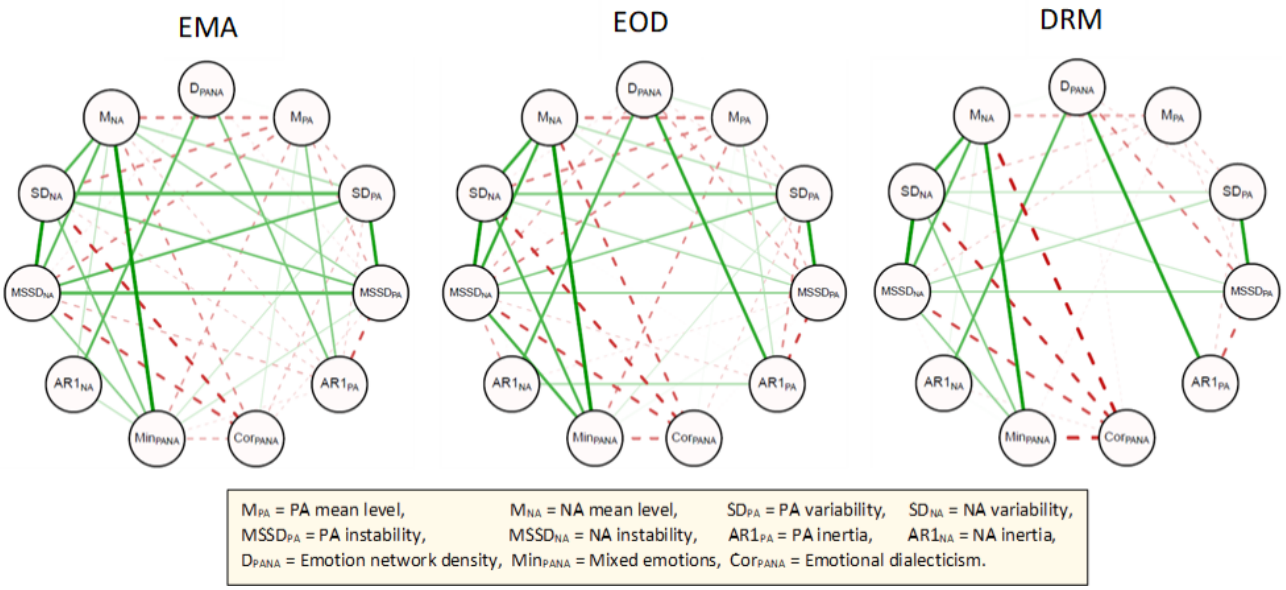
[g]NA: negative affect.

The goal of the primary analyses was to examine the comparability of measures that were computed in the way they would most likely be computed in applied research, that is, based on all available data for each ILA method. However, given that the DRM captures only 1 single day, an interesting question is whether the correspondence between EMA-based and DRM-based measures increases when measures of emotion dynamics are derived from EMAs for the exact day before the DRM was completed (ie, the exact day respondents were asked to rate in the DRM). Secondary analyses were conducted to address this question (based on 71 participants who had at least 3 EMA reports for the day before the DRM). The pattern of correlations with DRM-based measures was highly similar when measures of emotion dynamics were derived from all EMA data versus same-day EMA data (the median Cohen *q* for the difference in correlations was −0.014).

Patterns of Associations Among Emotion Dynamics Within Each ILA Method

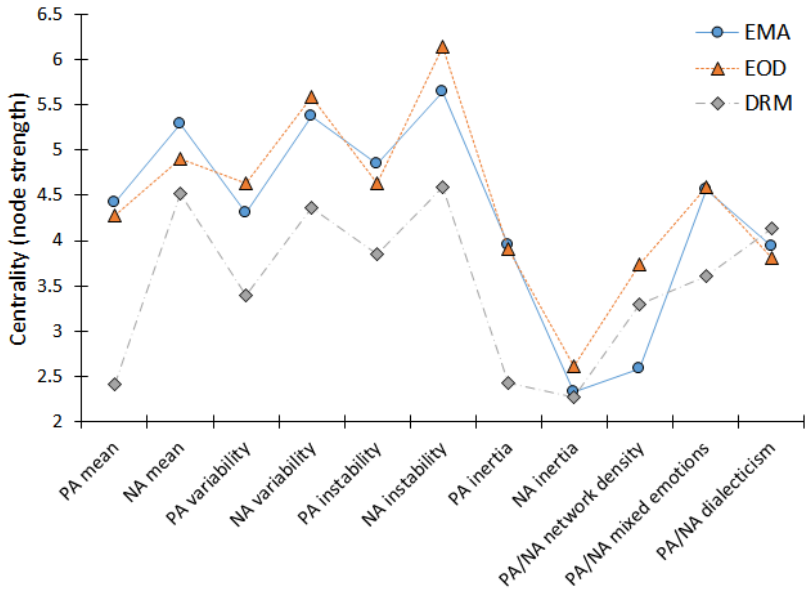
The network of pairwise associations among the various emotion dynamics is illustrated in Figure 1 for each ILA method. Examining the network of EMA-derived measures, several patterns are noteworthy. First, moderate-to-strong positive connections were evident among all variability and instability measures (ρ≥0.67, indicated by thick green lines). Second, the measure of mean NA levels showed positive connections with NA variability and NA instability and mixed emotions (ρ≥0.67). Third, although measures of mean PA levels and emotional dialecticism were weakly correlated with each other (ρ=0.33), both showed moderate-to-strong negative connections with mean NA levels, NA variability, and NA instability (ρ≤−0.54, indicated by thick red lines). Finally, emotion network density was positively associated with PA inertia and NA inertia (ρ≥0.59), whereas all 3 measures showed otherwise few connections in the network. Visual comparison of the correlation networks suggests that the patterns of association were similar for measures derived from EMAs and EOD diaries. However, DRM-derived measures generally showed fewer and weaker interconnections (Figure 1).

Figure 1. The network of pairwise correlations among emotion dynamics for each assessment method. Green lines represent positive relations, and red dashed lines represent negative relations. The thickness and transparency of the lines correspond with the magnitude of correlations. For clarity, lines are shown only for correlations that surpass the $P<.05$ significance threshold (above an absolute value of 0.21). EMA: ecological momentary assessment; EOD: end-of-day; DRM: day reconstruction method; PA: positive affect; NA: negative affect.



The magnitude of the interconnections is quantified in the node strength centralities of the measures (Figure 2). For emotion measures derived from EMA, those tapping emotion levels, variability, and instability occupied the most central places with the strongest interconnections; measures for NA showed consistently greater centralities than corresponding measures for PA. As can be seen in Figure 2, the ordering of node strengths corresponded very closely between EMA measures and those derived from EOD diaries, with a correlation of ρ=0.92 between these 2 ILA methods. The node strength centralities of DRM-derived measures were generally lower (corresponding with weaker interconnections among DRM-derived measures), and the ordering of node strengths showed moderate-to-high correspondence with those from EMAs (ρ=0.75) and EOD diaries (ρ=0.70).

Figure 2. Centrality (node strength) of emotion dynamics for each assessment method. EMA: ecological momentary assessment; EOD: end-of-day; DRM: day reconstruction method; PA: positive affect; NA: negative affect.



Relationships With Health Outcomes

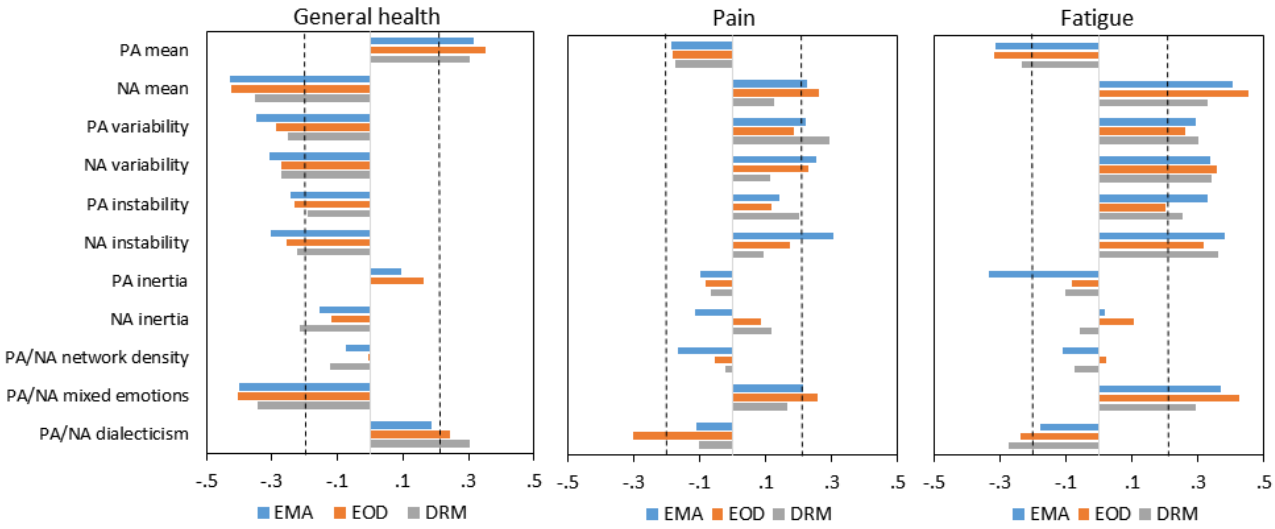
Finally, relationships between the measures of emotion dynamics and health outcome variables were examined. Overall, the emotion dynamics showed small- to medium-sized correlations with self-reports of general health and fatigue, and somewhat less pronounced correlations with pain (Figure 3). Comparisons of the correlation coefficients between ILA methods yielded very few differences: out of 33 Wald tests (for 11 emotion dynamics and 3 health outcomes), only 1 indicated a significant difference between the ILA methods (Wald

χ^2_2=6.19, *P*=.04 for the correlations between PA inertia and fatigue; post hoc pairwise comparisons showed that the correlation was significantly more negative for EMA than for EOD diary reports, z=−2.38, *P*=.02). Average effect sizes (absolute values of Cohen *q*) for the differences in correlations between the ILA methods were very small for EMAs versus EOD diaries (general health, mean absolute *q*=0.04; pain, mean absolute *q*=0.08; and fatigue, mean absolute *q*=0.09), EMAs versus DRM (general health, mean absolute *q*=0.07; pain, mean absolute *q*=0.10; and fatigue mean absolute *q*=0.08), and for EOD diaries versus DRM (general health, mean absolute *q*=0.07; pain, mean absolute *q*=0.09; and fatigue, mean absolute *q*=0.08).

As shown in Figure 3, for all ILA methods, the most pronounced relationships with health outcomes were evident for mean emotion levels, variability, and instability measures in expected directions. Higher mean PA levels were consistently correlated with better general health, less pain, and less fatigue, whereas higher mean NA, variability (PA and NA), and instability (PA and NA) correlated with poorer general health, more pain, and more fatigue. Higher values of emotional dialecticism were associated with better health outcomes, whereas, contrary to theoretical expectations [37], more mixed emotions were associated with worse health outcomes. Finally, inertia (PA and NA) and emotion network density measures showed almost no relationship with the health outcomes.

We also explored whether measures of emotion dynamics demonstrated incremental validity in predicting the health outcomes above and beyond average emotion levels. To examine this, a series of multiple regression analyses were conducted in which the health outcomes were regressed on a measure of emotion dynamics after controlling for PA and NA mean levels separately for each ILA method. In these models, the only significant predictor of general health was the mean level of NA, consistently for each ILA method (*P*<.05 in all instances). For pain as an outcome variable, no emotion measure (including mean PA and NA) was a significant predictor in the multiple regressions; the exception was that PA variability uniquely predicted pain in the DRM (*P*=.02). For fatigue, the only significant predictor was the mean level of NA, again consistently for each ILA method (*P*<.05 in all instances). In addition, PA variability uniquely predicted fatigue in the DRM (*P*=.03).

Figure 3. Correlations of measures of emotion dynamics with self-reports of general health, pain, and fatigue for each assessment method. Horizontal bars represent the magnitude and direction of correlation coefficients. Correlations exceeding a value of 0.21 (indicated by vertical dashed lines) are significant at *P*<.05. EMA: ecological momentary assessment; EOD: end-of-day; DRM: day reconstruction method; PA: positive affect; NA: negative affect.



Discussion

Emotions fluctuate and change over time, and ILA methods are uniquely suited to quantify the temporal dynamics of people's emotional lives. There has been a surge of interest in creating measures that tap a variety of emotion dynamics from ILA. To our knowledge, this is the first study to directly compare EMA, EOD diary, and DRM ILA methods in the measurement of intraindividual emotion dynamics. If different ILA methods produce noncorresponding measures of emotion dynamics, this would have important research implications in that it would question the validity of the measures and threaten the reproducibility of empirical research results.

EMAs and EOD diaries are arguably the 2 most commonly used ILA methods [4]. Previous research has documented that individual differences in mean levels of positive and negative experiences are highly correlated between these methods [38,39]. This finding was confirmed in this study. Expanding on previous research, we found that several measures of emotion dynamics derived from EOD diaries also reproduced those derived from EMAs very well. Specifically, measures of emotion variability, instability, mixed emotions, and emotional dialecticism showed substantial correspondence between the 2 methods, with comparable mean levels and moderate-to-high correlations. Considering the argument that momentary and daily fluctuations in emotions may be quite different conceptually, the level of agreement between the methods is perhaps somewhat surprising. For example, given that EMAs encompass the sum of within-day and between-day sources of intraindividual variation, whereas EOD diary ratings are limited to between-day variation, one might expect the average

variability to be higher in EMAs than in EOD diaries, which we did not find. Our results suggest that differences in the time scale and frequency of measurement inherent in EMA and EOD diary ratings do not dramatically impact measures of emotion variability, instability, mixed emotions, and emotional dialecticism.

However, this does not mean that EMA and EOD diary measures can be viewed as universally interchangeable. Measures of inertia and emotion network density were correlated at levels below 0.50 between the methods. In contrast to the other measures of emotion dynamics, inertia and emotion network density measures are specifically focused on temporal dependencies between successive self-report ratings; that is, they capture the rate (or speed) of changes rather than the magnitude of changes. This suggests that the measurement of how emotions evolve over time and at what rate may not be captured in the same way by EOD diaries and EMAs.

The DRM was originally developed as an alternative to EMAs for use in large and population-representative samples, yet few previous studies have directly compared the information obtained from the DRM with that obtained from EMAs [40-42] or EOD diaries [43,44]. In this study, the DRM demonstrated high correspondence with other ILA methods for measures of average emotion levels and mixed emotions and moderate correspondence for emotion variability. For more complex emotion dynamics, the DRM showed low correlations with EMAs and EOD diaries. Similarly, the networks of pairwise associations among the various emotion dynamics were only moderately concordant between the DRM and the other ILA methods. This suggests that the DRM may adequately capture people's average emotion levels (and, to some extent, their emotion variability) but may not serve as a direct replacement of other ILA methods in research on more intricate emotion dynamics.

Interestingly, all ILA methods produced very similar patterns of relationships with health outcomes. Higher mean PA and lower mean NA levels were associated with better physical health outcomes with medium (general health and fatigue) and small-to-medium (pain) effect sizes, replicating previous research on these mind-body relationships [45-47]. The consistency of these results across ILA methods suggests that each of them can equally contribute to understanding the linkages between people's general emotion levels and health outcomes in older adulthood.

Measures of emotion variability and more emotional instability derived from each ILA method were also consistently related to health outcomes in expected directions. These results are noteworthy considering that even though the last decades have witnessed a surge in research linking emotion variability with maladaptive outcomes, most of this research has focused on psychological well-being (eg, depression, anxiety) rather than physical health outcomes. In a large meta-analysis, Houben et al [9] found small-to-medium effect sizes for relationships between psychological well-being and measures of emotion variability (ρ=0.18) and instability (ρ=0.21). Our results suggest effect sizes of similar magnitude for physical health outcomes.

Measures of mixed emotions (MIN index) and emotional dialecticism (within-person correlation of PA and NA) were moderately negatively correlated with each other and showed opposite relationships with the health outcomes within each of the ILA methods. This finding would appear counterintuitive, given that the 2 measures aim at capturing conceptually similar concepts. However, previous studies have found a generally weak correspondence between these indices of mixed emotions and emotional dialecticism [13,26], suggesting that they measure different aspects of emotional experience. Consistent with our results, higher emotional dialecticism scores have previously been shown to predict fewer health symptoms [12], whereas mixed emotions assessed with the MIN index were associated with more physical disability across adulthood [13].

Measures that focus on temporal dependencies of emotional states (emotional inertia and emotion network density) were practically uncorrelated with the physical health outcomes in this study, contrary to previous meta-analytic findings that higher emotional inertia relates to worse psychological well-being, albeit with overall small effect sizes [9]. It is possible that temporal dependency measures play a lesser role in understanding physical health compared with mental health. However, previous simulation studies have also demonstrated that measures of inertia derived from ILAs tend to have very low reliability, and this may have attenuated the observed correlations with health outcomes. A recent Monte Carlo simulation by Du and Wang [17] suggests that even if the emotional states themselves are assessed with near-perfect reliability, up to 100 measurement occasions per person may be necessary to obtain reliabilities >0.70 for inertia measures, whereas individual differences in mean levels, variability, and instability measurement required substantially fewer measurement occasions to be reliably measured in the simulation study. Future research would benefit from implementing ILA methods across multiple waves (eg, using measurement bursts) as a means to estimate the test-retest reliability of measures of emotion dynamics and to correct their correlations with health outcomes for unreliability in empirical samples.

It is also noteworthy that once the explanatory power of mean levels of PA and NA was taken into account, measures of emotion dynamics showed little to no added value in the prediction of the health outcomes, regardless of the ILA method. This finding corresponds with recent findings by Dejonckheere et al [21], suggesting that more complex emotion dynamics add little to the prediction of psychological well-being and emotion disorders after controlling for mean affect levels. Our findings suggest that caution is similarly warranted when assuming that measures of complex emotion dynamics have unique predictive utility for understanding physical health parameters, even though this would need to be confirmed in larger studies.

Several limitations of this study need to be considered. First, the study was restricted to older individuals aged 50 years and above, and the results may not be generalizable to younger adults. However, evidence for convergent validity across ILA methods in this age group may be particularly important given that older people may have more problems with electronic ILA data collection and that potential cognitive problems in this age group may interfere with accurate self-reports [48]. Second, the

sample size of 90 respondents was relatively modest, although it was comparable with previous studies examining the correspondence of EMAs with EOD diaries [38] or the DRM [40,41]. A third limitation is that whereas EMAs and EOD diaries used the same wording of affect items (ie, happy, dejected/blue/downhearted) and the same response scale (a 0-100 visual analog scale), the DRM used a partially different wording (ie, happy, sad) and different response scale (a 7-point numeric response scale) to collect affect ratings. This may have artificially deflated the correspondence between measures of emotion dynamics derived from the DRM and the other ILA methods. Holding the items and response scales constant across ILA methods in future studies would enhance the rigor of comparisons by minimizing the potential impact of such extraneous factors. Fourth, it is also possible that completing multiple EMA ratings throughout the day impacted participants' EOD diary and DRM affect ratings, which may have artificially inflated the correspondence between the different measures. Previous literature found little evidence that recall ratings are impacted by momentary reporting [49], but we did not specifically examine this possibility here. Finally, even though we examined a variety of measures of emotion dynamics, the study was limited to measures that can be derived from 2 affect questions. Additional measures of emotion dynamics that have been proposed in the literature were not considered here because they require administration of multiple PA or NA items at each measurement occasion (examples are emotional granularity, which captures the extent to which individuals differentiate between multiple emotions of the same valence [50], and emodiversity, which captures the extent to which individuals experience a narrow or wide range of different emotions [51]).

In summary, EMAs and EOD diaries correspond moderately to highly with each other in the information they provide about individual differences in various emotion dynamics. Compared with these ILA methods, the DRM provides corresponding information about emotion levels and (to a lesser extent) variability, but not about more complex emotion dynamics. Our results caution researchers against viewing these ILA methods as universally interchangeable. Although measures of emotion dynamics derived from all ILA methods showed small-to-moderate relationships with physical health outcomes, the unique predictive ability of more complex emotion dynamics for understanding health outcomes remains to be established.

Acknowledgments

This study was supported by grants from the National Institute on Aging (R21 AG061364 and U01 AG054580) and the Social Security Administration. The sponsors were not involved in the conduct of this study or the preparation of this paper.

Conflicts of Interest

AS is a senior scientist with the Gallup Organization and a consultant with Adelphi Values. All other authors declare no conflicts of interest.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. Annu Rev Psychol 2003;54:579-616. [doi: 10.1146/annurev.psych.54.101601.145030] [Medline: 12499517]
3. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004 Dec 2;306(5702):1776-1780. [doi: 10.1126/science.1103572] [Medline: 15576620]
4. Schneider S, Stone AA. Ambulatory and diary methods can facilitate the measurement of patient-reported outcomes. Qual Life Res 2016 Mar;25(3):497-506 [FREE Full text] [doi: 10.1007/s11136-015-1054-z] [Medline: 26101141]
5. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. Proc Natl Acad Sci U S A 2018 Jul 3;115(27):E6106-E6115 [FREE Full text] [doi: 10.1073/pnas.1711978115] [Medline: 29915059]
6. Hamaker EL, Wichers M. No time like the present. Curr Dir Psychol Sci 2017 Feb 8;26(1):10-15. [doi: 10.1177/0963721416666518]
7. Trull TJ, Solhan MB, Tragesser SL, Jahng S, Wood PK, Piasecki TM, et al. Affective instability: measuring a core feature of borderline personality disorder with ecological momentary assessment. J Abnorm Psychol 2008 Aug;117(3):647-661. [doi: 10.1037/a0012532] [Medline: 18729616]
8. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
9. Houben M, van den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: a meta-analysis. Psychol Bull 2015 Jul;141(4):901-930. [doi: 10.1037/a0038822] [Medline: 25822133]
10. Kuppens P, Allen NB, Sheeber LB. Emotional inertia and psychological maladjustment. Psychol Sci 2010 Jul;21(7):984-991 [FREE Full text] [doi: 10.1177/0956797610372634] [Medline: 20501521]
11. Kuppens P, Verduyn P. Emotion dynamics. Curr Opin Psychol 2017 Oct;17:22-26. [doi: 10.1016/j.copsyc.2017.06.004] [Medline: 28950968]

12. Hershfield HE, Scheibe S, Sims TL, Carstensen LL. When feeling bad can be good: mixed emotions benefit physical health across adulthood. Soc Psychol Personal Sci 2013 Jan;4(1):54-61 [FREE Full text] [doi: 10.1177/1948550612444616] [Medline: 24032072]
13. Schneider S, Stone AA. Mixed emotions across the adult life span in the United States. Psychol Aging 2015 Jun;30(2):369-382 [FREE Full text] [doi: 10.1037/pag0000018] [Medline: 25894487]
14. Stone AA, Schwartz JE, Schkade D, Schwarz N, Krueger A, Kahneman D. A population approach to the study of emotion: diurnal rhythms of a working day examined with the day reconstruction method. Emotion 2006 Feb;6(1):139-149. [doi: 10.1037/1528-3542.6.1.139] [Medline: 16637757]
15. Neubauer AB, Scott SB, Sliwinski MJ, Smyth JM. How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. J Pers Soc Psychol 2020 Jul;119(1):185-203. [doi: 10.1037/pspp0000248] [Medline: 31070397]
16. Schneider S, Stone AA, Schwartz JE, Broderick JE. Peak and end effects in patients' daily recall of pain and fatigue: a within-subjects analysis. J Pain 2011 Feb;12(2):228-235 [FREE Full text] [doi: 10.1016/j.jpain.2010.07.001] [Medline: 20817615]
17. Du H, Wang L. Reliabilities of intraindividual variability indicators with autocorrelated longitudinal data: implications for longitudinal study designs. Multivariate Behav Res 2018;53(4):502-520. [doi: 10.1080/00273171.2018.1457939] [Medline: 29683725]
18. Flint RF. An introduction to the understanding america study internet panel. Soc Sec Bull 1949;60(9):1305. [doi: 10.1130/0016-7606(1949)60[1305:i]2.0.co;2]
19. Kapteyn A. Measurement and experimentation in the social sciences. SSRN J 2010:-. [doi: 10.2139/ssrn.1889312]
20. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019 Feb 5;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
21. Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, et al. Complex affect dynamics add limited information to the prediction of psychological well-being. Nat Hum Behav 2019 May;3(5):478-491. [doi: 10.1038/s41562-019-0555-0] [Medline: 30988484]
22. Eid M, Diener E. Intraindividual variability in affect: reliability, validity, and personality correlates. J Pers Soc Psychol 1999 Apr;76(4):662-676. [doi: 10.1037/0022-3514.76.4.662]
23. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, et al. A network approach to psychopathology: new insights into clinical longitudinal data. PLoS One 2013;8(4):e60188 [FREE Full text] [doi: 10.1371/journal.pone.0060188] [Medline: 23593171]
24. Lydon-Staley DM, Xia M, Mak HW, Fosco GM. Adolescent emotion network dynamics in daily life and implications for depression. J Abnorm Child Psychol 2019 Apr;47(4):717-729 [FREE Full text] [doi: 10.1007/s10802-018-0474-y] [Medline: 30203118]
25. Schimmack U. Pleasure, displeasure, and mixed feelings: are semantic opposites mutually exclusive? Cognit Emot 2001 Jan;15(1):81-97. [doi: 10.1080/02699930126097]
26. Larsen JT, Hershfield HE, Stastny BJ, Hester N. On the relationship between positive and negative affect: their correlation and their co-occurrence. Emotion 2017 Mar;17(2):323-336. [doi: 10.1037/emo0000231] [Medline: 27709977]
27. Kaplan KJ. On the ambivalence-indifference problem in attitude theory and measurement: a suggested modification of the semantic differential technique. Psychol Bull 1972;77(5):361-372. [doi: 10.1037/h0032590]
28. Priester JR, Petty RE. The gradual threshold model of ambivalence: relating the positive and negative bases of attitudes to subjective ambivalence. J Pers Soc Psychol 1996;71(3):431-449. [doi: 10.1037/0022-3514.71.3.431] [Medline: 8831157]
29. Dejonckheere E, Mestdagh M, Houben M, Erbas Y, Pe M, Koval P, et al. The bipolarity of affect and depressive symptoms. J Pers Soc Psychol 2018 Feb;114(2):323-341. [doi: 10.1037/pspp0000186] [Medline: 29389216]
30. Lindquist K, Barrett L. Emotional complexity. In: Lewis M, Haviland-Jones JM, Barrett LF, editors. Handbook of Emotions. Third Edition. New York, USA: Guilford Press; 2008:513-530.
31. Singer J, Willett J. Applied Longitudinal Data Analysis: Modeling Change and Event Occurence. Oxford, UK: Oxford University Press; 2003.
32. Ware J. SF-36 Health Survey: Manual & Interpretation Guide. Boston, UK: The Health Institute, New England Medical Center; 1993.
33. Mason R, Lind D, Marchal W. Statistics: An Introduction. New York, USA: Harcourt Brace Jovanovich Inc; 1983.
34. Muthén L, Muthén B. Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8). Los Angeles, CA: Muthén & Muthén; 2017.
35. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Network visualizations of relationships in psychometric data. J Stat Soft 2012;48(4):-. [doi: 10.18637/jss.v048.i04]
36. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Natl Acad Sci U S A 2004 Mar 16;101(11):3747-3752 [FREE Full text] [doi: 10.1073/pnas.0400087101] [Medline: 15007165]

37. Larsen J, Hemenover S, Norris C, Cacioppo J. Turning adversity to advantage: on the virtues of the coactivation of positivenegative emotions. In: Aspinwall LG, Staudinger UM, editors. A Psychology of Human Strengths: Fundamental Questions and Future. Washington, DC: American Psychological Association; 2003.
38. Broderick JE, Schwartz JE, Schneider S, Stone AA. Can end-of-day reports replace momentary assessment of pain and fatigue? J Pain 2009 Mar;10(3):274-281 [FREE Full text] [doi: 10.1016/j.jpain.2008.09.003] [Medline: 19070550]
39. Diener E, Larsen RJ, Levine S, Emmons RA. Intensity and frequency: dimensions underlying positive and negative affect. J Pers Soc Psychol 1985;48(5):1253-1265. [doi: 10.1037/0022-3514.48.5.1253]
40. Dockray S, Grant N, Stone AA, Kahneman D, Wardle J, Steptoe A. A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. Soc Indic Res 2010 Nov;99(2):269-283 [FREE Full text] [doi: 10.1007/s11205-010-9578-7] [Medline: 21113328]
41. Kim J, Kikuchi H, Yamamoto Y. Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. Br J Health Psychol 2013 Feb;18(1):155-167. [doi: 10.1111/bjhp.12000] [Medline: 23017062]
42. Han W, Feng X, Zhang M, Peng K, Zhang D. Mood states and everyday creativity: employing an experience sampling method and a day reconstruction method. Front Psychol 2019;10:1698 [FREE Full text] [doi: 10.3389/fpsyg.2019.01698] [Medline: 31379699]
43. Christodoulou C, Schneider S, Stone AA. Validation of a brief yesterday measure of hedonic well-being and daily activities: comparison with the day reconstruction method. Soc Indic Res 2014 Feb 1;115(3):907-917 [FREE Full text] [doi: 10.1007/s11205-013-0240-z] [Medline: 24683288]
44. Miron-Shatz T, Stone A, Kahneman D. Memories of yesterday's emotions: does the valence of experience affect the memory-experience gap? Emotion 2009 Dec;9(6):885-891. [doi: 10.1037/a0017823] [Medline: 20001131]
45. Steptoe A, Wardle J. Positive affect measured using ecological momentary assessment and survival in older men and women. Proc Natl Acad Sci U S A 2011 Nov 8;108(45):18244-18248 [FREE Full text] [doi: 10.1073/pnas.1110892108] [Medline: 22042845]
46. Conner TS, Barrett LF. Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. Psychosom Med 2012 May;74(4):327-337 [FREE Full text] [doi: 10.1097/PSY.0b013e3182546f18] [Medline: 22582330]
47. Pressman SD, Jenkins BN, Moskowitz JT. Positive affect and health: what do we know and where next should we go? Annu Rev Psychol 2019 Jan 4;70:627-650. [doi: 10.1146/annurev-psych-010418-102955] [Medline: 30260746]
48. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. J Psychiatr Res 2009 Jul;43(11):987-996 [FREE Full text] [doi: 10.1016/j.jpsychires.2009.01.014] [Medline: 19272611]
49. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 2003 Jul;104(1-2):343-351. [doi: 10.1016/s0304-3959(03)00040-x] [Medline: 12855344]
50. Kashdan TB, Barrett LF, McKnight PE. Unpacking emotion differentiation. Curr Dir Psychol Sci 2015 Feb 18;24(1):10-16. [doi: 10.1177/0963721414550708]
51. Quoidbach J, Gruber J, Mikolajczak M, Kogan A, Kotsou I, Norton MI. Emodiversity and the emotional ecosystem. J Exp Psychol Gen 2014 Dec;143(6):2057-2066. [doi: 10.1037/a0038025] [Medline: 25285428]

Abbreviations

ANOVA: analysis of variance
DRM: day reconstruction method
EMA: ecological momentary assessment
EOD: end-of-day
ILA: intensive longitudinal assessment
NA: negative affect
PA: positive affect
UAS: Understanding America Study
USC: University of Southern California

Edited by G Eysenbach; submitted 09.04.20; peer-reviewed by M Houben, E Granholm, C Simons; comments to author 04.07.20; revised version received 15.08.20; accepted 01.09.20; published 24.09.20.

Please cite as:
Schneider S, Junghaenel DU, Gutsche T, Mak HW, Stone AA
Comparability of Emotion Dynamics Derived From Ecological Momentary Assessments, Daily Diaries, and the Day Reconstruction Method: Observational Study
J Med Internet Res 2020;22(9):e19201
URL: http://www.jmir.org/2020/9/e19201/
doi:10.2196/19201
PMID:32969835

Original Paper

Exploring Types of Information Sources Used When Choosing Doctors: Observational Study in an Online Health Care Community

Shuang Zhang[1*], PhD; Jying-Nan Wang[1,2*], PhD; Ya-Ling Chiu[2,3*], PhD; Yuan-Teng Hsu[4*], PhD

[1]College of International Finance and Trade, Zhejiang Yuexiu University of Foreign Languages, Zhejiang, China

[2]Research Institute for Modern Economics and Management, Zhejiang Yuexiu University of Foreign Languages, Zhejiang, China

[3]College of International Business, Zhejiang Yuexiu University of Foreign Languages, Zhejiang, China

[4]Research Center of Finance, Shanghai Business School, Shanghai, China

[*]all authors contributed equally

Corresponding Author:
Yuan-Teng Hsu, PhD
Research Center of Finance
Shanghai Business School
No 2271 West Zhong Shan Rd
Shanghai, 200235
China
Phone: 86 21 64870020 ext 1404
Email: yuanteng.hsu@gmail.com

Abstract

Background: Patients attempt to make appropriate decisions based on their own knowledge when choosing a doctor. In this process, the first question usually faced is that of how to obtain useful and relevant information. This study investigated the types of information sources that are used widely by patients in choosing a doctor and identified ways in which the preferred sources differ in various situations.

Objective: This study aims to address the following questions: (1) What is the proportion in which each of the various information sources is used? (2) How does the information source preferred by patients in choosing a doctor change when there is a difference in the difficulty of medical decision making, in the level of the hospital, or in a rural versus urban situation? (3) How do information sources used by patients differ when they choose doctors with different specialties?

Methods: This study overcomes a major limitation in the use of the survey technique by employing data from the Good Doctor website, which is now China's leading online health care community, data which are objective and can be obtained relatively easily and frequently. Multinomial logistic regression models were applied to examine whether the proportion of use of these information sources changes in different situations. We then used visual analysis to explore the question of which type of information source patients prefer to use when they seek medical assistance from doctors with different specialties.

Results: The 3 main information sources were online reviews (OR), family and friend recommendations (FR), and doctor recommendations (DR), with proportions of use of 32.93% (559,345/1,698,666), 23.68% (402,322/1,698,666), and 17.48% (296,912/1,698,666), respectively. Difficulty in medical decision making, the hospital level, and rural-urban differences were significantly associated with patients' preferred information sources for choosing doctors. Further, the sources of information that patients prefer to use were found to vary when they looked for doctors with different medical specialties.

Conclusions: Patients are less likely to use online reviews when medical decisions are more difficult or when the provider is not a tertiary hospital, the former situation leading to a greater use of online reviews and the latter to a greater use of family and friend recommendations. In addition, patients in large cities are more likely to use information from online reviews than family and friend recommendations. Among different medical specialties, for those in which personal privacy is a concern, online reviews are the most common source. For those related to children, patients are more likely to refer to family and friend recommendations, and for those related to surgery, they value doctor recommendations more highly. Our results can not only contribute to aiding government efforts to further promote the dissemination of health care information but may also help health care industry managers develop better marketing strategies.

(J Med Internet Res 2020;22(9):e20910) doi:10.2196/20910

KEYWORDS

information source; decision making; online reviews; online health care community; doctor; health information

Introduction

Background

Promoting patient choice can encourage competition among health care providers, which is likely to make health care more responsive to patient needs, enhance equity in care, and improve efficiency or quality as a result of effects such as reductions in wait times and costs [1-3]. Some studies have pointed out that consumer-directed health care does not always control costs better than other systems and that it has no significant effect on quality improvement [4,5]. The focus of this study is not on exploring the question of whether consumer-directed health care policies should be implemented but rather on understanding more fully the types of information sources consumers use when choosing a doctor. When they are able to make a rational choice, patients themselves can find a quality provider by weighing information from different sources. In practice, however, it is difficult for most patients to make fully rational choices [6-10], a difficulty which may result from certain individual characteristics. One of the most important characteristics is the patient's health literacy (ie, the capacity to access, process, and understand basic health information) [8,11-13]. Whether patients intend to actively use health care information when they choose between providers is another relevant question [14-16]. Furthermore, if patients encounter barriers to accessing information, such as short times for making decisions, geographical barriers [15], distrust of information [17], and information overload [18,19], this can also lead to bias in the decision-making process. Therefore, understanding which types of information sources are used widely for choosing doctors and how patients' preferences for information sources change under various circumstances will be useful for mitigating the barriers to information dissemination and for improving the decision-making process for patients.

The information sources that patients can use when choosing a doctor or health care provider are diverse. The mainstream literature has explored how the patient's choice is affected by public reports that compare the quality of health care providers. Although patients generally agree that such comparative reports are important, relatively few can understand or use them [15,19-21]. In order to take into account the widest possible variety of information sources for choosing doctors, this study divided information sources into 6 categories: online reviews (OR), family and friend recommendations (FR), doctor recommendations (DR), random registration (RR), multiple reasons (MR), and others (OTS). The first 3 types of information sources require further explanation.

First, with the growth of mature online health care communities (OHC) [22-26], many people can use online reviews in the OHC to share their opinions on the quality of their doctors. The function of the online health care community is now more than merely providing people with opportunities to share their ideas on the experience of seeing a doctor. The OHC now provides nontraditional channels and approaches to health care services, such as the use of electronic communication to query or improve a patient's clinical condition [27,28]. Since this type of information is usually free to the public, patients can refer to it easily when choosing a doctor. However, some potential complications make online reviews less credible. For example, doctors with lower qualifications are less likely to be rated online, and doctors' online scores are often exaggerated at the upper end of the quality spectrum [29]. In addition, avoiding shilling attacks remains a major challenge for all online review platforms [30]. Thus, relying entirely on online reviews for choosing a doctor is not an optimal solution, and the patient's decision-making process sometimes requires additional reference to other, more reliable information sources.

Second, while online reviews provided by mere acquaintances or by individuals who do not know the patient at all can be considered weak-tie recommendation sources, recommendations from friends or family numbers are strong-tie sources [31]. The main advantage of strong-tie sources is that they allow for an evaluation of alternatives based on the individual's situation. Specifically, compared with online reviews, family and friend recommendations are more likely to be a good source of information regarding affective cues rather than instrumental cues [32]. Another advantage is that patients usually do not have to worry about the deliberate falsification of information from friend recommendations. Nevertheless, family and friend recommendations are usually based on personal experience rather than on professional advice, which is based on professional medical knowledge. If a patient needs professional advice to choose the right doctor, the most common practice is to refer to a specialist, chosen with the assistance of a general physician. In addition, with the development of information communication technology, online medical consultations are becoming more and more popular [27,33], allowing doctors to answer patients' questions online. It should be noted that since the doctor-patient relationship is one of the most complex among all interpersonal relationships [34], it is not straightforward to determine whether doctor recommendations should be categorized as strong-tie or weak-tie sources.

Taken together, since these 3 types of information sources (ie, online reviews, family and friend recommendations, and doctor recommendations) do not have the same characteristics, understanding patients' preference to use one or another of them in different situations can be useful for facilitating the transfer of information in the health care system.

Finally, we briefly introduce the 3 information sources that have not been fully explained. First, random registration means that patients do not deliberately choose a doctor by referring to any particular information. Second, the multiple reasons designation refers to a situation in which a patient uses multiple information sources for choosing a doctor; for example, patients may look up doctor reviews online and then ask friends which doctor is better. Third, we refer to situations that cannot be categorized in terms of the first 5 information sources as others, a step which ensures that our classification criteria cover all possibilities.

The Research Problem

Understanding how patients choose doctors in different situations is important. However, due to considerations of cost, research conducted in the form of traditional questionnaires is often limited in terms of sample size, regional representativeness, and the diversity of specialties represented in the sample. This study overcomes this difficulty by collecting a large amount of data from an online health care community. Specifically, we address 3 main research questions. First, while some studies have discussed the impact of online physician evaluations or hospital evaluations on patient decision making [8,11,35,36], there are no studies, to the best of our knowledge, that specifically address the proportions in which different information sources are used in selecting doctors. Thus, our first research objective is to determine the proportions of use of the various information sources used to select physicians. This will give us a fuller picture of the current state of public access to medical information.

Second, there are a number of factors that can affect the decision-making process of patients seeking medical care. From a decision-making difficulty perspective [32], if surgical treatment is involved, patients need to give more careful consideration to choosing a doctor. From the perspective of the patient's external environment, urban-rural differences may contribute to differences in information accessibility or correspond to differences in the patient's medical knowledge [28]. From the point of view of the reputation of medical institutions, an official announcement of the hospital ratings can influence the perception of reliability that patients have toward doctors [11,35]. Thus, the second research objective is to explore how the information source used in the patient's choice of doctor is affected by the treatment method, the size of the city where the hospital is located, and the hospital level.

Third, regarding most medical management or decision-making issues, the impact of the particular medical specialty area is clear. For example, compared with other specialties, physicians in the gynecology/obstetrics and pediatrics specialty areas are more willing to contribute to online health care communities [26]. This study employed a large amount of data collected from the online medical community, so the sample included data related to a wide variety of specialties. This facilitated our exploration of the third research question: how does the information source that patients use change when patients choose a doctor with a different specialty?

Methods

Data Collection and Processing

We collected a large amount of public data from the Good Doctor website to explore our research questions. Founded in 2006, the Good Doctor website is now China's leading OHC. As of December 2019, this website contains the online review profiles of 610,000 doctors. Of these, about 220,000 doctors who have been certified by this website are able to create their own pages and interact directly with patients online. Moreover, we chose this OHC for two reasons. First, the Good Doctor website has provided standard options for replying to the question “Why did you choose this doctor?” since October 2016, and this allowed us to directly observe the information source used in the patient's choice of doctor. Second, since it is one of the most popular OHCs in China, many recent studies have used data sources from this website for exploring various research questions [24,26,33,34], which implies that the data collected from the site are reliable and representative. Our data collection procedure is described in detail below.

As with most OHCs, patients can provide online reviews on the Good Doctor website to share their experiences of seeing doctors. Figure 1 represents a typical example of a review; each review contains textual content, the patient name after deidentification, a time stamp, the name of the disease, the treatment received, and the information source used for choosing that particular doctor. As the time stamps of online reviews are removed by the system after 2 years, we collected data in 2 phases for the purpose of making use of a longer sample period. First, in October 2017, we started to collect all reviews posted on the Good Doctor website from October 1, 2016, to September 30, 2017. In this phase, 664,491 reviews were added to our sample. Two years later, we repeated the same data collection procedure and obtained 1,059,300 reviews posted from October 1, 2017, to September 30, 2019. In addition to these reviews, we also collected data about the hospitals and departments to which the doctors belonged. The address and the level of each hospital can be obtained from other pages of the Good Doctor website. After online reviews were merged with the hospital information in our sample, 25,125 reviews were excluded due to the corresponding address of the hospital being missing in the data. Finally, we retained 1,698,666 reviews in our sample, including reviews with information on 111,042 doctors from various specialty areas, 4747 different hospitals, 1095 observed days, and 31 provinces or municipalities in China. The data collection process is displayed in Figure 2. It is important to note that since all patients have been deidentified in the data and only aggregated results are reported in this manuscript, this work meets the ethical requirements for academic research.

Figure 1. Screenshot of a review on the Good Doctor website.

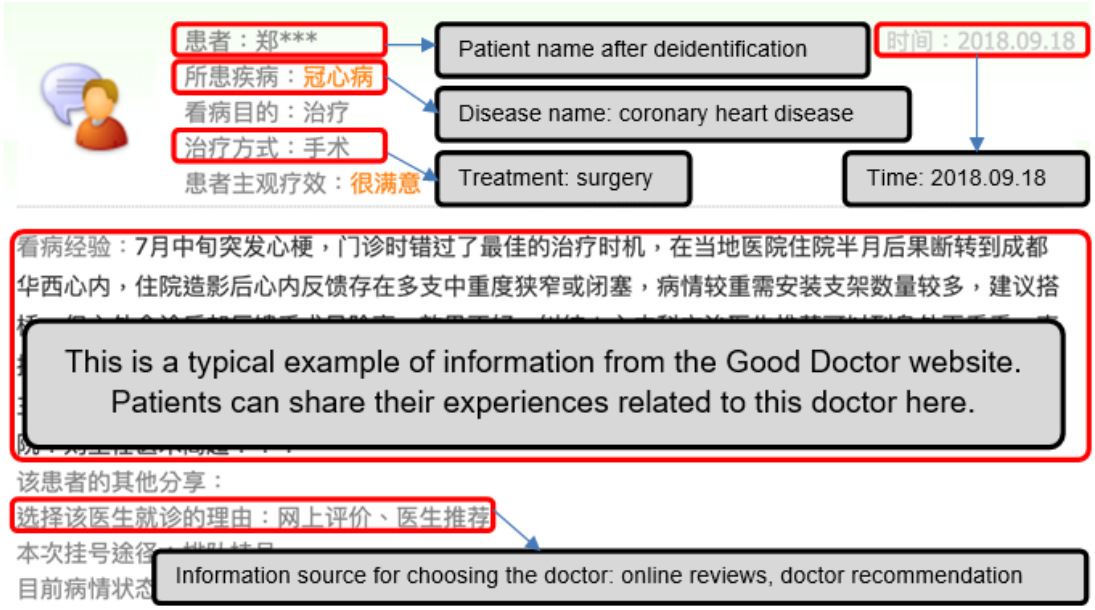
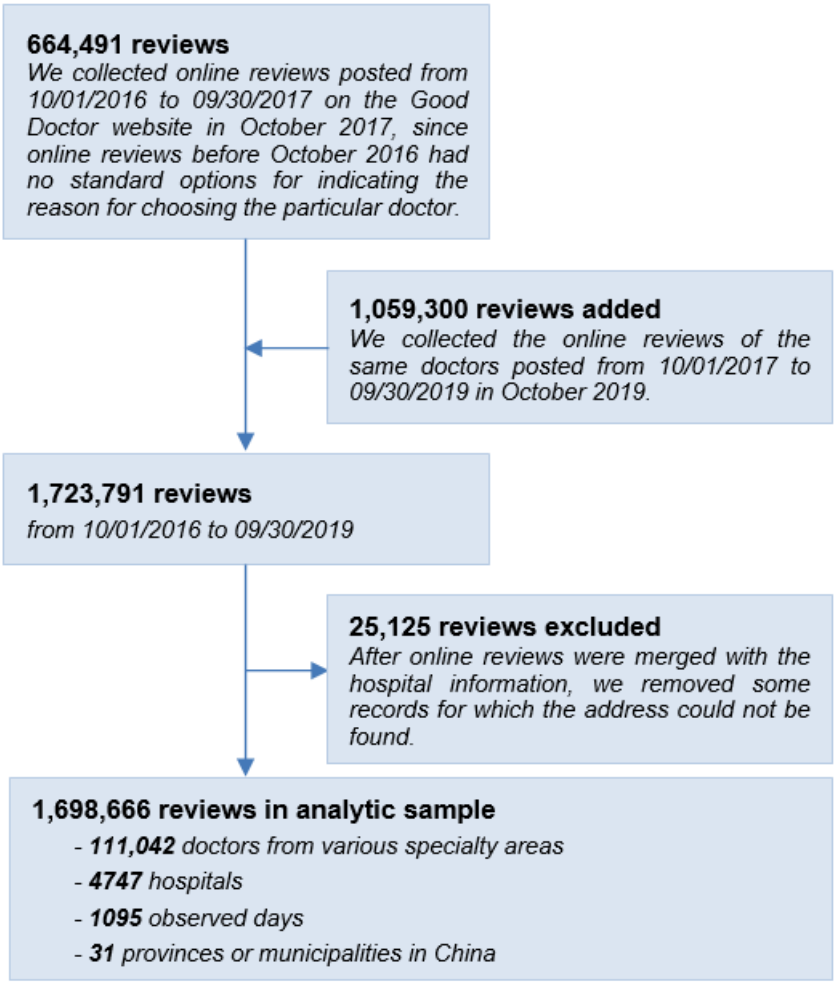


Figure 2. The sample collection process and characteristics.



Measures

In light of the kinds of data collected from the Good Doctor website, several variables involved in this study can be defined formally. First, when patients are filling out reviews, the Good Doctor website provides 4 defined options that can be indicated as the information source used to choose the doctor: online reviews, family and friend recommendations, doctor

recommendations, and random registration. In addition to these 4 options, the system allows patients to write any other information source, and patients can select more than one type of information source. For convenience, we denote the former option as others and the latter as multiple reasons. With this setup, we were able to ensure that each item in the sample was clearly categorized in terms of using a particular information source. It is also worth noting that we looked carefully at the actual content written in the OTS option, and we found most of the indicated sources were very similar in nature to the RR option. This study focuses on discussing the 3 most common information sources, OR, FR, and DR. The related results from reviews that indicate the other sources are included for the reader's reference only.

In order to investigate factors that may affect patients' decisions about using the information sources to choose doctors, this study considers 3 other variables that can also be observed from the website. First, the patient can indicate the way their condition was treated, such as medication, counseling, surgery, or other means, in their review. Because surgical treatment is usually a decision involving greater deliberation for members of the general public, a patient's consideration of whether to undergo surgery may result in a tendency to use certain information sources in choosing a doctor. Hence, we defined a dummy variable for treatment. When the treatment involved surgery, treatment=1; otherwise, treatment=0. Table 1 shows that up to 45.15% (766,933/1,698,666) of the patients in our sample were treated in a surgery-related manner. Second, we defined a dummy variable to indicate the hospital level, set to 1 if the doctor who was reviewed by the patient was from a tertiary hospital and set to 0 otherwise. Table 1 indicates that approximately 92.73% (1,575,216/1,698,666) of patients chose doctors from hospitals in the tertiary category, which is the official certification for the highest-quality hospitals. Third, because we want to understand how the degree of regional development affects the patients' use of information sources in choosing doctors, we further defined a dummy variable for city, denoting the level of the city. Specifically, city was set to 1 if the doctor was from Beijing, Shanghai, Shenzhen, or Guangzhou, and otherwise it was set to 0. Table 1 shows that 45.35% (770,384/1,698,666) of the doctors in our sample were in first-tier cities in China.

Table 1. Descriptive statistics of variables.

Variable	Observation, n	Proportion, % (95% CI)
Information source for choosing the doctor		
Online Reviews (OR)	559,345	32.93 (32.86-33.00)
Family and friend recommendations (FR)	402,322	23.68 (23.62-23.75)
Doctor recommendations (DR)	296,912	17.48 (17.42-17.54)
Multiple reasons (MR)	131,648	7.75 (7.71-7.79)
Random registration (RR)	111,800	6.58 (6.54-6.62)
Others (OTS)	196,639	11.58 (11.53-11.62)
Does the treatment include surgery? (TREATMENT)		
Yes (1)	766,933	45.15 (45.07-45.22)
No (0)	931,733	54.85 (54.78-54.93)
Is this doctor from a tertiary hospital? (HOSPITAL)		
Yes (1)	1,575,216	92.73 (92.69-92.77)
No (0)	123,450	7.27 (7.23-7.31)
Is this doctor from a big city? (CITY)		
Yes (1)	770,384	45.35 (45.28-45.43)
No (0)	928,282	54.65 (54.57-54.72)
Total	1,698,666	100.00

Statistical Analysis

To investigate the research question about which factors affect the selection of an information source when choosing a doctor, we looked at the impact of the above 3 dummy variables (ie, treatment, hospital, and city) on the choice of the information source. We used data visualization techniques to facilitate an understanding of the main findings of this research. Moreover, we formally adopted multinomial logistic regression to examine whether these findings were statistically significant. Multinomial logistic regression is an extension of binary logistic regression that allows the nominal dependent variable to belong to more than two categories and uses maximum likelihood estimation to evaluate the log odds of the outcomes. In practice, to deal with a multinomial logistic regression model with *K* possible outcomes, it is usual to estimate *K-1* independent binary logistic regression models, in which one outcome is chosen as a baseline and then the other *K-1* outcomes are compared to this baseline.

In this study, we consider 3 models with different baselines. To simplify the expression model, we denote IS_1, IS_2, … , IS_6 as OR, FR, DR, MR, RR, and OTS, respectively. Then, given OR

(one outcome of the dependent variable) as the baseline, we construct model 1 with the following form:

In the above equation,j=2, 3, 4, 5, or 6. Note that this model introduces 5 separate sets of regression coefficients for each possible outcome. For example, j=2 means the regression coefficients for the outcome are for FR. Hence, $\beta_{1,2}$, $\beta_{2,2}$, and $\beta_{3,2}$ represent how the log odds ratio of FR versus OR will change if 3 independent variables (ie, TREATMENT, HOSPITAL, and CITY) move from 0 to 1, respectively. Similar to model 1, we also define model 2 as having the following form:

In the above equation,j=1, 3, 4, 5, or 6. Similarly, we define model 3 as having the following form:

In model 3, j=1, 2, 4, 5, or 6.

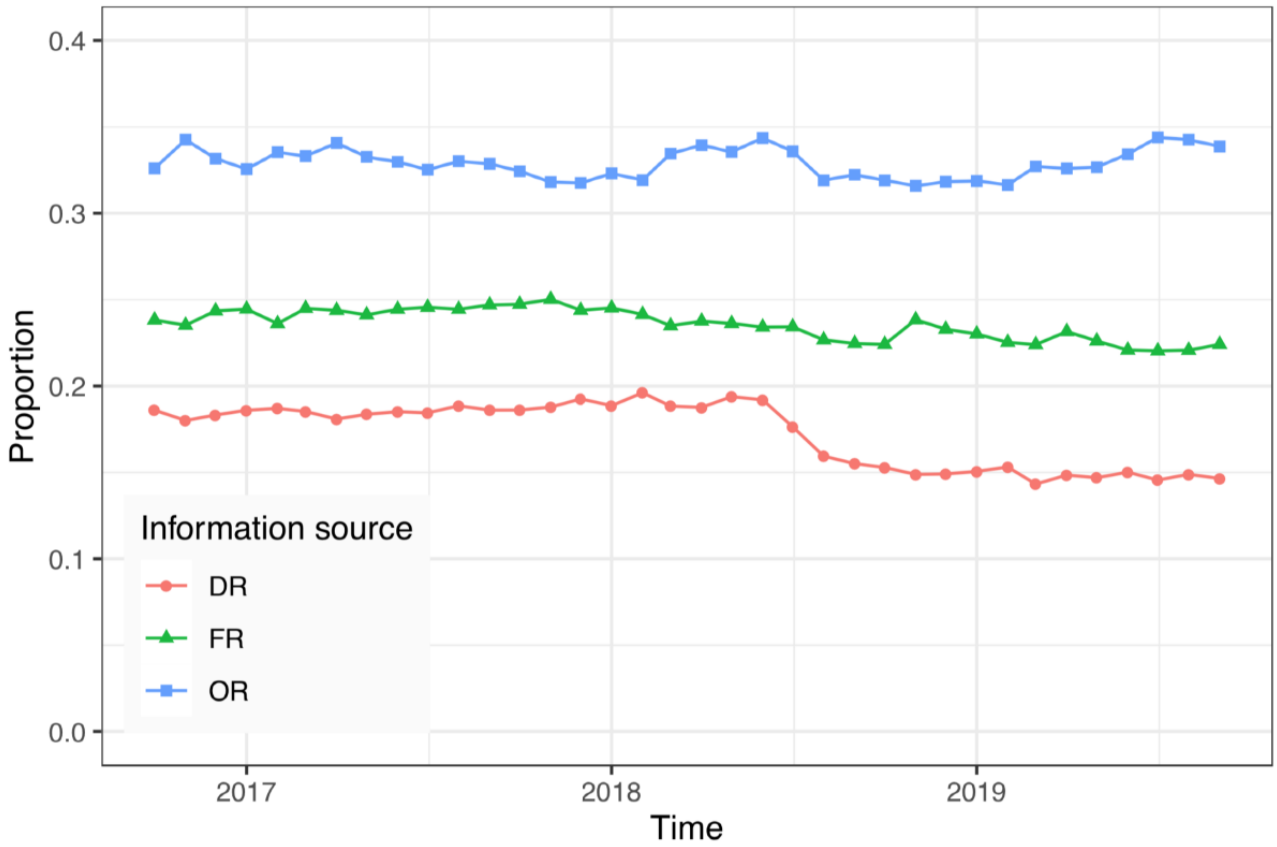
All of these models will be further discussed in future empirical studies.

Results

Descriptive Statistics

As shown in Table 1, patients relied primarily on the 3 main information sources, OR, FR, and DR, which accounted for 32.93% (95% CI 32.86%–33.00%; 559,345/1,698,666), 23.68% (95% CI 23.62%-23.75%; 402,322/1,698,666), and 17.48% (95% CI 17.42%-17.54%; 296,912/1,698,666), respectively, when choosing a doctor. To check whether these results were robust, we further divided the sample for the entire 3-year period into 36 subsamples of 1 month. We then similarly examined the proportions of the 3 information sources, OR, FR, and DR, in each subsample. Figure 3 displays the time-varying proportions of these 3 information sources in each month. The results showed proportion values between 31.59% (11,836/37,553) and 34.40% (13,676/39,760) for OR, between 22.04% (8762/39,760) and 25.02% (12,483/49,890) for FR, and between 14.32% (5547/38,723) and 19.61% (8001/40,797) for DR. These results indicate that the proportions of these 3 sources did not change much over time during the sample period.

Figure 3. The proportions of the 3 main information sources in each month. DR: doctor recommendations; FR: family and friend recommendations; OR: online reviews.



In order to clarify the effect of treatment, city, and hospital on proportions in which information sources were used, we first provide visualizations of the results for each of these factors. We take treatment as an example to illustrate how relevant graphics were generated. First, all 1,698,666 samples were divided into 2 subsamples based on treatment, with sample sizes of 766,933 (treatment=1) and 931,733 (treatment=0). We then calculated the weight of different information sources in each subsample (these outcomes are presented in Figure 4). This figure indicates that when patients' treatments included surgery compared with not involving surgery, patients showed greater preference for finding a suitable doctor by using DR (20.58%

vs 14.93%; 157,808/766,933 vs 139,104/931,733), FR (24.02% vs 23.41%; 184,293/766,933 vs 218,083/931,733), or MR (8.72% vs 6.95%; 66,859/766,933 vs 64,789/931,733) rather than OR (31.70% vs 33.94%; 243,096/766,933 vs 316,249/931,733), OTS (8.65% vs 13.99%; 66,322/766,933 vs 130,317/931,733), or RR (6.34% vs 6.78%; 48,609/766,933 vs 63,191/931,733). Similarly, Figure 5 shows that when doctors chosen by patients were from a large city compared with not being from a large city, patients showed greater preference for making a decision through OR (37.82% vs 28.87%; 291,332/770,384 vs 268,013/928,282) rather than other sources, especially FR (21.23% vs 25.72%; 163,536/770,384 vs 238,786/928,282). Finally, Figure 6 shows that when doctors chosen by patients were from a tertiary hospital compared with not being from the tertiary hospital, patients showed a strong preference for using OR (33.63% vs 23.95%; 529,779/1,575,216 vs 29,566/123,450) rather than FR (23.23% vs 29.46%; 365,956/1,575,216 vs 36,366/123,450) as their main information source.

While the visual graphs allow us to easily grasp the main findings, we still need to verify the findings' statistical significance using the multinomial logistic regression model.

Figure 4. The influence of treatment on reasons for choosing doctors. DR: doctor recommendations; FR: family and friend recommendations; MR: multiple reasons; OR: online reviews; OTS: others; RR: random registration.

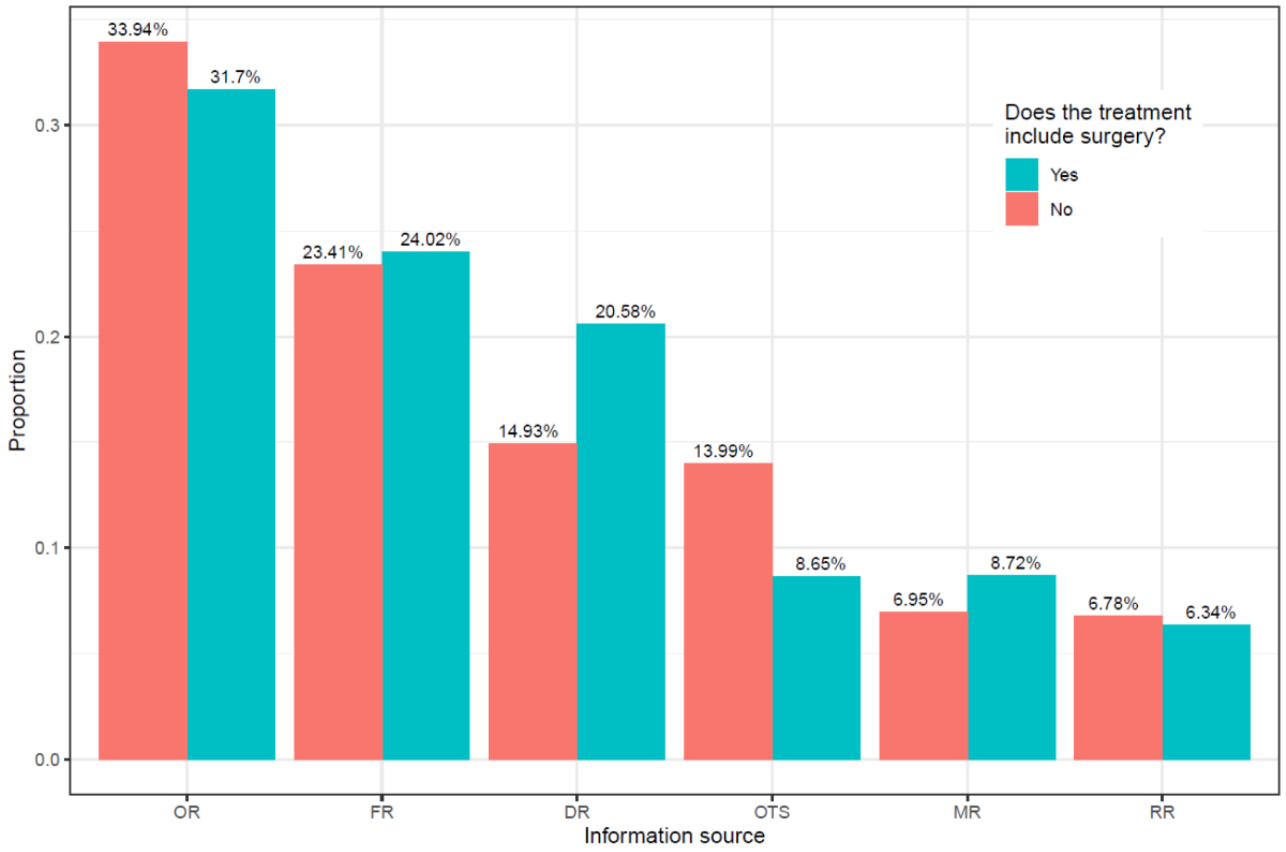


Figure 5. The influence of city on reasons for choosing doctors. DR: doctor recommendations; FR: family and friend recommendations; MR: multiple reasons; OR: online reviews; OTS: others; RR: random registration.

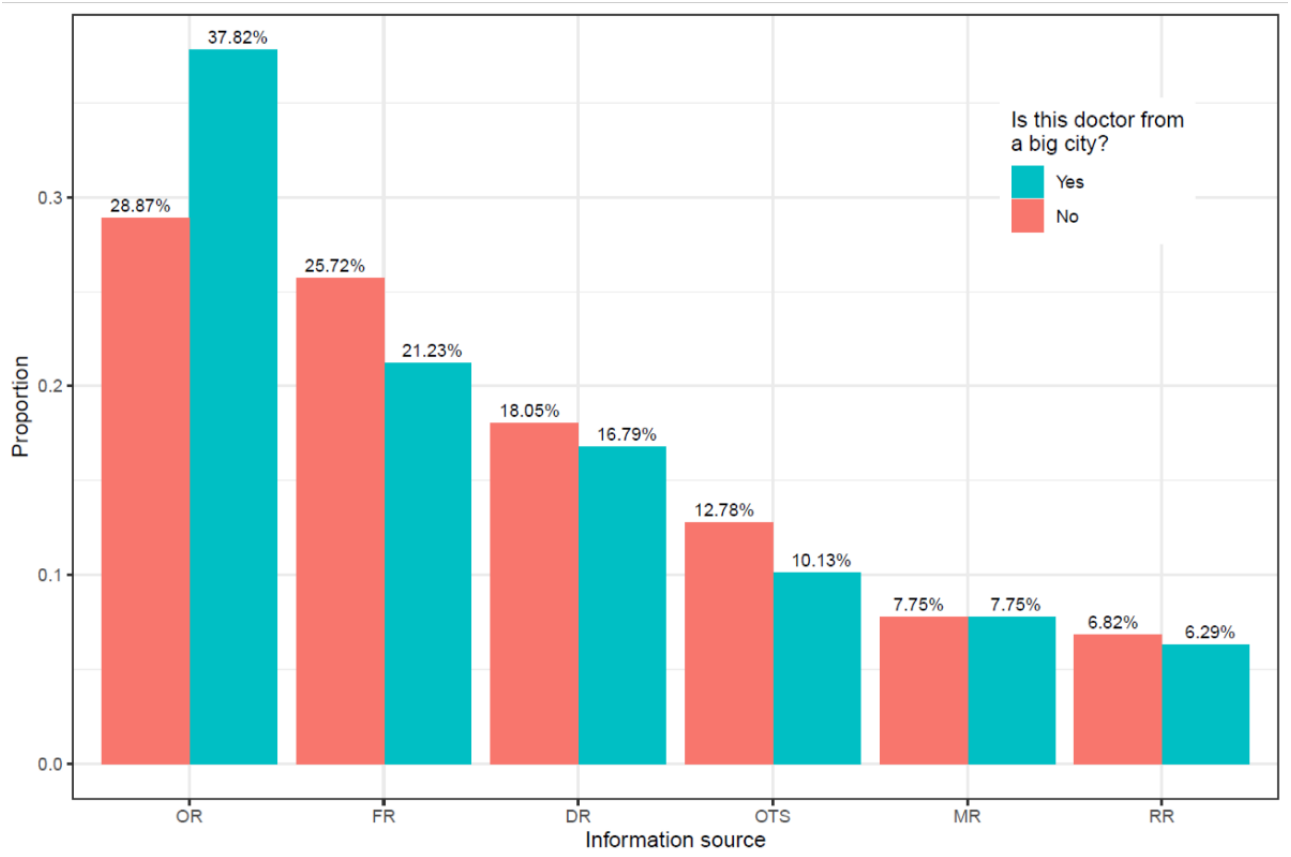
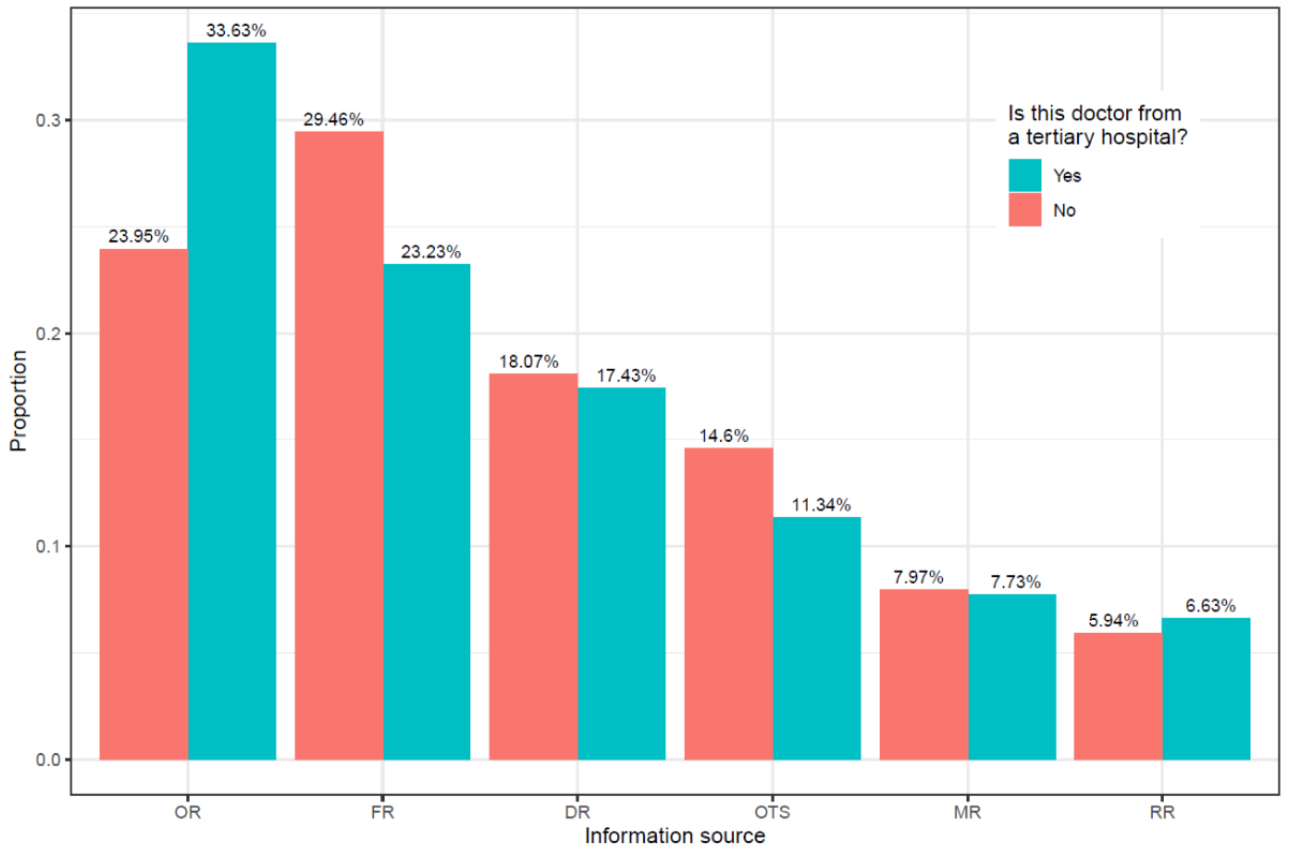


Figure 6. The influence of hospital on reasons for choosing doctors. DR: doctor recommendations; FR: family and friend recommendations; MR: multiple reasons; OR: online reviews; OTS: others; RR: random registration.



Multinomial Logistic Regressions

Table 2 presents the multinomial logistic regression estimation for model 1, model 2, and model 3 with the 1,698,666 samples. Since the outcome measure in this kind of analysis is the odds ratio, we briefly introduce that concept as follows. An odds ratio greater than 1 indicates that the possibility of the patient using the comparison information source relative to the possibility of the patient using the baseline information source increases as the dummy variable changes from 0 to 1. In other words, the comparison information source is more likely to be used by patients in that case. For example, in model 1, where OR is set as the baseline information source, if we consider FR as the comparison information source, we can observe that the odds ratios of the 3 dummy variables, treatment, city, and hospital, are 1.16, 0.63, and 0.58, respectively. This means that when treatment changes from 0 to 1, patients are more likely to use FR than OR, but when city or hospital changes from 0 to 1, patients are less likely to use FR than OR.

Table 2. Multinomial logistic regression results.

Model	Treatment (yes), odds ratio (95% CI)	City (yes), odds ratio (95% CI)	Hospital (yes), odds ratio (95% CI)
Model 1: Using online reviews as the baseline			
Family and friend recommendations	1.16 (1.15-1.17)	0.63 (0.62-0.63)	0.58 (0.57-0.59)
Doctor recommendations	1.54 (1.53-1.56)	0.69 (0.68-0.69)	0.69 (0.68-0.70)
Multiple reasons	1.39 (1.37-1.41)	0.74 (0.73-0.75)	0.69 (0.68-0.71)
Random registration	1.04 (1.03-1.05)	0.70 (0.69-0.71)	0.82 (0.79-0.84)
Others	0.70 (0.69-0.71)	0.63 (0.63-0.64)	0.59 (0.58-0.60)
Model 2: Using family and friend recommendations as the baseline			
Online reviews	0.86 (0.85-0.87)	1.59 (1.58-1.61)	1.73 (1.71-1.76)
Doctor recommendations	1.33 (1.31-1.34)	1.09 (1.08-1.10)	1.19 (1.17-1.21)
Multiple reasons	1.20 (1.18-1.21)	1.18 (1.17-1.20)	1.20 (1.17-1.23)
Random registration	0.89 (0.88-0.91)	1.12 (1.11-1.14)	1.41 (1.38-1.45)
Others	0.60 (0.59-0.61)	1.01 (1.00-1.02)	1.02 (1.00-1.04)
Model 3: Using doctor recommendations as the baseline			
Online reviews	0.65 (0.64-0.65)	1.46 (1.45-1.47)	1.45 (1.43-1.48)
Family and friend recommendations	0.75 (0.75-0.76)	0.92 (0.91-0.93)	0.84 (0.82-0.85)
Multiple reasons	0.90 (0.89-0.91)	1.08 (1.07-1.10)	1.01 (0.98-1.03)
Random registration	0.67 (0.66-0.68)	1.03 (1.01-1.04)	1.19 (1.15-1.22)
Others	0.45 (0.45-0.46)	0.93 (0.91-0.94)	0.85 (0.84-0.87)

Based on the above interpretation, the significance of the odds ratio representations in different situations is clear, so we will not provide additional explanation for each number in Table 2. Nevertheless, there are two points that deserve further elaboration. First, for testing of the statistical significance of each odds ratio, the corresponding 95% confidence interval is also presented in the table. The absence of 1 in the confidence interval for a variable's odds ratio indicates that this variable is significantly associated with the patient's preference for using the comparison or baseline information sources. In Table 2, all odds ratios are significant, except for those in model 3, which has 1 odds ratio that is not significant. Second, we should note that the results in Table 2 are consistent with the findings in Figures 4-6. For example, Figure 4 shows that patients prefer to use OR if the doctor is in a big city. Likewise, in model 1 of Table 2, all odds ratios associated with the city variable are less than 1, meaning that relative to any other information source, patients are more likely to use OR when city changes from 0 to 1. Next, we explore the relationship between different medical specialties and the patients' preference for particular information sources.

Impacts of Different Medical Specialties

In order to investigate the impact of different medical specialties on patient preferences for using particular information sources to choose doctors, we grouped doctors based on the Chinese names of their departments as listed on the Good Doctor website. Because the total number of categories into which medical specialties can be subdivided is very high, and because even almost identical medical specialties may have differences in Chinese terminology, the number of groups was greater than 2000. While we could try to determine which specialties might be placed in the same category, objectively setting the criteria for such consolidation is a challenge. Hence, we decided not to perform a subjective consolidation process, but instead kept only the 36 medical specialties that had more than 10,000 records for the following analysis. A Chinese and English comparison table of these 36 medical specialty names is

available in Multimedia Appendix 1 for readers' reference. This method of categorization can bring several benefits. First, this method does not include any subjective interpretations and retains the full names of medical specialties. It retains complete information about specialties and allows users to explore more possible issues. Second, these medical specialties have a large enough sample size to ensure statistical analytical reliability, and this also shows that these specialties are the most common in China. Third, the current results allow users to build a new category according to their requirements and find the relevant proportions of information sources.

For example, if we combine neurosurgery and neurology as a new specialty category and measure the proportion of OR for this specialty, according to the outcomes in Table 3, we can easily find that the proportion of OR is 33.45%, which is calculated with (61,199 × 0.3071 + 37,711 × 0.3789) ÷ (61,199 + 37,711). For each medical specialty, we calculated the proportion of each information source and the corresponding 95% confidence interval. The relevant results are presented in Table 3. For example, 90,693 patients saw a urologist, and the proportion of those using OR, FR, DR, MR, OTS, and RR to select a doctor was 34.72% (31,489/90,693), 19.86% (18,012/90,693), 19.14% (17,359/90,693), 8.03% (7283/90,693), 10.07% (9133/90,693), and 8.18% (7419/90,693), respectively. We summarized the proportion of the 36 medical specialties for which the selection of a doctor was made using the 3 most important information sources (ie, OR, FR, and DR) in Figure 7, and we gave the names of the medical specialties with the top 5 highest proportions for selection using each information source. Specifically, in terms of OR, the top 5 medical specialties that showed the highest percentage were reconstructive surgery (9241/16,657, 55.48%), plastic surgery (6209/11,214, 55.37%), andrology (7099/13,680, 51.89%), skin–sexually transmitted disease (7632/18,834, 40.52%), and oral and maxillofacial surgery (5105/12,834, 39.78%). In terms of FR, the order was traditional Chinese medicine (5451/14,282, 38.17%), reproductive medicine center (5048/14,187, 35.58%), reproductive center (4321/12,301, 35.13%), pediatrics (10,370/32,723, 31.69%), and obstetrics (5029/16,425, 30.62%). In terms of DR, the order was pediatric surgery (4818/19,626, 24.55%), cardiac surgery (2742/11,184, 24.52%), general surgery 2 (4043/17,210, 23.49%), neurosurgery (13,197/61,199, 22.74%), and general surgery 1 (6657/31,489, 21.14%). From the above results, it is clear that when patients choose a doctor with a different specialty, they may use different sources of information. A related discussion is included in the next section.

Table 3. Comparison of reasons for choosing doctors among various specialties.

Specialty	Observation	OR[a], % (95% CI)	FR[b], % (95% CI)	DR[c], % (95% CI)	MR[d], % (95% CI)	OTS[e], % (95% CI)	RR[f], % (95% CI)
Urology	90,693	34.72 (34.41-35.03)	19.86 (19.60-20.12)	19.14 (18.89-19.40)	8.03 (7.85-8.20)	10.07 (9.87-10.26)	8.18 (8.01-8.36)
Dermatology	77,961	39.13 (38.79-39.47)	19.38 (19.11-19.66)	11.38 (11.15-11.60)	6.68 (6.50-6.85)	14.37 (14.13-14.62)	9.06 (8.86-9.26)
Gynecology	65,551	34.41 (34.04-34.77)	24.35 (24.02-24.68)	15.24 (14.97-15.52)	6.81 (6.62-7.01)	12.83 (12.58-13.09)	6.35 (6.17-6.54)
Orthopedics	64,714	31.73 (31.37-32.09)	23.84 (23.51-24.17)	18.42 (18.13-18.72)	8.38 (8.16-8.59)	10.55 (10.32-10.79)	7.08 (6.88-7.28)
Neurosurgery	61,199	30.71 (30.35-31.08)	25.31 (24.96-25.65)	22.74 (22.41-23.07)	8.84 (8.61-9.06)	7.71 (7.50-7.92)	4.70 (4.53-4.87)
Ophthalmology	51,074	33.31 (32.90-33.72)	21.21 (20.86-21.57)	20.28 (19.93-20.63)	7.90 (7.67-8.13)	10.81 (10.54-11.08)	6.49 (6.28-6.71)
Neurology	37,711	37.89 (37.40-38.38)	21.00 (20.59-21.42)	15.06 (14.70-15.42)	7.05 (6.80-7.31)	12.02 (11.69-12.35)	6.98 (6.72-7.24)
Pediatrics	32,723	22.35 (21.90-22.81)	31.69 (31.18-32.19)	15.99 (15.59-16.39)	7.96 (7.67-8.26)	17.36 (16.95-17.77)	4.64 (4.41-4.87)
General surgery 1	31,489	29.34 (28.84-29.84)	24.86 (24.39-25.34)	21.14 (20.69-21.59)	8.25 (7.95-8.55)	10.13 (9.80-10.46)	6.27 (6.00-6.54)
Obstetrics and gynecology	29,641	23.94 (23.46-24.43)	30.11 (29.59-30.64)	16.39 (15.96-16.81)	7.40 (7.10-7.70)	15.85 (15.43-16.26)	6.31 (6.03-6.59)
Thoracic surgery	29,563	33.43 (32.89-33.96)	25.58 (25.08-26.07)	18.89 (18.44-19.33)	9.10 (8.77-9.42)	7.81 (7.51-8.12)	5.20 (4.95-5.46)
Cardiology	26,792	29.43 (28.88-29.98)	25.62 (25.10-26.14)	17.26 (16.81-17.71)	7.03 (6.73-7.34)	13.52 (13.11-13.92)	7.14 (6.84-7.45)
Otolaryngology	25,257	39.19 (38.59-39.79)	19.67 (19.18-20.16)	14.68 (14.24-15.12)	7.14 (6.82-7.46)	12.76 (12.35-13.17)	6.56 (6.26-6.87)
Endocrinology	24,283	32.86 (32.27-33.45)	24.33 (23.79-24.87)	14.18 (13.74-14.62)	7.24 (6.91-7.57)	13.17 (12.75-13.60)	8.22 (7.87-8.56)
Gastroenterology	23,038	35.85 (35.23-36.47)	21.69 (21.16-22.23)	15.00 (14.54-15.46)	7.09 (6.76-7.42)	13.11 (12.67-13.54)	7.26 (6.92-7.59)
Breast surgery	21,887	32.59 (31.97-33.21)	25.41 (24.84-25.99)	14.83 (14.36-15.30)	7.33 (6.98-7.67)	12.30 (11.86-12.73)	7.54 (7.19-7.89)
Hepatobiliary surgery	20,382	35.55 (34.89-36.20)	24.53 (23.94-25.12)	16.58 (16.07-17.09)	8.48 (8.10-8.86)	8.50 (8.11-8.88)	6.37 (6.03-6.70)
Pediatric surgery	19,626	25.42 (24.81-26.03)	19.23 (18.68-19.79)	24.55 (23.95-25.16)	7.87 (7.49-8.24)	13.55 (13.07-14.03)	9.38 (8.97-9.78)
Skin-STD[g]	18,834	40.52 (39.82-41.22)	16.49 (15.96-17.02)	12.69 (12.21-13.17)	6.95 (6.59-7.31)	15.48 (14.96-15.99)	7.87 (7.49-8.26)
Otorhinolaryngology – head and neck surgery	18,595	37.96 (37.26-38.65)	18.98 (18.41-19.54)	16.85 (16.32-17.39)	7.36 (6.99-7.74)	11.22 (10.77-11.68)	7.63 (7.24-8.01)
Division of rheumatology	17,420	29.25 (28.57-29.92)	25.55 (24.90-26.19)	18.23 (17.66-18.81)	7.63 (7.24-8.03)	13.94 (13.42-14.45)	5.40 (5.07-5.74)
General surgery 2	17,210	27.38 (26.71-28.05)	23.67 (23.04-24.31)	23.49 (22.86-24.13)	8.65 (8.23-9.07)	9.81 (9.36-10.25)	7.00 (6.61-7.38)

Specialty	Observation	OR[a], % (95% CI)	FR[b], % (95% CI)	DR[c], % (95% CI)	MR[d], % (95% CI)	OTS[e], % (95% CI)	RR[f], % (95% CI)
Reconstructive surgery	16,657	55.48 (54.73-56.24)	15.78 (15.23-16.34)	9.52 (9.08-9.97)	8.62 (8.19-9.05)	6.79 (6.41-7.17)	3.80 (3.51-4.09)
Obstetrics	16,425	24.47 (23.81-25.13)	30.62 (29.92-31.33)	15.40 (14.85-15.95)	8.10 (7.68-8.51)	17.21 (16.63-17.78)	4.21 (3.90-4.51)
Stomatology	14,541	29.94 (29.20-30.69)	20.32 (19.67-20.98)	16.80 (16.19-17.41)	7.31 (6.89-7.73)	16.68 (16.07-17.28)	8.95 (8.48-9.41)
Spine surgery	14,534	33.38 (32.61-34.14)	26.60 (25.88-27.32)	15.46 (14.87-16.05)	8.72 (8.26-9.18)	8.54 (8.08-8.99)	7.31 (6.88-7.73)
Traditional Chinese medicine	14,282	27.62 (26.88-28.35)	38.17 (37.37-38.96)	12.18 (11.65-12.72)	6.44 (6.04-6.84)	12.39 (11.85-12.93)	3.20 (2.91-3.49)
Reproductive medicine center	14,187	23.18 (22.49-23.88)	35.58 (34.79-36.37)	15.44 (14.84-16.03)	8.26 (7.81-8.71)	12.77 (12.22-13.31)	4.77 (4.42-5.12)
Andrology	13,680	51.89 (51.06-52.73)	13.03 (12.46-13.59)	10.18 (9.67-10.68)	8.93 (8.45-9.40)	10.00 (9.50-10.50)	5.98 (5.58-6.38)
Oral and maxillofacial surgery	12,834	39.78 (38.93-40.62)	14.82 (14.21-15.43)	15.14 (14.52-15.76)	7.52 (7.06-7.98)	11.82 (11.26-12.38)	10.92 (10.38-11.46)
Otolaryngology	12,661	37.15 (36.31-38.00)	24.30 (23.56-25.05)	13.40 (12.81-14.00)	7.94 (7.47-8.41)	10.12 (9.59-10.64)	7.08 (6.64-7.53)
Reproductive center	12,301	25.36 (24.59-26.13)	35.13 (34.28-35.97)	15.04 (14.41-15.67)	7.99 (7.51-8.47)	11.62 (11.05-12.18)	4.86 (4.48-5.24)
Psychiatry	11,980	38.10 (37.23-38.97)	22.71 (21.96-23.46)	13.01 (12.41-13.62)	5.97 (5.54-6.39)	14.14 (13.52-14.76)	6.07 (5.64-6.50)
Plastic surgery	11,214	55.37 (54.45-56.29)	19.23 (18.51-19.96)	8.94 (8.41-9.46)	7.45 (6.97-7.94)	6.40 (5.95-6.86)	2.60 (2.31-2.90)
Cardiac surgery	11,184	29.17 (28.32-30.01)	20.71 (19.96-21.46)	24.52 (23.72-25.31)	8.95 (8.42-9.48)	10.39 (9.82-10.96)	6.27 (5.82-6.72)
Anorectal	10,872	34.57 (33.67-35.46)	23.52 (22.72-24.32)	15.06 (14.38-15.73)	8.31 (7.80-8.83)	11.40 (10.8-11.99)	7.15 (6.66-7.63)

[a]OR: online reviews.

[b]FR: family and friend recommendations.

[c]DR: doctor recommendations.

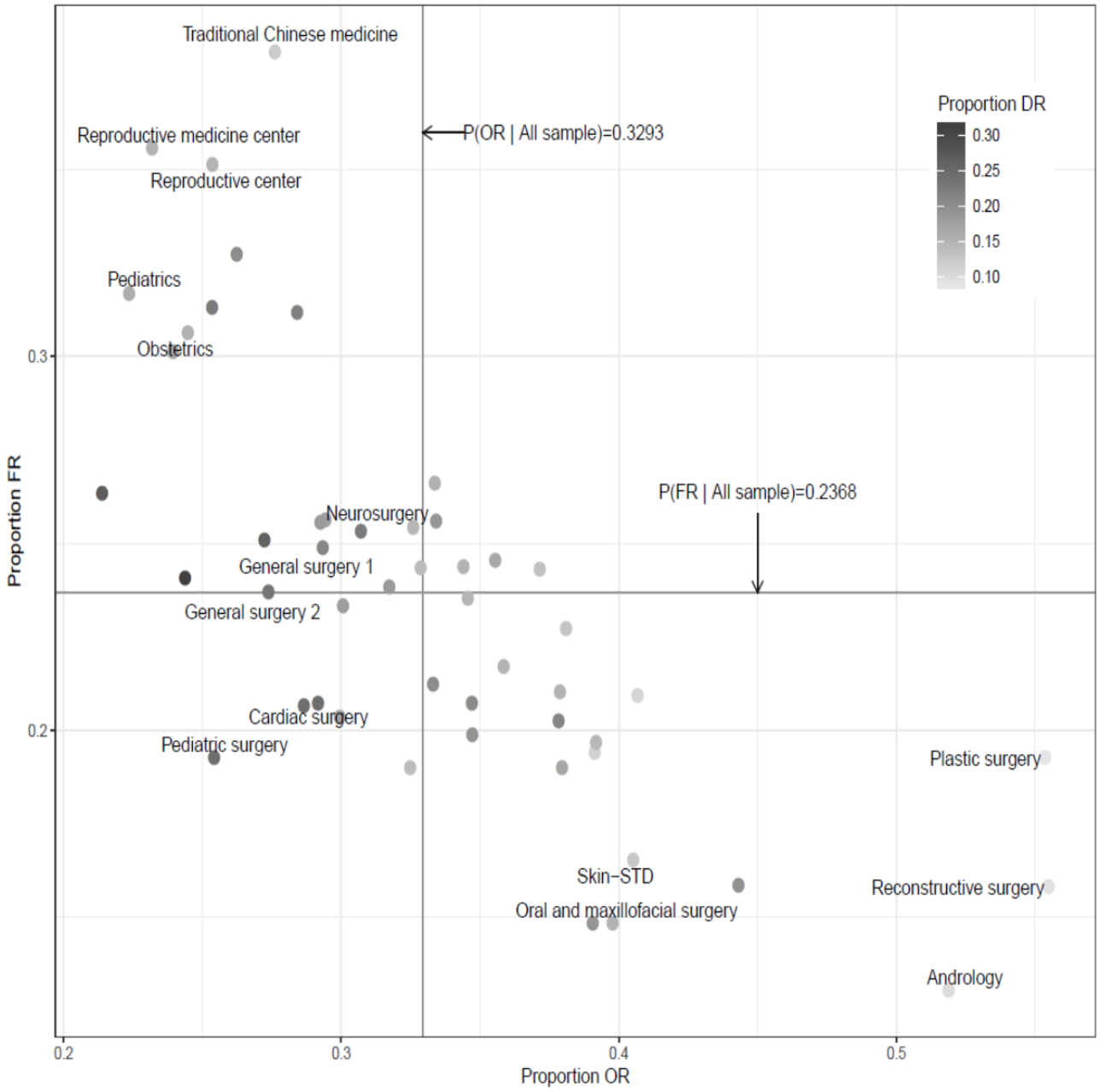
[d]MR: multiple reasons.

[e]OTS: others.

[f]RR: random registration.

[g]STD: sexually transmitted disease.

Figure 7. Proportions in the use of information sources for different medical specialties. DR: doctor recommendations; FR: family and friend recommendations; OR: online reviews; STD: sexually transmitted disease.



Discussion

Principal Results

The current study used 1,698,666 samples collected from the Good Doctor Website, including information related to 111,042 doctors, 4747 different hospitals, 1095 observation days, and 31 provinces or municipalities in China; the diversity of these data ensures the empirical results are representative. The main findings can be summarized in three points. First, our results showed that the 3 primary sources of information used by Chinese patients were OR, FR, and DR, accounting for 74.09% (1,258,579/1,698,666) of all usage of sources. Surprisingly, the proportion of use of DR is very low, at only 17.48% (296,912/1,698,666), which may be related to the development of the family doctor service system. In many countries, initial consultations are made through the family doctor service system, followed by referrals to specialists, with the expectation that health resources will be used effectively to control health costs and improve health outcomes [37,38]. Indeed, the Chinese government is also improving the family doctor service system [39,40], but it will not be easy to change the general public's perceptions regarding medical care in a short time. In 2009, China started a new round of health care system reform, in which the development of community health service centers (CHSCs) is a key measure. In the following years, CHSCs had become significantly more functional, but residents did not have sufficient confidence in the primary health care system, and therefore the tiered diagnosis and treatment system was not established as expected. In 2011, the government began to establish a system of family doctors in a number of developed cities based on the existing CHSC service model. Local residents are encouraged to contract with family doctors and use

CHSC-based initial consultations for referral services. More information on the Chinese health care system can be found in a recent World Health Organization report [41]. Since our results show that 32.93% (559,345/1,698,666) of patients refer to OR in making their selections, the full use of strategies related to the population's online tendencies may be a direction that health care authorities could consider in promoting the development of the family doctor service system. For example, governments could further liberalize telemedicine regulations to make it easier for people to conduct initial consultations with family doctors over the internet.

Second, understanding the varied preferences of patients in using information sources in different situations can be of aid in the effective dissemination of health care information. For example, if the patient needs surgery, they usually need professional advice or a more trusted information source, such as DR or FR. In other words, when decision making becomes more difficult, people use weak-tie sources like OR less often [32]. Furthermore, as shown in Figure 4, patients are more likely to consider using multiple information sources or to avoid randomly choosing a doctor if they need surgery. In addition, the impact of urban-rural differences is evident. It may be that the information gap leads patients in nonmetropolitan areas to use OR less often than patients in large cities, or it may be that patients in large cities are less likely to have access to strong-tie sources due to more distant relationships and must therefore rely on OR. It is also possible that patients in large cities have a higher level of medical literacy, so they are better able to organize and comprehend information from the internet rather than having to seek advice from family or friends. Finally, we explore the question of whether doctors being from tertiary hospitals has an effect on patient preferences in using information sources. If the physician is not from a tertiary hospital, the patient is more likely to choose a doctor using FR rather than OR. This phenomenon can be explained in terms of brand impact. The effects of a good brand can convince consumers that it features consistent quality, which in turn can reduce consumer perceptions of risk [42]. Tertiary hospitals are like a better brand of the health care system in China. When a doctor is from a tertiary hospital, the patient's risk perception lowers and they are more willing to use online reviews to choose a doctor. Conversely, if the doctor is not from a tertiary hospital, the patient's risk perception rises, and they may need to use more trusted information sources to help them make decisions.

Third, some meaningful outcomes can be found when we look at the 5 medical specialties with the highest weighting in the use of each of the information sources indicated in Figure 7. Related to the use of DR, it is consistent with the above findings that the highest 5 specialties were related to surgery; they include pediatric surgery and cardiac surgery. Related to the use of OR, we find that the medical professions for which patients most prefer to use this information source were associated with personal privacy. Specifically, when patients need the help of this type of medical specialist, they usually gather information about the doctor through the internet and do not want their family or friends to know about it, so specialties with high OR have a low FR. For example, when patients needed the help of an andrologist, 51.89% (7099/13,680) referred to online reviews, and only 13.03% (1783/13,680) sought advice from family or friends. Corresponding to the use of FR, in addition to traditional Chinese medicine, the highest 5 specialties were related to children. It seems likely that friends and family may not necessarily have relevant experience with a particular medical problem, but if they have children, they can certainly share their own medical experience with patients. Another possible reason is that patients tend to be cautious about child-related issues, so they prefer to use strong-tie sources rather than weak-tie sources [32], which also reflects the fact that these specialties related to the high use of FR are also related to low use of OR. As for traditional Chinese medicine, the reason it is related to such a high use of FR is unclear. Perhaps it is because traditional Chinese medicine is very popular in the Chinese world [43], and therefore it is easy to get useful information from friends and family. In addition, in most cases, patients seeking a traditional Chinese medicine doctor usually do not have an urgent medical condition, so they usually have sufficient time to gather information from family or friends to make a choice.

In summary, based on the study's findings, we make 4 specific recommendations for hospital administrators or policymakers in the government health care sector. First, avoiding disruption of the health care system with rumors or untrue advertisements on the internet is an important task because online reviews are currently the main source of information for the general public when they go to the clinic, especially for some medical specialties that are related to personal privacy. Second, the rate of referrals by family doctors still appears to be low, and consideration could be given to the internet habits of the population in conjunction with the family doctor system when the effort is made to change the general public’s medical habits, thereby mitigating overcrowding in hospitals. Third, the gap between urban and rural residents in obtaining health information persists. This may be due to differences in ease of access to medical information or may reflect differences in individuals’ ability to interpret medical knowledge, but people in nonmetropolitan areas make less frequent use of online information when choosing a doctor. The government should continue to strengthen the information infrastructure in rural areas and actively promote health literacy locally. Fourth, people's trust in nontertiary hospitals remains low, so most patients rely primarily on the experience of family and friends when they need to choose a doctor in a nontertiary hospital. The fact that doctors in nontertiary hospitals are also fully capable of treating patients if the disease is not particularly serious is something that administrators must make people aware of.

Limitations and Future Work

This study has some limitations and points to possible future research questions. First, our sample was drawn from patients or their families who have shared thoughts about their medical visits on the OHC, which means that most of these patients have the ability to use the internet. Thus, our sample selection may have automatically excluded patients who did not have access to the internet, which may have led to the outcomes overestimating the proportion in which patients used online reviews as their information source and underestimating the impact of urban-rural differences. Second, because of ethical

considerations, this study did not use any variables related to patients' personal information. If some personal information about patients can be included in the research model (with prior permission), further insight into the behavioral characteristics involved in the choice of doctor will be gained. Third, because the sample used in this study is entirely from China, differences in culture or social environment may result in the findings of the study not being applicable to other countries or regions. We believe that the difficulty of medical decision making, hospital level, and rural-urban differences are still significantly associated with patients' preferred information source in most cases. However, the impacts of medical specialties may be significantly different in other countries and regions, a possibility which requires further investigation in the future. Finally, this study examines which information sources patients use to choose their doctors, but it does not attempt to determine how good the decisions are. Specifically, we used a large amount of actual data to explore which information sources patients preferred to use to help them make decisions in different situations. Whether or not those decisions actually led them to the right doctors was a question beyond the scope of our study. Future research could employ textual analysis of the online review content to provide insight into which information sources are used to help patients make better decisions in different contexts.

Conclusions

Patients' abilities to make appropriate decisions when seeking a doctor often depend on their own knowledge or characteristic ability to compile relevant information, and the way that they obtain useful information is the first question they must face before a decision can be made. This study investigated the types of information sources that are currently widely used in choosing doctors and showed how the preferred sources vary from situation to situation, thus contributing to an understanding of how to help patients obtain the information they will need in the future. This study makes several specific contributions. First, we know based on a large amount of data that patients currently use online reviews, family and friend recommendations, and doctor recommendations to get the information they need when choosing a doctor. Second, different circumstances correspond to differences in patients' preferences for information sources. Specifically, when medical decisions become more difficult, as when surgery is required, or when a medical facility is not rated as tertiary, patients are less likely to refer to online evaluations, referring more often in the former case to recommendations given by doctors and in the latter case asking more often for recommendations from family and friends. In addition, rural-urban differences are also associated with differences in patient preferences. Patients in large cities are more likely to use information from online reviews rather than recommendations from family and friends. Third, we explored differences in patient preference for information sources in relation to a variety of medical specialties. For specialties related to personal privacy, online reviews were the most common source of information; for specialties related to children, patients were more likely to refer to the opinions of their family and friends, and for specialties related to surgery, they sought out the advice of doctors more often. Of particular interest is that most traditional Chinese medicine patients chose doctors on the recommendations of their family and friends. These results may not only help the government to further promote the dissemination of medical information but may also aid managers in the health care industry in developing better marketing strategies.

Acknowledgments

This article presents independent research. We thank the editor and anonymous referee for very helpful comments and suggestions that greatly improved the paper. Y-T Hsu gratefully acknowledges financial support from the Ministry of Education Project of Humanities and Social Sciences (No. 20YJCZH199).

Conflicts of Interest

None declared.

Multimedia Appendix 1
English and Chinese medical specialty name comparison table.
[DOCX File , 24 KB - jmir_v22i9e20910_app1.docx]

References

1. Dixon A, Robertson R, Bal R. The experience of implementing choice at point of referral: a comparison of the Netherlands and England. Health Econ Policy Law 2010 Jul;5(3):295-317 [FREE Full text] [doi: 10.1017/S1744133110000058] [Medline: 20462469]
2. Grytten J, Sørensen RJ. Patient choice and access to primary physician services in Norway. Health Econ Policy Law 2009 Jan;4(Pt 1):11-27 [FREE Full text] [doi: 10.1017/S1744133108004623] [Medline: 19099615]
3. Vrangbaek K, Robertson R, Winblad U, Van de Bovenkamp H, Dixon A. Choice policies in Northern European health systems. Health Econ Policy Law 2012 Jan;7(1):47-71 [FREE Full text] [doi: 10.1017/S1744133111000302] [Medline: 22221928]

4. Parente ST, Feldman R, Christianson JB. Evaluation of the effect of a consumer-driven health plan on medical care expenditures and utilization. Health Serv Res 2004 Aug;39(4 Pt 2):1189-1210 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00282.x] [Medline: 15230920]
5. Buntin MB, Damberg, Haviland A, Kapur K, Lurie N, McDevitt R, et al. Consumer-directed health care: early evidence about effects on cost and quality. Health Aff (Millwood) 2006;25(6):w516-w530 [FREE Full text] [doi: 10.1377/hlthaff.25.w516] [Medline: 17062591]
6. Lubalin JS, Harris-Kojetin LD. What do consumers want and need to know in making health care choices? Med Care Res Rev 1999;56 Suppl 1:67-102; discussion 103. [doi: 10.1177/1077558799056001S04] [Medline: 10354679]
7. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433 [FREE Full text] [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
8. Fasolo B, Reutskaja E, Dixon A, Boyce T. Helping patients choose: how to improve the design of comparative scorecards of hospital quality. Patient Educ Couns 2010 Mar;78(3):344-349 [FREE Full text] [doi: 10.1016/j.pec.2010.01.009] [Medline: 20171822]
9. Foster MM, Earl PE, Haines TP, Mitchell GK. Unravelling the concept of consumer preference: implications for health policy and optimal planning in primary care. Health Policy 2010 Oct;97(2-3):105-112 [FREE Full text] [doi: 10.1016/j.healthpol.2010.04.005] [Medline: 20466449]
10. Kooreman P, Prast H. What Does Behavioral Economics Mean for Policy? Challenges to Savings and Health Policies in the Netherlands. De Economist 2010 May 27;158(2):101-122 [FREE Full text] [doi: 10.1007/s10645-010-9141-6]
11. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? A systematic review. Med Care 2009 Jan;47(1):1-8 [FREE Full text] [doi: 10.1097/MLR.0b013e3181808bb5] [Medline: 19106724]
12. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/1077558707070640020301] [Medline: 17406019]
13. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychol Bull 2009 Nov;135(6):943-973 [FREE Full text] [doi: 10.1037/a0017327] [Medline: 19883143]
14. Harris KM. How do patients choose physicians? Evidence from a national survey of enrollees in employment-related health plans. Health Serv Res 2003 Apr;38(2):711-732 [FREE Full text] [doi: 10.1111/1475-6773.00141] [Medline: 12785569]
15. Kolstad JT, Chernew ME. Quality and consumer decision making in the market for health insurance and health care services. Med Care Res Rev 2009 Feb;66(1 Suppl):28S-52S. [doi: 10.1177/1077558708325887] [Medline: 19029288]
16. McGlone TA, Butler ES, McGlone VL. Factors influencing consumers' selection of a primary care physician. Health Mark Q 2002;19(3):21-37 [FREE Full text] [doi: 10.1300/J026v19n03_03] [Medline: 12077807]
17. Schwartz LM, Woloshin S, Birkmeyer JD. How do elderly patients decide where to go for major surgery? Telephone interview survey. BMJ 2005 Oct 08;331(7520):821 [FREE Full text] [doi: 10.1136/bmj.38614.449016.DE] [Medline: 16192286]
18. Hibbard JH, Greene J, Daniel D. What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. Med Care Res Rev 2010 Jun;67(3):275-293. [doi: 10.1177/1077558709356300] [Medline: 20093399]
19. Damman OC, Hendriks M, Rademakers J, Delnoij DMJ, Groenewegen PP. How do healthcare consumers process and evaluate comparative healthcare information? A qualitative study using cognitive interviews. BMC Public Health 2009 Nov 20;9:423 [FREE Full text] [doi: 10.1186/1471-2458-9-423] [Medline: 19930564]
20. Fotaki M, Roland M, Boyd A, McDonald R, Scheaff R, Smith L. What benefits will choice bring to patients? Literature review and assessment of implications. J Health Serv Res Policy 2008 Jul;13(3):178-184 [FREE Full text] [doi: 10.1258/jhsrp.2008.007163] [Medline: 18573768]
21. Mukamel DB, Mushlin AI. The impact of quality report cards on choice of physicians, hospitals, and HMOs: a midcourse evaluation. Jt Comm J Qual Improv 2001 Jan;27(1):20-27. [doi: 10.1016/s1070-3241(01)27003-5] [Medline: 11147237]
22. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24. [doi: 10.2196/jmir.2360] [Medline: 23372115]
23. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38. [doi: 10.2196/jmir.2003] [Medline: 22366336]
24. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134. [doi: 10.2196/jmir.4365] [Medline: 26032933]
25. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: International comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
26. Wang JN, Chiu YL, Yu H, Hsu YT. Understanding a Nonlinear Causal Relationship Between Rewards and Physicians' Contributions in Online Health Care Communities: Longitudinal Study. J Med Internet Res 2017 Dec 21;19(12):e427. [doi: 10.2196/jmir.9082] [Medline: 29269344]

27. Miscione G. Telemedicine in the Upper Amazon: Interplay with Local Health Care Practices. MIS Quarterly 2007 Jan;31(2):403-425 [FREE Full text] [doi: 10.2307/25148797]
28. Goh JM, Gao G, Agarwal R. The Creation of Social Value: Can an Online Health Community Reduce Rural-Urban Health Disparities? MISQ 2016 Jan 1;40(1):247-263 [FREE Full text] [doi: 10.25300/MISQ/2016/40.1.11]
29. Gao G, Greenwood BN, Agarwal R, McCullough JS. Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perceptions of Quality? MISQ 2015 Mar 3;39(3):565-589. [doi: 10.25300/MISQ/2015/39.3.03]
30. Gunes I, Kaleli C, Bilge A, Polat H. Shilling attacks against recommender systems: a comprehensive survey. Artif Intell Rev 2012 Nov 2;42(4):767-799. [doi: 10.1007/s10462-012-9364-9]
31. Brown JJ, Reingen PH. Social Ties and Word-of-Mouth Referral Behavior. J Consum Res 1987 Dec 28;14(3):350-341. [doi: 10.1086/209118]
32. Duhan DF, Johnson SD, Wilcox JB, Harrell GD. Influences on Consumer Use of Word-of-Mouth Recommendation Sources. Journal of the Academy of Marketing Science 1997 Sep 01;25(4):283-295 [FREE Full text] [doi: 10.1177/0092070397254001]
33. Guo S, Guo X, Fang Y, Vogel D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. Journal of Management Information Systems 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
34. Guo S, Guo X, Zhang X, Vogel D. Doctor–patient relationship strength's impact in an online healthcare community. Information Technology for Development 2017 Mar 09;24(2):279-300 [FREE Full text] [doi: 10.1080/02681102.2017.1283287]
35. Dijs-Elsinga J, Otten W, Versluijs MM, Smeets HJ, Kievit J, Vree R, et al. Choosing a hospital for surgery: the importance of information on quality of care. Med Decis Making 2010;30(5):544-555 [FREE Full text] [doi: 10.1177/0272989X09357474] [Medline: 20110514]
36. Rothenfluh F, Germeni E, Schulz PJ. Consumer Decision-Making Based on Review Websites: Are There Differences Between Choosing a Hotel and Choosing a Physician? J Med Internet Res 2016 Jun 16;18(6):e129. [doi: 10.2196/jmir.5580] [Medline: 27311623]
37. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Feb 28;297(8):831-841 [FREE Full text] [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
38. Mendis S, Johnston SC, Fan W, Oladapo O, Cameron A, Faramawi MF. Cardiovascular risk management and its impact on hypertension control in primary care in low-resource settings: a cluster-randomized trial. Bull World Health Organ 2010 Jun;88(6):412-419 [FREE Full text] [doi: 10.2471/BLT.08.062364] [Medline: 20539854]
39. Lian S, Chen Q, Yao M, Chi C, Fetters MD. Training Pathways to Working as a General Practitioner in China. Fam Med 2019 Mar;51(3):262-270 [FREE Full text] [doi: 10.22454/FamMed.2019.329090] [Medline: 30861081]
40. Huang J, Lu W, Wang L, Zhang T, Liu C, Liu S, et al. A preliminary effect analysis of family doctor and medical insurance payment coordination reform in Changning District of Shanghai, China. BMC Fam Pract 2019 May 10;20(1):60 [FREE Full text] [doi: 10.1186/s12875-019-0949-0] [Medline: 31077151]
41. Meng Q, Yang H, Chen W, Sun Q, Liu X. People's Republic of China health system review. Health Sys Trans. 2015. URL: https://iris.wpro.who.int/bitstream/handle/10665.1/11408/9789290617280_eng.pdf [accessed 2020-08-27]
42. Ha HY. The Effects of Consumer Risk Perception on Pre-purchase Information in Online Auctions: Brand, Word-of-Mouth, and Customized Information. J Comput Mediated Commun 2002;8(1):JCMC813 [FREE Full text] [doi: 10.1111/j.1083-6101.2002.tb00160.x]
43. Chen FP, Chen TJ, Kung YY, Chen YC, Chou LF, Chen FJ, et al. Use frequency of traditional Chinese medicine in Taiwan. BMC Health Serv Res 2007 Feb 23;7:26 [FREE Full text] [doi: 10.1186/1472-6963-7-26] [Medline: 17319950]

Abbreviations

CHSC: community health service center
DR: doctor recommendations
FR: family and friend recommendations
MR: multiple reasons
OHC: online health care community
OR: online reviews
OTS: others
RR: random registration

Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by K(Na, CF Wu; comments to author 22.06.20; revised version received 28.07.20; accepted 03.08.20; published 16.09.20.

Please cite as:
Zhang S, Wang JN, Chiu YL, Hsu YT
Exploring Types of Information Sources Used When Choosing Doctors: Observational Study in an Online Health Care Community
J Med Internet Res 2020;22(9):e20910
URL: http://www.jmir.org/2020/9/e20910/
doi:10.2196/20910
PMID:32936080

Original Paper

Impact of MyDiabetesPlan, a Web-Based Patient Decision Aid on Decisional Conflict, Diabetes Distress, Quality of Life, and Chronic Illness Care in Patients With Diabetes: Cluster Randomized Controlled Trial

Catherine Yu[1,2,3,4], BSc, MHSc, MD, FRCPC; Dorothy Choi[1], BSc; Brigida A Bruno[5], BSc, MD; Kevin E Thorpe[3,6], MMath; Sharon E Straus[2,4], HSBc, MSc, MD, FRCPC; Paul Cantarutti[7], MD; Karen Chu[8], MD; Paul Frydrych[9], MD; Amy Hoang-Kim[1], MSc, PhD; Noah Ivers[10,11], MD, PhD; David Kaplan[11,12], MSc, MD; Fok-Han Leung[1], MD; John Maxted[13], MD; Jeremy Rezmovitz[14], MD; Joanna Sale[15], MSc, PhD; Sumeet Sodhi-Helou[16], MPH, MD; Dawn Stacey[17], BScN, MScN, PhD; Deanna Telner[18], MD

[1]St. Michael's Hospital (Unity Health Toronto), Toronto, ON, Canada
[2]Department of Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[4]Li Ka Shing Knowledge Institute, St. Michael's Hospital (Unity Health Toronto), Toronto, ON, Canada
[5]Department of Medicine, University of Toronto, Toronto, ON, Canada
[6]Applied Health Research Centre, Li Ka Shing Knowledge Institute, St. Michael's Hospital (Unity Health Toronto), Toronto, ON, Canada
[7]Southlake Regional Health Centre, Newmarket, ON, Canada
[8]Bridgepoint Active Healthcare (Sinai Health System), Toronto, ON, Canada
[9]Mount Dennis Weston Health Centre, Humber River Family Health Team, Toronto, ON, Canada
[10]Department of Family and Community Medicine, Women's College Hospital, Toronto, ON, Canada
[11]University of Toronto, Toronto, ON, Canada
[12]North York General Hospital, Toronto, ON, Canada
[13]Markham Stouffville Hospital, Markham, ON, Canada
[14]Sunnybrook Health Sciences Centre, Toronto, ON, Canada
[15]Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital (Unity Health Toronto), Toronto, ON, Canada
[16]Toronto Western Family Health Team, Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada
[17]School of Nursing, University of Ottawa, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[18]South East Toronto Family Health Team (Toronto East Health Network), Toronto, ON, Canada

Corresponding Author:
Catherine Yu, BSc, MHSc, MD, FRCPC
St. Michael's Hospital (Unity Health Toronto)
30 Bond St
Toronto, ON
Canada
Phone: 1 416 864 6060
Email: catherine.yu@unityhealth.to

Abstract

Background: Person-centered care is critical for delivering high-quality diabetes care. Shared decision making (SDM) is central to person-centered care, and in diabetes care, it can improve decision quality, patient knowledge, and patient risk perception. Delivery of person-centered care can be facilitated with the use of patient decision aids (PtDAs). We developed *MyDiabetesPlan*, an interactive SDM and goal-setting PtDA designed to help individualize care priorities and support an interprofessional approach to SDM.

Objective: This study aims to assess the impact of *MyDiabetesPlan* on decisional conflict, diabetes distress, health-related quality of life, and patient assessment of chronic illness care at the individual patient level.

Methods: A two-step, parallel, 10-site cluster randomized controlled trial (first step: provider-directed implementation only; second step: both provider- and patient-directed implementation 6 months later) was conducted. Participants were adults 18 years and older with diabetes and 2 other comorbidities at 10 family health teams (FHTs) in Southwestern Ontario. FHTs were randomly assigned to *MyDiabetesPlan* (n=5) or control (n=5) through a computer-generated algorithm. *MyDiabetesPlan* was integrated into intervention practices, and clinicians (first step) followed by patients (second step) were trained on its use. Control participants received static generic Diabetes Canada resources. Patients were not blinded. Participants completed validated questionnaires at baseline, 6 months, and 12 months. The primary outcome at the individual patient level was decisional conflict; secondary outcomes were diabetes distress, health-related quality of life, chronic illness care, and clinician intention to practice interprofessional SDM. Multilevel hierarchical regression models were used.

Results: At the end of the study, the intervention group (5 clusters, n=111) had a modest reduction in total decisional conflicts compared with the control group (5 clusters, n=102; −3.5, 95% CI −7.4 to 0.42). Although there was no difference in diabetes distress or health-related quality of life, there was an increase in patient assessment of chronic illness care (0.7, 95% CI 0.4 to 1.0).

Conclusions: Use of goal-setting decision aids modestly improved decision quality and chronic illness care but not quality of life. Our findings may be due to a gap between goal setting and attainment, suggesting a role for optimizing patient engagement and behavioral support. The next steps include clarifying the mechanisms by which decision aids impact outcomes and revising *MyDiabetesPlan* and its delivery.

Trial Registration: ClinicalTrials.gov NCT02379078; https://clinicaltrials.gov/ct2/show/NCT02379078

(J Med Internet Res 2020;22(9):e16984) doi:10.2196/16984

KEYWORDS

shared decision making; goals of care; decision aid; diabetes mellitus; decisional conflict; quality of life; patient assessment of chronic illness care; diabetes distress; randomized clinical trials

Introduction

Person-centered care, whereby health care providers are encouraged to partner with patients to co‐design and deliver personalized care [1], is critical to delivering high-quality diabetes care [2]. Shared decision making (SDM) is central to person-centered care [3] and can be facilitated with the use of patient decision aids (PtDAs), such as Diabetes Medication Choice [4] or patient activation programs [5]. A meta-analysis of 16 studies using SDM in diabetes care found an association with improved decision quality, patient knowledge, and patient risk perception [6]. Decisional conflict is a measure of decision quality and reflection of the individual's uncertainty in choosing an option [7]. Decisions with low decisional conflict have been associated with less regret [8] and less emotional and psychological distress [9]. Thus, engagement of the person in the decision-making process to make a high-quality decision is an important step in delivering person-centered care.

Other markers of person-centered care in diabetes management include health-related quality of life [10], diabetes distress [11], and perception of chronic illness care [12]. Health-related quality of life reflects the combined impact of an individual's physical, psychological, and social well-being on health-related quality of life, and in patients living with diabetes, lower quality of life is associated with poorer clinical outcomes such as glycemic control [13]. Decisional conflict is a reflection of an individual's uncertainty in choosing an option, whereas diabetes distress refers to an individual's emotional state as a result of the burden of self-care tasks related to diabetes self-management and is associated with reduced self-care, reduced quality of life, and poor glycemic control [11,14]. Similarly, an individual's positive perception of chronic illness care is associated with improved self-management behaviors and glycemic control [12]. Thus, these outcomes reflect person-centered care.

Decision Aids to Support Person-Centered Care

Delivery of person-centered care and optimization of these outcomes can be facilitated with the use of PtDAs. A systematic review of 105 studies found that PtDAs improved decision quality and process and reduced decisional conflict but had no impact on quality of life [15]. For trials evaluating PtDAs for diabetes decisions, patients were more likely to change their medication. With the evolution of technology, PtDAs have expanded from static pamphlets, booklets, or videos to include interactive video- and computer-based programs; the latter enable individualized content tailored to the patient's characteristics and needs [16,17].

Given the complexity of diabetes care and multiple competing priorities, decision aids that support goal setting are particularly relevant. To date, one randomized controlled trial (RCT) of a goal-setting and SDM aid (which offered individually tailored risk information and treatment options for multiple risk factors to help patients prioritize between clinical issues) in 344 patients with uncomplicated type 2 diabetes found no difference in patient empowerment for setting and achieving goals [18]. However, this intervention neither addressed patient-important priorities and preferences specifically nor used a provider-specific point-of-care tool at the time of consultation.

Objectives

To address this gap, we developed *MyDiabetesPlan* and a multicomponent PtDA toolkit, which includes patient-directed, provider-directed, and point-of-care tools. *MyDiabetesPlan* is an interactive SDM and goal-setting PtDA designed to help individualize care priorities and support an interprofessional

approach to SDM, in the context of complex guideline recommendations for patients with type 1 or type 2 diabetes and other comorbidities. The overall aim of this study was to assess the impact of *MyDiabetesPlan* on decisional conflict, diabetes distress, health-related quality of life, and patient assessment of chronic illness care in individual patients in primary care practice groups randomized to *MyDiabetesPlan*.

Methods

Research Program Overview

We previously reported on how the development and refinement of *MyDiabetesPlan*, an interprofessional shared decision-making (IPSDM) toolkit, following the principles of user-centered design [19,20]. In this paper, we describe our assessment of the effectiveness of *MyDiabetesPlan* through a two-step cluster RCT followed by individual interviews. We used the Consolidated Standards of Reporting Trials (CONSORT) checklist (CONSORT-eHealth [21] and CONSORT extension for cluster trials [22]) to report this paper (Multimedia Appendix 1).

Study Design

The study protocol and methods are described in previous studies [19,23]. We conducted a two-step, parallel, cluster RCT with a 1:1 allocation ratio. We selected a clustered design and randomized at the level of primary care practice groups to avoid contamination (eg, clinicians using an IPSDM approach with control patients). In brief, the first step was provider-directed (*MyDiabetesPlan* was delivered to physicians, nurses, dietitians, or pharmacists), whereas the second step (occurring 6 months later) was provider- and patient-directed (patients were asked to use *MyDiabetesPlan* by themselves before the appointment; this was then reviewed by the provider team). We chose a two-step approach because a prior feasibility study [20] identified that patients required clinician assistance for completing their initial *MyDiabetesPlan*. Outcome measures were administered at the first step (baseline), second step (6 months later), and follow-up (12 months later).

Setting and Participants

All primary care practice groups in Southern Ontario that had interprofessional staff (eg, nurse, dietitian, or pharmacist) and electronic medical records (EMRs, to identify patients with diabetes) were invited to participate via email, telephone, and in-person or virtual presentation to the executive or medical director; groups without interprofessional staff or EMRs were excluded. All primary care physicians in these group practices were invited to participate. A research coordinator identified patients with diabetes (type 1 or type 2) and 2 other comorbidities (heart disease, stroke, hypertension, cancer, chronic lung disease, arthritis, inflammatory bowel disorders, and urinary incontinence) from each consenting physician's practice using a combination of keywords, International Classification of Diseases and billing codes. Patients were excluded if they did not speak English, had documented cognitive deficits, were unable to provide informed consent, had limited life expectancy (<1 year), or were not available for follow-up. Potentially eligible patients were identified via EMR query, and eligibility was further confirmed by chart review; from this group, participants were randomly selected and invited to participate and provided consent by telephone.

Intervention

MyDiabetesPlan was described previously [19]. It is a web-based PtDA in which patients populate their cardiometabolic and psychosocial profiles and general care priorities; *MyDiabetesPlan* then generates individualized diabetes-specific goals and strategies based on these inputs that the patients then select, resulting in an action plan. After randomization, at study onset, clinicians at intervention sites underwent a one-on-one 60-min tutorial in their clinic room by the research coordinator, with access to a one-page *how-to* guide and 2-min video. During subsequent clinical encounters, a member of the interprofessional team (nurse or dietitian) logged into *MyDiabetesPlan* and completed it with the patient; the physician subsequently reviewed the resultant action plan with the patient. At 6 months, patients at intervention sites were provided with a patient-directed *how-to* guide and video and directed to update *MyDiabetesPlan* according to their progress before the appointment. The research coordinator followed up with participants by email and telephone at study onset, followed by quarterly debriefing sessions, in both individual and group formats.

Control

Clinicians in the control sites received paper copies of the executive summary of the Diabetes Canada clinical practice guidelines [24] and a postcard outlining web-based clinical information resources. After 6 months, patients in the control sites received a Diabetes Canada patient education pamphlet [25] regarding diabetes self-management and a postcard outlining web-based additional patient resources.

Outcome Measures

The primary outcome was decisional conflict [26]; secondary outcomes were diabetes distress [27], health-related quality of life [28], chronic illness care [29], and intention to engage in IPSDM [30] (Table 1). Decisional conflict was measured using the Decisional Conflict Scale (DCS), which can predict individuals' intentions and subsequent behavior. It has a test-retest coefficient of 0.81 and internal consistency coefficients ranging from 0.78 to 0.92 [7]. These outcomes were assessed at the individual participant level, at baseline, and at 6 months and 12 months (after an appointment) through a web-based survey or by mail.

Table 1. Outcome measures and validated scales.

Outcome	Scale	Description and psychometric properties
Patient outcomes		
Decisional conflict	DCS[a] (16-item, 5 subscales; O'Connor, 1995) [7]	This scale consists of 16 items with 5 response categories (0=strongly agree, 4=strongly disagree), where higher scores indicate greater decisional conflict. The scale includes subscales for uncertainty, informed, values clarity, support, and effective decision. Test-retest correlation and Cronbach alpha exceed .78. It correlated with related constructs of knowledge, regret, and discontinuance and had excellent predictive validity. A clinically significant effect size is 0.30 to 0.40; scores lower than 25 are associated with implementing decision; scores exceeding 37.5 are associated with decision delay or feeling unsure about implementation. The primary outcome decisional conflict has been demonstrated to be responsive to change over time and thus will yield meaningful results when measured at baseline and throughout the study intervention.
Diabetes distress	DDS[b] (Polonsky et al, 2005) [27]	The DDS is a 17-item instrument that assesses emotional distress and functioning specific to living with diabetes. Responses are scored on a 6-point Likert-type scale from 1=no problem to 6=serious problem. Scores can range from 17 to 102, with higher scores indicating poorer diabetes-related quality of life and lower scores indicating better diabetes-related quality of life. This instrument has been found to have high internal reliability with a Cronbach alpha of .93, good convergent validity with the CESD[c] (r=0.56) and self-care behaviors including lower adherence to eating recommendations (r=0.30) and lower levels of physical activity (r=0.20).
Health-related quality of life	SF[d]-12 (Ware, 1996) [31]	The SF-12 is a 12-item version of the SF-36. The SF-12 is a widely used and validated generic measure of health-related quality of life. It is a multidimensional measure of perceived health, assessing physical functioning, physical role, bodily pain, general health, vitality, social functioning, emotional role, and mental health. Scores ranges from 0 to 100, with higher scores reflecting better health. Its validity was demonstrated in studies of patients with various chronic conditions [32,33].
Chronic illness care	PACIC[e] Scale (Glasgow et al, 2005) [29]	The PACIC Scale assesses the degree to which care is congruent with the Chronic Care Model from the perspective of the patient. Specifically, it was designed to measure patient activation, goal setting, problem solving/contextual counseling, delivery system design/decision support, and follow-up/coordination. The PACIC Scale has been used to evaluate a variety of chronic health conditions, including type 2 diabetes [29,34,35]. It has moderate test-retest reliability (r=0.58 during the course of 3 months) and correlates moderately with measures of primary care and patient activation (r=0.32-0.60, median=0.50, P<.001).
Clinician outcome		
Intention to engage in IPSDM[f]	CPD[g] Reaction Questionnaire (Legare et al, 2014) [36]	This 11-item questionnaire is based on the Theory of Planned Behavior, encompassing instrumental attitude, affective attitude, subjective norm, and perceived behavioral control. It has a reliability that ranges from 0.67 to 0.93 [37]

[a]DCS: Decisional Conflict Scale.

[b]DDS: Diabetes Distress Scale.

[c]CESDS: Center for Epidemiological Studies Depression Scale.

[d]SF: Short Form.

[e]PACIC: Patient Assessment of Chronic Illness Care.

[f]IPSDM: interprofessional shared decision-making.

[g]CPD: Continuing Professional Development.

Sample Size Calculation

With at least 40 patients per physician, 50% patient participation rate, and an anticipated patient attrition rate of 25% [38], we estimated that approximately 15 patients per practice would be able to participate. On the basis of a previous study using the DCS [7], we selected a standardized effect size of 0.4 with an SD of 0.6, α of .05, and β of .10. Previous data have shown that ρ (intraclass coefficient) for decisional conflict for patients living with diabetes clustered within primary care physicians is 0.013 [39]. Therefore, accounting for clustering, 56 patients per intervention/control group, or 4 sites per intervention/control group would be required.

Randomization

Practices were simultaneously randomized and allocated by a biostatistician to either intervention or control using computer-generated randomization in a 1:1 ratio. Each practice was given a code, and the biostatistician was blinded to the allocation status. After assignment, investigators, research coordinators, and trial participants were no longer blinded to group allocation owing to the nature of the intervention. The list of all eligible patients from each cluster was randomly

ordered; patients were recruited from this list until the target sample size was met.

Data Collection

The practice and sociodemographic characteristics of clinicians and patients were obtained at baseline. Outcome data were collected using participant questionnaires [23] distributed through web or by mail according to patient preference for pragmatic reasons at baseline, 6 months, and 12 months.

Analysis

A modified intent-to-treat analysis was conducted. For the primary and secondary outcomes, a linear mixed effect model was used to analyze the total score for each scale where site was the random effect with adjustment for baseline value. To account for missing data, we conducted a fully adjusted mixed effect model using repeated measurements [40-42]. The baseline variables we adjusted for were baseline DCS score, age, sex, ethnicity, educational attainment, employment, and living arrangements as well as a history of cancer and heart, musculoskeletal, respiratory, mental health, kidney, eye, and nerve disease.

The impact of sociodemographic variables on these outcomes (decisional conflict, diabetes distress, health-related quality of life, and chronic illness care) was assessed. Specifically, we fit a main effects model that adjusted for age, sex, ethnicity, education, employment status, and living arrangements as well as a fully adjusted model that included all interactions between treatment and the preceding variables. As these subgroup analyses were explorative, these were performed on the complete case data because the use of repeated measure data would necessitate the need for interactions with time as well. *P* values for the treatment effect in the baseline adjusted models used Satterthwaite approximation for the denominator degrees of freedom, whereas the tests of interactions (subgroup effects) used likelihood ratio tests from a full maximum likelihood estimation [43]. Irrespective of the test result on subgroups, the treatment effects were then shown by subgroup, estimated from the second model specified above, along with 95% CI and a *P* value that tested each interaction in this fully adjusted model. Descriptive statistics were used to describe *the intervention effect by MyDiabetesPlan* use. Analysis of variance was used to assess differences in the clinician's intention to practice IPSDM scores between the intervention and control groups at baseline, 6 months, and 12 months. Analysis was performed in R version 3.5.2 [44], and the packages lme4 (version 1.1-21) and lmerTest (version 3.1-0) [43] were used to fit and report the mixed effect models.

Research Ethics

The study was approved by the Research Ethics Boards of Bridgepoint Health (15-009-BP), Markham Stouffville Hospital (Canadian Institutes of Health Research [CIHR] protocol, v1, January 2013), Michael Garron Hospital (609-1410-Mis-245), North York General Hospital (13-0265), Southlake Regional Health Centre (0055-1314), St. Michael's Hospital (13-014; includes Humber River Regional Hospital), Sunnybrook Health Sciences Health Centre (345-2013), University Health Network (16-6044), and Women's College Hospital (2013-2058, 2014-0043-B).

Funding

The study was funded by the CIHR Knowledge to Action Operating Grant (Funding reference number KAL 290086), which had no role in the design, collection, management, analysis or interpretation of data, or in the writing or publication of the manuscript. SS is supported by a Tier 1 Canada Research Chair. NI was supported by New Investigator Awards from CIHR and from the Department of Family and Community Medicine, University of Toronto.

Results

Setting and Participants

A total of 10 primary care practice groups were recruited from December 2014 to November 2015; patients were recruited from December 2015 to September 2016, followed by October 2016 to September 2017. Recruitment metrics and the CONSORT flow diagram are shown in Figure 1. The practice and sociodemographic characteristics of clinicians and patients are shown in Table 2. In the intervention group, 50.0% (51/102) and 46% (33/72) of participants completed the questionnaire via web-based platform for time points 1 and 3, respectively. In the control group, 47.7% (53/111) and 51% (40/79) of participants completed the questionnaire via web-based platform for time points 1 and 3, respectively (Multimedia Appendix 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. DM: diabetes mellitus; FHT: family health team.

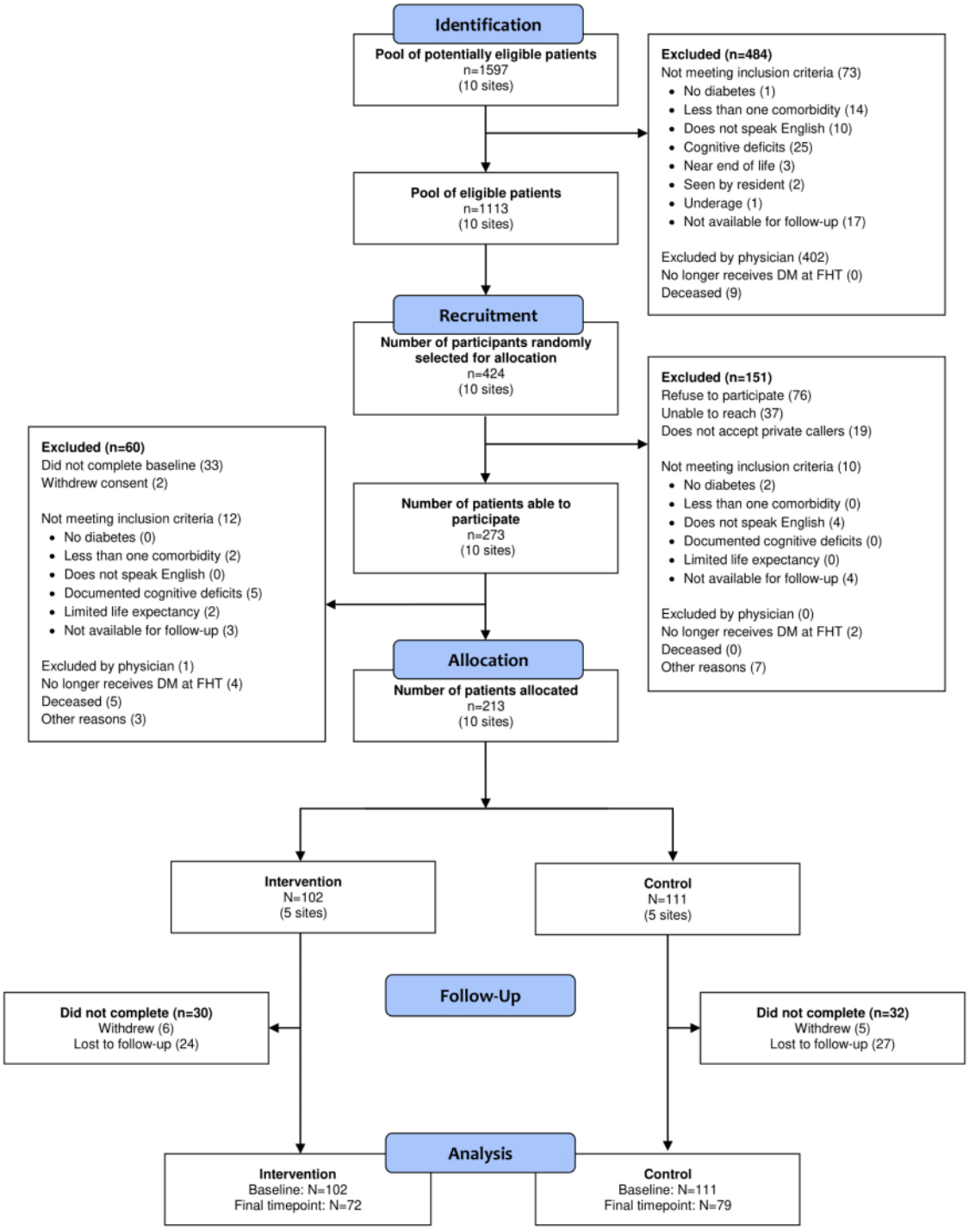


Table 2. Clinician and patient characteristics.

Characteristics	Control, n (%)	Intervention, n (%)
Clinician	24 (100)	29 (100)
Sex at birth		
Female	11 (46)	21 (72)
Male	10 (42)	7 (25)
Prefer not to answer	3 (12)	1 (3)
Duration in practice (years)		
2-5	7 (29)	5 (17)
6-10	5 (21)	8 (28)
≥11	12 (50)	16 (55)
Number of patients with diabetes seen per week		
<10	12 (50)	17 (59)
≥10	12 (50)	10 (34)
Unsure	0 (0)	2 (7)
Type of health team		
Community	20 (83)	5 (17)
Academic	4 (17)	24 (83)
Patients	111 (100)	102 (100)
Age (years)		
18-44	7 (6.3)	2 (2.0)
45-54	9 (8.1)	11 (11.0)
55-64	28 (25.2)	20 (20.0)
65-74	38 (34.2)	47 (47.0)
75-84	24 (21.6)	16 (16.0)
≥85	5 (4.5)	4 (4.0)
Sex at birth		
Female	46 (41.4)	56 (54.9)
Male	65 (58.6)	46 (45.1)
Language		
English	103 (92.8)	81 (81.0)
Other	8 (7.2)	19 (19.0)
Ethnicity		
White	75 (67.6)	62 (63.3)
Black	8 (7.2)	5 (5.1)
Asian	8 (7.2)	19 (18.6)
Indigenous	3 (2.7)	4 (4.1)
Latin American	2 (1.8)	1 (1.0)
Other	15 (13.5)	7 (7.1)
Education		
Bachelor's	17 (16.0)	23 (23.2)
Below bachelor	5 (4.7)	3 (3.0)
College	26 (24.5)	30 (30.3)
High school	31 (29.2)	19 (19.2)

Characteristics	Control, n (%)	Intervention, n (%)
Postgraduation	12 (11.3)	13 (13.1)
Below high school	15 (1.2)	11 (11.1)
Employment		
Retired	63 (58.3)	54 (55.1)
Full time with employee health benefits	15 (13.9)	22 (22.4)
Full time/part time without employee health benefits	8 (7.4)	8 (8.2)
Government assistance/disability	6 (5.6)	3 (3.1)
Unemployed	5 (4.6)	2 (2.0)
Stay-at-home parent, student, volunteer	5 (4.6)	2 (2.0)
Other	4 (3.7)	7 (7.1)
Prefer not to answer	2 (1.9)	0 (0.0)
Income, Can $ (US $)		
<10,000 (7603)	9 (8.7)	6 (7.6)
10,000-19,000 (7603-14,446)	18 (17.5)	6 (7.6)
20,000-29,000 (15,206-22,048)	8 (7.8)	5 (6.3)
30,000-39,000 (22,809-29,651)	13 (12.6)	7 (8.9)
40,000-49,000 (22,543-37,254)	10 (9.7)	7 (8.9)
50,000-59,000 (38,015-44,857)	8 (7.8)	5 (6.3)
60,000-69,000 (45,617-52,460)	3 (2.9)	6 (7.6)
70,000-79,000 (53,220-60,063)	6 (5.8)	6 (7.6)
80,000-89,000 (60,823-67,666)	2 (1.9)	6 (7.6)
90,000-99,000 (68,426-75,269)	8 (7.8)	6 (7.6)
100,000-149,000 (76,029-113,283)	7 (6.8)	8 (10.1)
≥150,000 (114,044)	11 (10.7)	11 (13.9)
Living arrangements		
Alone	30 (27.3)	26 (25.7)
With family members	24 (21.8)	28 (27.7)
With partner/spouse	46 (41.8)	38 (37.6)
With roommates	2 (1.8)	3 (3.0)
Other	8 (7.3)	6 (5.9)

Attrition Analysis

A total of 62 patients withdrew or were lost to follow-up. Of these patients, we had demographic data on 34 patients (control, n=10; intervention, n=24; Multimedia Appendix 1). There were proportionately more non–English-speaking patients (9/34, 26%) with high school education (11/34, 32%) who withdrew, compared with those who remained in the study.

Primary Outcome: Decisional Conflict

Total decisional conflict was modestly reduced in the intervention group at 12 months compared with the control group (−3.5 of a total score of 100, 95% CI −7.4 to 0.4, *P*=.08; Table 3). At 12 months, the *Uninformed* subscale was reduced in the intervention group (−3.9, 95% CI −8.8 to −1.02, *P*=.11). Similarly, at 12 months, the *Unclear Values* subscale was reduced in the intervention group (−3.6, 95% CI −9.6 to 2.28, *P*=.21).

Secondary Outcomes

Patient Chronic Care, Diabetes Distress, and Quality of Life

Patient assessment of chronic illness care increased in the intervention group compared with the control group (0.7 of a total score of 5, 95% CI 0.4 to 1.0, *P*<.001). There was a small difference in diabetes distress (−0.2, 95% CI −0.4 to 0.05, *P*=.12) and quality of life (1.2, 95% CI −3.2 to 5.5, *P*=.57; Table 3).

Table 3. Scores at baseline, 6 and 12 months, and treatment effect at 6 and 12 months for decisional conflict, patient assessment of chronic illness care, diabetes distress, and quality of life.

Outcome measures	Score, mean (SD)						Treatment effect			
	Control			Intervention			6 months		12 months	
	Baseline	6 months	12 months	Baseline	6 months	12 months	Mean, 95% CI	*P* value	Mean, 95% CI	*P* value
DCS[a] (out of 100; higher score represents more decisional conflict)	23.56 (15.00)	21.10 (12.79)	19.58 (9.11)	25.53 (14.73)	21.97 (14.87)	17.35 (11.21)	−1.82, −6.02 to 2.38	.38	−3.49, −7.4 to −0.42	.08
PACIC[b] (out of 5)	3.16 (0.95)	3.41 (1.05)	3.22 (1.08)	2.82 (1.10)	3.16 (1.10)	3.68 (0.99)	0.15, −0.19 to 0.50	.35	0.71, 0.38 to 1.04	<.001
DDS[c] (out of 6)	1.93 (0.83)	1.88 (0.78)	1.90 (0.75)	2.08 (1.02)	1.92 (1.09)	1.86 (0.87)	−0.08, −0.34 to 0.18	.53	−0.18, −0.42 to 0.05	.12
Quality of Life (SF[d]-12; out of 100)	89.69 (12.48)	87.77 (12.87)	86.99 (10.69)	87.35 (14.25)	88.88 (13.56)	87.94 (12.87)	3.47, −1.05 to 7.98	.12	1.18, −3.18 to 5.54	.57

[a]DCS: Decisional Conflict Scale.

[b]PACIC: Patient Assessment of Chronic Illness Care.

[c]DDS: Diabetes Distress Scale.

[d]SF: Short Form.

Intervention Effect of MyDiabetesPlan

A total of 52 patients completed (eg, *generated a complete plan)* *MyDiabetesPlan* two or more times during the study period. The greatest reduction in decisional conflict occurred in patients who completed *MyDiabetesPlan* more than 2 times, whereas the smallest reduction in decisional conflict occurred in patients who completed it 2 times (Table 4).

Table 4. Change in decisional conflict over 12 months in the intervention group using MyDiabetesPlan by number of completed plans.

Completed plans, n	Participants, n (N=68)	Decisional conflict score, mean (SD)		
		0 months	12 months	Change in score
<1	6	23.4 (22.7)	14.9 (11.8)	−8.5 (22.4)
1	20	23.2 (12.9)	13.9 (11.6)	−9.3 (10.6)
2	29	26.9 (10.9)	21.2 (11.5)	−5.7 (12.5)
>2	13	32.1 (22.5)	19.8 (14.6)	−12.3 (20.9)

Subgroup Analyses

For decisional conflict, we found weak evidence for any interaction (*P*=.07) that appeared to be driven by age>65 years (*P*=.01; Multimedia Appendix 1). We found stronger evidence when we examined the "uninformed" subscale (*P*=.03), driven by age>65 years (*P*=.003), and income>Can $50,000 (US $379,000; *P*=.04). Similarly, for diabetes distress, there was weak evidence for interaction (*P*=.14) driven by age>65 years (*P*=.01) and unemployment status (*P*=.03). There was weak evidence for interactions for PACIC (*P*=.01; lives alone: *P*=.11). There was little evidence of discernible interactions for the Short Form-12.

Intention to Engage in IPSDM

There was little evidence of differences between the 2 groups of clinicians in intention to practice SDM (Multimedia Appendix 1).

Harms

There were no harms associated with participation in this study.

Discussion

Principal Findings

We found that *MyDiabetesPlan*, an interprofessional goal-setting decision aid for people living with diabetes, resulted in a modest reduction in decisional conflict (specifically the *uninformed* subscale) and increased patient assessment of chronic illness care but had no impact on diabetes distress or health-related quality of life. *MyDiabetesPlan* reduced decisional conflict and diabetes distress most prominently in participants older than 65 years. There was no impact of *MyDiabetesPlan* use on clinicians' intention to practice SDM.

Our finding regarding the impact of *MyDiabetesPlan* on decisional conflict is generally consistent with the literature, although our results did not meet statistical significance. A 2017

systematic review found that decision aids reduced decisional conflict related to uncertainty caused by unclear values (−8.81/100; 95% CI −11.99 to −5.63; 23 studies; n=5068; high-quality evidence) and feeling uninformed (−9.28/100; 95% CI −12.20 to −6.36; 27 studies; n=5707; high-quality evidence) [45]. In our study, we found smaller reductions: a 3.5-point reduction in the total scale, a 3.9-point reduction in the *uninformed* subscale, and a 3.7-point reduction in the *unclear values* subscale. This discrepancy may be due to the goal-setting nature of *MyDiabetesPlan*: rather than offering risks and benefits of a single discrete decision, it offered risks and benefits of multiple potential strategies. Thus, answering a question in this context, such as "I am clear about which benefits matter most to me" may be more challenging.

Comparison With Previous Work

Consistent with a previous review [45], we also found no impact of *MyDiabetesPlan* on general or condition-specific health-related quality of life such as diabetes distress (though there was a signal for an effect in participants older than 65 years). Authors have postulated that this may be owing to the fact that decision aids are often used in situations where the options have no clear health outcome advantage [45]. A subanalysis of this review identified 11 studies involving 2684 patients that examined the impact of PtDA on health-related quality of life [46]. Of the 11 trials, 6 trials neither reported difference in health-related quality of life between PtDA and control nor over time. This confirmed the lack of impact, suggesting that health-related quality of life may be an uninformative end point unless a specific hypothesis for its impact can be made [46]. Another potential reason for this is that the outcome is too distant from the act of using *MyDiabetesPlan*: patients may select a goal (such as increasing physical activity) but may not enact the goal (ie, *actually increasing their physical activity*), and *thus may not experience any change in health-related quality of life.* Although *MyDiabetesPlan* may have reduced decisional conflict and increased patient activation and thus increased goal setting, there may be a gap between goal setting and goal attainment; bridging this gap with additional behavioral supports to optimize goal attainment may then result in improvements in health-related quality of life.

Our finding that *MyDiabetesPlan* reduced decisional conflict and diabetes distress, particularly in participants older than 65 years, is consistent with the literature [47]. This systematic review of 22 studies examining the impact of PtDa in adults aged 65 years and older found that people exposed to a decision aid had greater knowledge (5 studies; mean difference 6.5, 95% CI 0.76 to 12.25) and reduced decisional conflict (11 studies; mean difference −3.17 out of 100, 95% CI −4.44 to 1.90). These findings are reassuring, particularly in light of concerns of social inequities and the digital health divide [48], and provide support that a digital health innovation such as *MyDiabetesPlan* is appropriate for a complex older population with multiple comorbidities.

The strengths of this study include rigorous adherence to RCT methodology (including central randomization, similar intervention and control groups at baseline, appropriate length of follow-up, use of validated scales, intention-to-treat analysis), its finding of reducing decisional conflict and improving chronic care delivery, especially in those older than 65 years, and its generalizability. We designed this study to assess *the feasibility of implementing MyDiabetesPlan* in interprofessional primary care settings; as such, it is primarily a pragmatic trial along the explanatory-pragmatic continuum [49]. Thus, our study results mainly reflect what would happen if *MyDiabetesPlan* would be implemented in the usual clinical practice of interprofessional primary care.

Limitations

Study limitations include the lack of blinding of participants (patients and clinicians) owing to the nature of the intervention, use of both paper- and web-based data collection methods, attrition rate of 29%, and less-than-anticipated *MyDiabetesPlan* use, resulting in potential bias, and its lack of clinical outcome measures. Although different data collection methods may introduce respondent bias, recent literature has shown that response rates do not differ between paper- and web-based respondents [50] and that there was no difference in socioeconomic characteristics between paper- and web-based respondents [51]. Our attrition rate and reduced *MyDiabetesPlan* use were amplified by our complex study population, many of whom either withdrew owing to competing health concerns or were lost to follow-up, presumably for similar reasons. Challenges to conducting research in this population are well documented, and solutions for future studies include providing transportation compensation, conducting home visits, and encouraging greater engagement of family members [52]. Although we did not assess clinical outcome measures, we assessed proximal patient-reported outcomes appropriate for a decision-aid intervention, outcomes that have typically been underused in the literature [53], and outcomes that have been associated with clinical outcomes such as A_{1c} [12-14]. Assessment of researcher-selected clinical outcomes such as A_{1c} in a trial where patients are encouraged to select their own personalized goals is incongruent with the principles of person-centered outcomes research [54]. Future studies may consider the use of goal attainment scaling (which has some demonstrated validity evidence in the geriatric care setting) [55] or composite clinical outcomes, though this is not without its limitations [56].

Conclusions

MyDiabetesPlan modestly reduced decisional conflict and increased patient assessment of chronic illness care but had no impact on diabetes distress or health-related quality of life. The next steps in this research program are to engage with knowledge users (patients and their caregivers, clinicians, managers, and policy makers) to discuss the implications of these findings, modify *MyDiabetesPlan* and its mode of delivery, and consideration of clinical, organizational, and health care system contexts to plan scale-up to an implementation trial.

Acknowledgments

All authors are members of the *MyDiabetesPlan* team. The authors thank our knowledge users Eva Butler, Jan Hux, Ted Daisley, Heather Whetstone, Pauline Wijeyesekera, and Art Scrannage for their invaluable insights, perspective, and time. We thank John Loach, Calvin Ke, Imran Somji, Joyce Hui, and Sasha Jovicic for assistance and input into *MyDiabetesPlan* design. We thank Sue Hall, Marc Settino, Rabiya Siddiqui from University of Toronto Practice-Based Research Network (UTOPIAN), Alanna Rigobon, Celia Cassiani, Lakmini Pinnaduwage, and Catherine Spagnuolo for assistance with site coordination. We thank all our patients, clinicians, and site administrators at our participating sites for their support and time: Bridgepoint Family Health Team (FHT), Health for All FHT, Humber River FHT, North York FHT, South East Toronto FHT, Southlake Regional Health Centre FHT, St. Michael's Hospital FHT, Sunnybrook Academic FHT, Toronto Western Hospital FHT, and Women's College Hospital FHT.

Authors' Contributions

CY conceived and designed the study and drafted the manuscript. DC and BB conducted data cleaning and drafted portions of the manuscript. KT assisted with study design, conducted statistical analyses, and contributed to manuscript development. SS assisted with study design and revised the manuscript critically for intellectual content.

PC, KC, PF, NI, AH, DK, FL, JM, JR, SS, and DT participated in the design of the study, site coordination, and revised the manuscript critically for intellectual content. JS and DS participated in the design of the study and revised the manuscript critically for intellectual content. All authors have read and approved the final manuscript.

Site leads for this were as follows: Paul Cantarutti, Southlake Regional Health Centre, Newmarket, Canada; Karen Chu, Bridgepoint Active Healthcare, Toronto, Canada; Paul Frydrych, Mount Dennis Weston Health Centre, Toronto, Canada; Noah Ivers, Family Practice Health Centre, Women's College Hospital & University of Toronto, Toronto, Canada; David Kaplan, North York Family Health Team, Toronto, Canada; Fok-Han Leung, St. Michael's Hospital, Toronto, Canada; John Maxted, Markham Stouffville Hospital, Markham, Canada; Jeremy Rezmovitz, Sunnybrook Health Sciences Centre, Toronto, Canada; Sumeet Sodhi, Toronto Western Family Health Team, University Health Network, Toronto, Canada; Deanna Telner, South East Toronto Family Health Team, Toronto, Canada. The site coordinator for this study was as follows: Amy Hoang-Kim, St. Michael's Hospital, Toronto, Canada. The content experts were as follows: Joanna Sale, Musculoskeletal Health and Outcomes Research - Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Canada; Dawn Stacey RN, PhD, School of Nursing, University of Ottawa, Senior Scientist, Ottawa Hospital Research Institute, Ottawa, Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental file including CONSORT (Consolidated Standards of Reporting Trials) checklists (cluster, eHealth) and supplemental Tables 1-4.

[PDF File (Adobe PDF File), 708 KB - jmir_v22i9e16984_app1.pdf]

References

1. Santana MJ, Manalili K, Jolley RJ, Zelinsky S, Quan H, Lu M. How to practice person-centred care: a conceptual framework. Health Expect 2018 Apr;21(2):429-440 [FREE Full text] [doi: 10.1111/hex.12640] [Medline: 29151269]
2. Williams JS, Walker RJ, Smalls BL, Hill R, Egede LE. Patient-centered care, glycemic control, diabetes self-care, and quality of life in adults with type 2 diabetes. Diabetes Technol Ther 2016 Oct;18(10):644-649 [FREE Full text] [doi: 10.1089/dia.2016.0079] [Medline: 27541872]
3. Scholl I, Zill JM, Härter M, Dirmaier J. An integrative model of patient-centeredness - a systematic review and concept analysis. PLoS One 2014;9(9):e107828 [FREE Full text] [doi: 10.1371/journal.pone.0107828] [Medline: 25229640]
4. Mullan RJ, Montori VM, Shah ND, Christianson TJH, Bryant SC, Guyatt GH, et al. The diabetes mellitus medication choice decision aid: a randomized trial. Arch Intern Med 2009 Sep 28;169(17):1560-1568. [doi: 10.1001/archinternmed.2009.293] [Medline: 19786674]
5. Rost KM, Flavin KS, Cole K, McGill JB. Change in metabolic control and functional status after hospitalization. Impact of patient activation intervention in diabetic patients. Diabetes Care 1991 Oct;14(10):881-889. [doi: 10.2337/diacare.14.10.881] [Medline: 1773686]
6. Kashaf MS, McGill ET, Berger ZD. Shared decision-making and outcomes in type 2 diabetes: a systematic review and meta-analysis. Patient Educ Couns 2017 Dec;100(12):2159-2171. [doi: 10.1016/j.pec.2017.06.030] [Medline: 28693922]
7. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making 1995;15(1):25-30. [doi: 10.1177/0272989X9501500105] [Medline: 7898294]

8. Becerra-Perez M, Menear M, Turcotte S, Labrecque M, Légaré F. More primary care patients regret health decisions if they experienced decisional conflict in the consultation: a secondary analysis of a multicenter descriptive study. BMC Fam Pract 2016 Nov 10;17(1):156 [FREE Full text] [doi: 10.1186/s12875-016-0558-0] [Medline: 27832752]
9. Kaplan AL, Crespi CM, Saucedo JD, Connor SE, Litwin MS, Saigal CS. Decisional conflict in economically disadvantaged men with newly diagnosed prostate cancer: baseline results from a shared decision-making trial. Cancer 2014 Sep 1;120(17):2721-2727 [FREE Full text] [doi: 10.1002/cncr.28755] [Medline: 24816472]
10. Solli O, Stavem K, Kristiansen IS. Health-related quality of life in diabetes: the associations of complications with EQ-5D scores. Health Qual Life Outcomes 2010 Feb 4;8:18 [FREE Full text] [doi: 10.1186/1477-7525-8-18] [Medline: 20132542]
11. Wong EM, Afshar R, Qian H, Zhang M, Elliott TG, Tang TS. Diabetes distress, depression and glycemic control in a Canadian-based specialty care setting. Can J Diabetes 2017 Aug;41(4):362-365. [doi: 10.1016/j.jcjd.2016.11.006] [Medline: 28462795]
12. Liu L, Li Y, Sha K, Wang Y, He X. Patient assessment of chronic illness care, glycemic control and the utilization of community health care among the patients with type 2 diabetes in Shanghai, China. PLoS One 2013;8(9):e73010 [FREE Full text] [doi: 10.1371/journal.pone.0073010] [Medline: 24039847]
13. Shim YT, Lee J, Toh MP, Tang WE, Ko Y. Health-related quality of life and glycaemic control in patients with type 2 diabetes mellitus in Singapore. Diabet Med 2012 Aug;29(8):e241-e248. [doi: 10.1111/j.1464-5491.2012.03689.x] [Medline: 22507291]
14. Fisher L, Polonsky WH, Hessler DM, Masharani U, Blumer I, Peters AL, et al. Understanding the sources of diabetes distress in adults with type 1 diabetes. J Diabetes Complications 2015;29(4):572-577 [FREE Full text] [doi: 10.1016/j.jdiacomp.2015.01.012] [Medline: 25765489]
15. Légaré F, Stacey D, Turcotte S, Cossi M, Kryworuchko J, Graham ID, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. Cochrane Database Syst Rev 2014 Sep 15(9):CD006732. [doi: 10.1002/14651858.CD006732.pub3] [Medline: 25222632]
16. Kim J, Whitney A, Hayter S, Lewis C, Campbell M, Sutherland L, et al. Development and initial testing of a computer-based patient decision aid to promote colorectal cancer screening for primary care practice. BMC Med Inform Decis Mak 2005 Nov 28;5:36 [FREE Full text] [doi: 10.1186/1472-6947-5-36] [Medline: 16313676]
17. Bailey RA, Pfeifer M, Shillington AC, Harshaw Q, Funnell MM, VanWingen J, et al. Effect of a patient decision aid (PDA) for type 2 diabetes on knowledge, decisional self-efficacy, and decisional conflict. BMC Health Serv Res 2016 Jan 14;16:10 [FREE Full text] [doi: 10.1186/s12913-016-1262-4] [Medline: 26762150]
18. Denig P, Schuling J, Haaijer-Ruskamp F, Voorham J. Effects of a patient oriented decision aid for prioritising treatment goals in diabetes: pragmatic randomised controlled trial. Br Med J 2014 Sep 25;349:g5651 [FREE Full text] [doi: 10.1136/bmj.g5651] [Medline: 25255799]
19. Yu CH, Stacey D, Sale J, Hall S, Kaplan DM, Ivers N, et al. Designing and evaluating an interprofessional shared decision-making and goal-setting decision aid for patients with diabetes in clinical care--systematic decision aid development and study protocol. Implement Sci 2014 Jan 22;9:16 [FREE Full text] [doi: 10.1186/1748-5908-9-16] [Medline: 24450385]
20. Yu CH, Ke C, Jovicic A, Hall S, Straus SE, IP-SDM Team. Beyond pros and cons - developing a patient decision aid to cultivate dialog to build relationships: insights from a qualitative study and decision aid development. BMC Med Inform Decis Mak 2019 Sep 18;19(1):186 [FREE Full text] [doi: 10.1186/s12911-019-0898-5] [Medline: 31533828]
21. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]
22. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. Br Med J 2012 Sep 4;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
23. Yu CH, Ivers NM, Stacey D, Rezmovitz J, Telner D, Thorpe K, et al. Impact of an interprofessional shared decision-making and goal-setting decision aid for patients with diabetes on decisional conflict--study protocol for a randomized controlled trial. Trials 2015 Jun 27;16:286 [FREE Full text] [doi: 10.1186/s13063-015-0797-8] [Medline: 26116444]
24. Cheng AY, Barnes T. Executive summary: Canadian diabetes association 2013 clinical practice guidelines for the prevention and management of diabetes in Canada. Can J Diabetes 2013;37(Suppl 3):S291-S360 [FREE Full text]
25. Staying Healthy With Diabetes. Diabetes Canada | Clinical Practice Guidelines. 2012. URL: https://guidelines.diabetes.ca/docs/patient-resources/staying-healthy-with-diabetes.pdf [accessed 2020-09-14]
26. User Manual - Decision Conflict Scale. Patient Decision Aids - Ottawa Hospital Research Institute. 1993. URL: http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf [accessed 2020-09-04]
27. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. Diabetes Care 2005 Mar;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]
28. Ware J. SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1. Boston, MA: Health Assessment Lab; 1994.
29. Glasgow RE, Whitesides H, Nelson CC, King DK. Use of the patient assessment of chronic illness care (PACIC) with diabetic patients: relationship to patient characteristics, receipt of care, and self-management. Diabetes Care 2005 Nov;28(11):2655-2661. [doi: 10.2337/diacare.28.11.2655] [Medline: 16249535]

30. Francis J, Eccles M, Johnston M, Walker A, Grimshaw J, Foy R, et al. Constructing Questionnaires Based on the Theory of Planned Behavior? A Manual for Health Services Researchers. Newcastle upon Tyne, UK: Centre for Health Services Research; 2004.
31. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
32. Bohannon RW, Maljanian R, Landes M. Test-retest reliability of short form (SF)-12 component scores of patients with stroke. Int J Rehabil Res 2004 Jun;27(2):149-150. [doi: 10.1097/01.mrr.0000127350.25287.08] [Medline: 15167113]
33. Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. Qual Life Res 2009 Aug;18(6):727-735. [doi: 10.1007/s11136-009-9483-1] [Medline: 19424821]
34. Schmittdiel J, Mosen DM, Glasgow RE, Hibbard J, Remmers C, Bellows J. Patient assessment of chronic illness care (PACIC) and improved patient-centered outcomes for chronic conditions. J Gen Intern Med 2008 Jan;23(1):77-80 [FREE Full text] [doi: 10.1007/s11606-007-0452-5] [Medline: 18030539]
35. Szecsenyi J, Rosemann T, Joos S, Peters-Klimm F, Miksch A. German diabetes disease management programs are appropriate for restructuring care according to the chronic care model: an evaluation with the patient assessment of chronic illness care instrument. Diabetes Care 2008 Jun;31(6):1150-1154. [doi: 10.2337/dc07-2104] [Medline: 18299443]
36. Légaré F, Borduas F, Freitas A, Jacques A, Godin G, Luconi F, CPD-KT team. Development of a simple 12-item theory-based instrument to assess the impact of continuing professional development on clinical behavioral intentions. PLoS One 2014;9(3):e91013 [FREE Full text] [doi: 10.1371/journal.pone.0091013] [Medline: 24643173]
37. Légaré F, Guerrier M, Nadeau C, Rhéaume C, Turcotte S, Labrecque M. Impact of DECISION + 2 on patient and physician assessment of shared decision making implementation in the context of antibiotics use for acute respiratory infections. Implement Sci 2013 Dec 26;8:144 [FREE Full text] [doi: 10.1186/1748-5908-8-144] [Medline: 24369771]
38. Heneghan C, Perera R, Ward A, Fitzmaurice D, Meats E, Glasziou P. Assessing differential attrition in clinical trials: self-monitoring of oral anticoagulation and type II diabetes. BMC Med Res Methodol 2007 May 2;7:18 [FREE Full text] [doi: 10.1186/1471-2288-7-18] [Medline: 17474976]
39. Légaré F, O'Connor AM, Graham ID, Wells GA, Tremblay S. Impact of the Ottawa decision support framework on the agreement and the difference between patients' and physicians' decisional conflict. Med Decis Making 2006;26(4):373-390. [doi: 10.1177/0272989X06290492] [Medline: 16855126]
40. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 6;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
41. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J Clin Epidemiol 2013 Sep;66(9):1022-1028. [doi: 10.1016/j.jclinepi.2013.03.017] [Medline: 23790725]
42. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods 2001 Dec;6(4):330-351. [Medline: 11778676]
43. Kuznetsova A, Brockhoff PB, Christensen RH. Package: tests in linear mixed effects models. J Stat Soft 2017;82(13):-. [doi: 10.18637/jss.v082.i13]
44. The R Project for Statistical Computing. 2018. URL: https://www.R-project.org/ [accessed 2019-05-03]
45. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
46. Rutherford C, King MT, Butow P, Legare F, Lyddiatt A, Souli I, et al. Is quality of life a suitable measure of patient decision aid effectiveness? Sub-analysis of a Cochrane systematic review. Qual Life Res 2019 Mar;28(3):593-607. [doi: 10.1007/s11136-018-2045-7] [Medline: 30426276]
47. van Weert JC, van Munster BC, Sanders R, Spijker R, Hooft L, Jansen J. Decision aids to help older people make health decisions: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2016 Apr 21;16:45 [FREE Full text] [doi: 10.1186/s12911-016-0281-8] [Medline: 27098100]
48. Hall AK, Bernhardt JM, Dodd V, Vollrath MW. The digital health divide: evaluating online health information access and use among older adults. Health Educ Behav 2015 Apr;42(2):202-209 [FREE Full text] [doi: 10.1177/1090198114547815] [Medline: 25156311]
49. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. Br Med J 2015 May 8;350:h2147. [doi: 10.1136/bmj.h2147] [Medline: 25956159]
50. Horevoorts NJ, Vissers PA, Mols F, Thong MS, van de Poll-Franse LV. Response rates for patient-reported outcomes using web-based versus paper questionnaires: comparison of two invitational methods in older colorectal cancer patients. J Med Internet Res 2015 May 7;17(5):e111 [FREE Full text] [doi: 10.2196/jmir.3741] [Medline: 25953059]
51. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. J Med Internet Res 2018 Jan 23;20(1):e24 [FREE Full text] [doi: 10.2196/jmir.8353] [Medline: 29362206]

52. Knechel NA. The challenges of enrolling older adults into intervention studies. Yale J Biol Med 2013 Mar;86(1):41-47 [FREE Full text] [Medline: 23482244]
53. Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. Br Med J 2010 Nov 1;341:c5707 [FREE Full text] [doi: 10.1136/bmj.c5707] [Medline: 21041324]
54. Costlow MR, Landsittel DP, James AE, Kahn JM, Morton SC. Model for a patient-centered comparative effectiveness research center. Clin Transl Sci 2015 Apr;8(2):155-159 [FREE Full text] [doi: 10.1111/cts.12257] [Medline: 25588873]
55. Stolee P, Rockwood K, Fox RA, Streiner DL. The use of goal attainment scaling in a geriatric care setting. J Am Geriatr Soc 1992 Jun;40(6):574-578. [doi: 10.1111/j.1532-5415.1992.tb02105.x] [Medline: 1587973]
56. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? J Am Med Assoc 2003 May 21;289(19):2554-2559. [doi: 10.1001/jama.289.19.2554] [Medline: 12759327]

Abbreviations

CIHR: Canadian Institutes of Health Research
CONSORT: Consolidated Standards of Reporting Trials
DCS: Decisional Conflict Scale
EMR: electronic medical record
FHT: Family Health Team
IPSDM: interprofessional shared decision-making
PtDA: patient decision aid
RCT: randomized controlled trial
SDM: shared decision making

Edited by G Eysenbach; submitted 15.11.19; peer-reviewed by J Dirmaier, K Eden; comments to author 10.03.20; revised version received 04.05.20; accepted 11.08.20; published 30.09.20.

Please cite as:
Yu C, Choi D, Bruno BA, Thorpe KE, Straus SE, Cantarutti P, Chu K, Frydrych P, Hoang-Kim A, Ivers N, Kaplan D, Leung FH, Maxted J, Rezmovitz J, Sale J, Sodhi-Helou S, Stacey D, Telner D
Impact of MyDiabetesPlan, a Web-Based Patient Decision Aid on Decisional Conflict, Diabetes Distress, Quality of Life, and Chronic Illness Care in Patients With Diabetes: Cluster Randomized Controlled Trial
J Med Internet Res 2020;22(9):e16984
URL: http://www.jmir.org/2020/9/e16984/
doi:10.2196/16984
PMID:32996893

Original Paper

Patient Portal Barriers and Group Differences: Cross-Sectional National Survey Study

Kea Turner[1*], MPH, MA, PhD; Alecia Clary[2*], MSW, PhD; Young-Rock Hong[3], MPH, PhD; Amir Alishahi Tabriz[4*], MPH, MD, PhD; Christopher M Shea[5*], MA, MPA, PhD

[1]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States

[2]Center for Healthcare Transformation, Avalere Health, Washington, DC, United States

[3]Department of Health Services Research, Management, and Policy, College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

[4]Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[*]these authors contributed equally

Corresponding Author:
Kea Turner, MPH, MA, PhD
Department of Health Outcomes and Behavior
Moffitt Cancer Center
12902 USF Magnolia Drive, MRC-CANCONT
Tampa, FL, 33612
United States
Phone: 1 813 745 5213
Email: kea.turner@gmail.com

Abstract

Background: Past studies examining barriers to patient portal adoption have been conducted with a small number of patients and health care settings, limiting generalizability.

Objective: This study had the following two objectives: (1) to assess the prevalence of barriers to patient portal adoption among nonadopters and (2) to examine the association between nonadopter characteristics and reported barriers in a nationally representative sample.

Methods: Data from this study were obtained from the 2019 Health Information National Trends Survey. We calculated descriptive statistics to determine the most prevalent barriers and conducted multiple variable logistic regression analysis to examine which characteristics were associated with the reported barriers.

Results: The sample included 4815 individuals. Among these, 2828 individuals (58.73%) had not adopted a patient portal. Among the nonadopters (n=2828), the most prevalent barriers were patient preference for in-person communication (1810/2828, 64.00%), no perceived need for the patient portal (1385/2828, 48.97%), and lack of comfort and experience with computers (735/2828, 25.99%). Less commonly, individuals reported having no patient portal (650/2828, 22.98%), no internet access (650/2828, 22.98%), privacy concerns (594/2828, 21.00%), difficulty logging on (537/2828, 18.99%), and multiple patient portals (255/2828, 9.02%) as barriers. Men had significantly lower odds of indicating a preference for speaking directly to a provider compared with women (odds ratio [OR] 0.75, 95% CI 0.60-0.94; *P*=.01). Older age (OR 1.01, 95% CI 1.00-1.02; *P*<.001), having a chronic condition (OR 1.83, 95% CI 1.44-2.33; *P*<.001), and having an income lower than US $20,000 (OR 1.61, 95% CI 1.11-2.34; *P*=.01) were positively associated with indicating a preference for speaking directly to a provider. Hispanic individuals had significantly higher odds of indicating that they had no need for a patient portal (OR 1.59, 95% CI 1.24-2.05; *P*<.001) compared with non-Hispanic individuals. Older individuals (OR 1.05, 95% CI 1.04-1.06; *P*<.001), individuals with less than a high school diploma (OR 3.15, 95% CI 1.79-5.53; *P*<.001), and individuals with a household income of less than US $20,000 (OR 2.78, 95% CI 1.88-4.11; *P*<.001) had significantly higher odds of indicating that they were uncomfortable with a computer.

Conclusions: The most common barriers to patient portal adoption are preference for in-person communication, not having a need for the patient portal, and feeling uncomfortable with computers, which are barriers that are modifiable and can be intervened

upon. Patient characteristics can help predict which patients are most likely to experience certain barriers to patient portal adoption. Further research is needed to tailor implementation approaches based on patients' needs and preferences.

(J Med Internet Res 2020;22(9):e18870) doi:10.2196/18870

KEYWORDS

patient portal; personal health record; electronic health record; implementation

Introduction

Patient portals have demonstrated promise in improving patient engagement and outcomes [1-7] but remain underutilized [8-13]. Patient portals (web applications tethered to the electronic health record [EHR]) offer patients numerous ways to better engage in their own care, such as viewing and downloading health information and securely messaging their health care providers [14-16]. To encourage patient portal use among patients and providers, the US government has taken steps to promote patient portal adoption [17,18]. Stage 2 of the Meaningful Use EHR Incentive Program, which is now part of the Merit-based Incentive Payment Program, requires eligible providers to ensure that a certain percentage of patients are downloading and viewing health information and securely messaging their care team [17,19]. Eligible hospitals are also required to ensure that patients download and view health information under the Promoting Interoperability for Hospitals Program. Additionally, the 21st Century Cures Act encourages providers to offer access to patient portals through digital health applications in order to enhance a patient's ability to maintain a longitudinal health record (ie, integrate portal data from multiple providers) and to share the record with other health care providers [18]. Despite these policy initiatives, patient portal adoption has been slow.

Studies have consistently shown that patient portal usage has increased over time but remains low overall [10,20,21]. A recent nationally representative study (weighted n=254,183 individuals) found a significant increase in the adoption of patient portals in the United States, from 12.5% in 2011 to 25.0% in 2017 (P<.001) [10]. However, an overall adoption rate of 25% is modest and means that many patients are still not using patient portals. To increase adoption, some health care systems have started offering access to patient portals through smartphone apps [22]. Smartphone app access allows patients to integrate data from multiple patient portals and may increase access for patients who do not have access to a home computer. However, a recent study examining patient portal access through a smartphone app found that the rate of new users did not significantly change over time (P=.18) and that the proportion of patient portal adopters who logged into the smartphone app was low (population mean [] 0.7%, SD 0.2%-2.1%) [23]. This finding suggests that there are other barriers that are affecting patient portal adoption and that additional implementation strategies, beyond accessibility through smartphone apps, may be needed to enhance adoption.

Several studies have identified patient- and provider-level barriers that are associated with low adoption of patient portals. Studies have consistently shown that lower socioeconomic status, older age, rural residence, male gender, black race, Hispanic ethnicity, and public or no insurance are associated with lower adoption of patient portals [8,24-30]. On the other hand, patients with a usual source of care, those having better patient-provider communication, and those with multiple chronic conditions are more likely to adopt patient portals [29,31,32]. Studies have also enumerated many barriers to adoption, such as computer literacy, lack of internet access, privacy concerns, difficulty logging in, and presence of different portals for different providers [33-43]. Many of these studies, however, involved small samples, limiting the ability to discern which barriers are most prominent and which patient subgroups are most likely to experience a specific barrier.

To address this gap, this study had the following two objectives: (1) to assess the prevalence of barriers to patient portal adoption among nonadopters and (2) to examine the association between nonadopter characteristics and reported barriers in a nationally representative sample. By clarifying which barriers are most common and which patient subgroups are most affected by these barriers, future studies can develop targeted implementation approaches to advance patient portal adoption.

Methods

Study Design

This was a cross-sectional observational study conducted in 2019. The unit of analysis was the individual.

Data

Data for this study were obtained from the Health Information National Trends Survey (HINTS), which is administered by the National Cancer Institute (NCI). The HINTS collects data on individuals' use of and access to health-related information, and health-related knowledge, awareness, and behaviors. The sampling frame included all civilian noninstitutionalized adults (aged over 18 years) living in the United States, and it was considered a nationally representative sample. The sampling strategy was two-staged. First, a stratified sample was selected based on a file of residential addresses maintained by the NCI, and then, one adult within each selected household was sampled. The HINTS 5, Cycle 3 survey, which was used for this study, was administered from January through April 2019 via a paper-based survey and an experimental web survey. The overall response rate was 30.3%.

Study Population

The HINTS 5, Cycle 3 sample included 5438 individuals. We removed individuals with missing data for key variables (eg, complete case analysis) since the rate of missingness was less than 7% for study variables. We examined whether individuals with missing data were more likely to report nonadoption of patient portals and did not find a relationship; therefore, we concluded that the data were missing at random. We excluded

individuals who reported not visiting a health care provider in the past 12 months (n=542). After removing individuals with missing data and individuals who had not visited a provider in the past 12 months, the analytic sample included 4815 individuals (weighted n=227,463,350).

Measures

Adoption of Patient Portals

We divided the sample based on adoption and nonadoption of patient portals according to a survey question ("How many times did you access your online medical record in the last 12 months?"). Individuals selecting zero times were categorized as "nonadopters" and individuals selecting one or more times were categorized as "adopters."

Barriers to Patient Portal Adoption

The survey asked a series of yes/no questions eliciting reasons why patients have not adopted patient portals, including patient preference to speak to a health care provider directly, lack of internet access, concerns about privacy, lack of patient portals, trouble remembering passwords, lack of experience with computers, and having more than one patient portal. Participants were allowed to select more than one barrier. Each of these barriers was categorized as a binary variable.

Individual Characteristics

The survey captured several measures of patient characteristics. We included characteristics associated with patient portal adoption that have been reported in previous studies [8,24-30], including binary measures of gender, black race, Hispanic ethnicity, marital status, insurance status, rural residence, presence of a chronic condition (eg, diabetes, hypertension, and heart disease), and having a regular provider. We included age as a continuously measured variable. We also included three categorical variables, including income (eg, less than US $20,000, US $20,000-$34,999, US $35,000-$49,999, US $50,000-$74,999, and US $75,000 or more), education (eg, less than high school diploma, high school diploma, college degree, and postgraduate degree), and satisfaction with care (eg, excellent, very good, good, fair, and poor). Each of these variables has been shown to be a predictor of patient portal adoption in previous research [8,24-30].

Analytic Approach

First, we described the characteristics of the study population, conducting bivariate analyses to compare the characteristics of the adopters and nonadopters of patient portals. Second, we adopted a series of multiple logistic regression models to examine which characteristics were associated with barriers that were experienced by at least 10% of participants. We chose 10% as a cutoff to ensure that we had an adequate sample size for each logistic regression. We present the results for the three most common barriers in the Results section and provide the results for the less common barriers in Multimedia Appendix 1 and Multimedia Appendix 2. The statistical analyses were conducted using Stata (version 16; StataCorp). We used jackknife replicate weights to account for the complex survey design (ie, stratified cluster sample) in the variance calculations. We also applied sampling weights to develop nationally representative estimates. We adhered to the guidelines for weighting and variance estimation from the NCI [44]. To ensure adequate reporting of our study, we followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [45]. This study was exempted by the Advarra Institutional Review Board owing to the use of publicly available data.

Results

Sample Characteristics

The sample included 4815 individuals (weighted n=227,463,350). The majority of participants were female (2746/4815, 57.03%), white (3755/4815, 77.99%), and non-Hispanic (4024/4815, 83.57%) (Table 1). The mean age was 56.3 years (SD 16.7 years). Less than half of the participants had a college degree (1307/4815, 27.14%) or postgraduate degree (963/4815, 20.00%) and a household income of US $75,000 or more per year (1719/4815, 35.70%).

There were more individuals who had not adopted a patient portal (ie, nonadopters; 2828/4815, 58.73%) than individuals who had adopted a patient portal (ie, adopters; 1987/4815, 41.27%) (Table 1). Nonadopters were significantly more likely to be male (P<.001), be of black race (P=.001), have Hispanic ethnicity (P=.02), have lower education attainment (P<.001), and have lower income (P<.001). Nonadopters were also significantly more likely to live in a rural area (P=.002), be unmarried (P=.001), be uninsured (P=.001), not have a usual source of care (P<.001), and rate their quality of care lower (P<.001) compared with adopters. Age and having a chronic condition did not significantly impact the decision to adopt a patient portal.

Table 1. Sample characteristics.

Characteristic	Total population (N=4815; weighted: 227,463,350)	Adopters (N=1987; weighted: 90,644,145)	Nonadopters (N=2828; weighted: 136,800,000)	*P* value
Gender, n (%)				<.001
Male	2069 (42.97%)	781 (39.31%)	1288 (45.54%)	
Female	2746 (57.03%)	1206 (60.69%)	1540 (54.46%)	
Race, n (%)				.001
Black	772 (16.03%)	255 (12.83%)	517 (18.28%)	
White	3755 (77.99%)	1649 (82.99%)	2106 (74.47%)	
Ethnicity, n (%)				.02
Hispanic	791 (16.43%)	265 (13.34%)	526 (18.60%)	
Non-Hispanic	4024 (83.57%)	1722 (86.66%)	2302 (81.40%)	
Age, μ[a] (SD)	56.3 (16.79)	54.1 (16.01)	57.8 (17.03)	.08
Education, n (%)				<.001
Less than HS[b]	280 (5.82%)	28 (1.41%)	252 (8.91%)	
HS diploma	2265 (47.04%)	739 (37.19%)	1526 (53.96%)	
College degree	1307 (27.14%)	651 (32.76%)	656 (23.20%)	
Postgraduate degree	963 (20.00%)	569 (28.64%)	394 (13.93%)	
Income, n (%)				<.001
Less than US $20,000	762 (15.83%)	154 (7.75%)	608 (21.50%)	
US $20,000 to $34,999	547 (11.36%)	164 (8.25%)	383 (13.54%)	
US $35,000 to $49,999	575 (11.94%)	233 (11.73%)	342 (12.09%)	
US $50,000 to $74,999	786 (16.32%)	363 (18.27%)	423 (14.96%)	
US $75,000 or more	1719 (35.70%)	929 (46.75%)	790 (27.93%)	
Rural, n (%)				.002
Yes	521 (10.82%)	169 (8.51%)	352 (12.48%)	
No	4294 (89.18%)	1818 (91.49%)	2476 (87.55%)	
Marital status, n (%)				.001
Married	2656 (55.16%)	1258 (63.31%)	1398 (49.43%)	
Unmarried	2159 (44.84%)	729 (36.69%)	1430 (50.57%)	
Chronic condition, n (%)				.06
Yes	2633 (54.68%)	1051 (52.89%)	1582 (55.94%)	
No	2182 (45.32%)	936 (47.11%)	1246 (44.06%)	
Insurance status, n (%)				.001
Insured	4559 (94.68%)	1941 (97.68%)	2618 (92.57%)	
Uninsured	256 (5.32%)	46 (2.32%)	210 (7.42%)	
Regular provider, n (%)				<.001
Yes	3392 (70.45%)	1614 (81.23%)	1778 (62.87%)	
No	1423 (29.55%)	373 (18.77)	1050 (37.13%)	
Quality of care, n (%)				<.001
Excellent	1829 (37.99%)	764 (38.45%)	1065 (37.66%)	
Very good	1806 (37.51%)	797 (40.11%)	1009 (35.68%)	
Good	862 (17.90%)	344 (17.31%)	518 (18.32%)	
Fair	197 (4.91%)	70 (3.52%)	127 (4.49%)	

Characteristic	Total population (N=4815; weighted: 227,463,350)	Adopters (N=1987; weighted: 90,644,145)	Nonadopters (N=2828; weighted: 136,800,000)	*P* value
Poor	120 (2.49%)	12 (0.60%)	108 (3.82%)	

[a]μ: population mean.

[b]HS: high school

Prevalence of Barriers to Patient Portal Adoption Among Nonadopters

Among nonadopters (n=2828), the most prevalent barrier to patient portal adoption was patient preference for in-person communication (1810/2828, 64.00%) (Table 2). The second most common barrier was no perceived need for the patient portal (1385/2828, 48.97%). The third most common barrier was lack of comfort and experience with computers (735/2828, 25.99%). Less commonly, individuals reported having no patient portal (650/2828, 22.98%), no internet access (650/2828, 22.98%), privacy concerns (594/2828, 21.00%), difficulty logging on (537/2828, 18.99%), and multiple patient portals (255/2828, 9.02%) (Multimedia Appendix 1 and Multimedia Appendix 2).

Nonadopter Characteristics and Barriers to Patient Portal Adoption

For the first barrier, men had significantly lower odds of indicating a preference for speaking directly to a provider compared with women (odds ratio [OR] 0.75, 95% CI 0.60-0.94; *P*=.01) (Table 2). Conversely, older age (OR 1.01, 95% CI 1.00-1.02; *P*<.001), having a chronic condition (OR 1.83, 95% CI 1.44-2.33; *P*<.001), having a regular provider (OR 1.45, 95% CI 1.14-1.84; *P*=.003), and having an income lower than US $20,000 (OR 1.61, 95% CI 1.11-2.34; *P*=.01) were positively associated with indicating a preference for speaking directly to a provider. In terms of education, individuals with less than a high school education (OR 2.03, 95% CI 1.17-3.50; *P*=.011), a high school diploma (OR 2.16, 95% CI 1.57-2.97; *P*<.001), and a college degree (OR 1.40, 95% CI 1.01-1.94; *P*=.04) had significantly higher odds of preferring to speak directly to a provider compared with individuals having a postgraduate degree. Individuals who rated their quality of care as “very good” (OR 0.58, 95% CI 0.36-0.95; *P*=.03) or “good” (OR 0.64, 95% CI 0.47-0.87; *P*=.004) had significantly lower odds of preferring to speak directly to a provider compared with individuals who rated their quality of care as “poor.”

For the second barrier, Hispanic individuals had significantly higher odds of indicating that they had no need for a patient portal (OR 1.59, 95% CI 1.24-2.05; *P*<.001) compared with non-Hispanic individuals (Table 2). In contrast, older individuals (OR 0.99, 95% CI 0.99-1.00; *P*=.04), individuals who rated their quality of care as “excellent” (OR 0.33, 95% CI 0.12-0.88; *P*=.03), and individuals with a household income of US $20,000 to $34,999 (OR 0.61, 95% CI 0.44-0.85; *P*=.003) or less than US $20,000 (OR 0.68, 95% CI 0.50-0.93; *P*=.02) had significantly lower odds of indicating that they had no need for a patient portal.

For the third barrier, older individuals (OR 1.05, 95% CI 1.04-1.06; *P*<.001), individuals with a chronic condition (OR 1.42, 95% CI 1.08-1.86; *P*=.01), and individuals who rated their quality of care as “good” (OR 1.55, 95% CI 1.11-2.15; *P*=.009) had significantly higher odds of indicating that they were uncomfortable with a computer (Table 2). Individuals with less than a high school diploma (OR 3.15, 95% CI 1.79-5.53; *P*<.001) and a high school diploma (OR 2.79, 95% CI 1.80-4.33; *P*<.001) had significantly higher odds of indicating that they were uncomfortable with a computer. Individuals with a household income of less than US $20,000 (OR 2.78, 95% CI 1.88-4.11; *P*<.001), US $20,000 to $34,999 (OR 2.17, 95% CI 1.45-3.27; *P*<.001), US $35,000 to $49,999 (OR 1.94, 95% CI 1.30-2.90; *P*=.001), and US $50,000 to $74,999 (OR 1.57, 95% CI 1.30-2.90; *P*=.001) had significantly higher odds of indicating that they were uncomfortable with a computer compared with individuals having a household income of US $75,000 or more. Black individuals (OR 0.93, 95% CI 0.89-0.98; *P*=.007) were less likely to indicate that they were uncomfortable with a computer compared with white individuals.

Table 2. Nonadopter characteristics and the three most common barriers to patient portal adoption.

Characteristic	Model 1: Speaking directly to a provider (N=1926; weighted: 88,630,105)			Model 2: No need for a patient portal (N=1893; weighted: 87,531,372)			Model 3: Uncomfortable with a computer (N=1893; weighted: 87,531,372)		
	OR[a]	95% CI	*P* value	OR	95% CI	*P* value	OR	95% CI	*P* value
Gender									
Male	0.75	0.60-0.94	.01	1.03	0.85-1.25	.75	1.18	0.93-1.49	.18
Female (ref[b])	N/A[c]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race									
Black	1.01	0.96-1.06	.73	1.05	1.00-1.10	.05	0.93	0.89-0.98	.007
White (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ethnicity									
Hispanic	0.90	0.67-1.21	.48	1.59	1.24-2.05	<.001	0.81	0.60-1.10	.19
Non-Hispanic (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age	1.01	1.00-1.02	<.001	0.99	0.99-1.00	.04	1.05	1.04-1.06	<.001
Education									
Less than HS[d]	2.03	1.17-3.50	.01	0.68	0.43-1.07	.10	3.15	1.79-5.53	<.001
HS diploma	2.16	1.57-2.97	<.001	0.98	0.73-1.32	.91	2.79	1.80-4.33	<.001
College degree	1.40	1.01-1.94	.04	1.1	0.81-1.50	.55	1.39	0.86-2.26	.18
Postgraduate (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Income									
Less than US $20,000	1.61	1.11-2.34	.01	0.68	0.50-0.93	.02	2.78	1.88-4.11	<.001
US $20,000 to $34,999	1.32	0.90-1.94	.16	0.61	0.44-0.85	.003	2.17	1.45-3.27	<.001
US $35,000 to $49,999	0.97	0.68-1.38	.86	0.79	0.58-1.08	.14	1.94	1.30-2.90	< .001
US $50,000 to $74,999	1.19	0.86-1.64	.29	0.99	0.74-1.32	.92	1.57	1.06-2.31	.02
US $75,000 or more (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rural									
Yes	1.05	0.74-1.49	.78	1.06	0.79-1.42	.70	1.02	0.72-1.45	.91
No (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Marital status									
Married	1.04	0.82-1.32	.74	0.86	0.69-1.06	.16	1.20	0.93-1.56	.16
Unmarried (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chronic condition									
Yes	1.83	1.44-2.33	<.001	0.93	0.75-1.16	.53	1.42	1.08-1.86	.01
No (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Insurance status									
Uninsured	0.88	0.53-1.47	.63	0.85	0.53-1.35	.49	0.60	0.31-1.15	.12
Insured (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Regular provider									
Yes	1.45	1.14-1.84	.003	1.19	0.96-1.48	.12	.98	0.75-1.30	.92
No (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Quality of care									
Excellent	0.52	0.19-1.42	.20	0.33	0.12-0.88	.03	1.76	0.62-4.99	.29

Characteristic	Model 1: Speaking directly to a provider (N=1926; weighted: 88,630,105)			Model 2: No need for a patient portal (N=1893; weighted: 87,531,372)			Model 3: Uncomfortable with a computer (N=1893; weighted: 87,531,372)		
	OR[a]	95% CI	*P* value	OR	95% CI	*P* value	OR	95% CI	*P* value
Very good	0.58	0.36-0.95	.03	0.69	0.44-1.08	.10	1.41	0.83-2.39	.21
Good	0.64	0.47-0.87	.004	0.92	0.71-1.20	.54	1.55	1.11-2.15	.009
Fair	0.93	0.71-1.22	.60	0.99	0.79-1.24	.94	1.09	0.82-1.46	.54
Poor (ref)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Constant	0.66	0.32-1.36	.26	0.94	0.50-1.77	.86	0.01	0.00-0.01	<.001

[a]OR: odds ratio.

[b]ref: reference.

[c]N/A: not applicable.

[d]HS: high school.

Discussion

Principal Findings

The goal of this study was to assess the prevalence of barriers to patient portal adoption in a nationally representative sample and to understand which patient subgroups are most likely to experience a given barrier. Our study found that the most common barriers to patient portal adoption are an individual's preference to speak to a provider in person, not having a need for the patient portal, and feeling uncomfortable with computers, which are barriers that are modifiable and can be intervened upon. Less frequently, patients reported concerns with privacy, internet access, difficulty logging on, and having multiple patient portals. Our study identified that patient characteristics can help predict which patients are most likely to experience certain barriers to patient portal adoption. Older adults and women, for example, commonly reported preference for in-person communication as a barrier to patient portals. We discuss implications for policy and practice below.

Consistent with previous studies, our study indicated that patient preference for in-person communication with providers serves as a barrier to patient portal adoption [38,46-49]. This study extends prior studies by demonstrating that this barrier may be more common (1810/2828, 64.00%) than previously thought and that certain patient demographics are associated with preferring in-person communication (eg, women, older patients, and patients with lower income and education). Dissemination strategies, which target information to a specific audience, may be needed to demonstrate that patient portals are meant to complement rather than replace in-person communication with providers [50]. For developing better messaging, however, there is a need for more implementation studies identifying effective practices for using the patient portal as a means to bolster patient-physician communication during visits. There have been successful examples of this, including collecting and displaying patient-reported outcomes through the portal, and using the portal to facilitate advanced care planning discussions [12,38,51-53]. These strategies should be replicated in additional settings.

Almost half of the individuals in our study indicated that not having a need for a patient portal was a barrier to adoption, which has emerged as a barrier in past research [33,34,36,37]. Our study found that Hispanic and younger individuals were more likely to not see a need for the patient portal. A prior study recommended using real-life patient stories that demonstrate how a patient portal can be used to make the case for why a patient portal might be valuable [54]. There also remains a need to do more usability testing with patient portals and apply user-centered design approaches to better understand what features within the patient portal would be valuable to patients [55-57]. Studies have identified a number of usability issues, including not having information presented in multiple languages, lack of educational resources, poor data visualization and lack of contextualization for laboratory values, and lack of personalization [58-60]. In response, some systems have tested creative strategies, such as offering tailored patient education, and have used motivational strategies (eg, social comparisons and gamification) to enhance the relevance of the patient portal and ensure that the portal is meeting user needs [55,61]. Future studies are needed to test strategies that align the patient portal with patients' information needs. Some systems have also trained providers on the portal and created time within clinic workflow to show patients what the portal is and how to use it, and these strategies may enhance patients' perceived need for the portal [62,63]. Further testing is needed to see whether this is an effective implementation approach.

Our study also found that lack of comfort with computers (735/2828, 25.99%) was a common barrier to patient portal adoption, a finding similar to that in past studies [33-35,37,64-66]. Consistent with prior work, we found that older individuals and individuals with lower income and education attainment were more likely to report lack of comfort with computers as a barrier to patient portal adoption. Several studies have tested strategies, such as having health care systems offer patient portal demonstrations to patients, as means to increase comfort with computers and ultimately patient portal adoption [2,62,67]. Researchers have also recommended providing additional support to older individuals who may lack comfort with computers, such as printed handouts and an option to call a toll free line for additional technical assistance [67,68]. Past studies have found that in-person training can improve eHealth literacy among older adults [69-71]; however, few studies have been performed in clinical settings or have included

clinical outcomes. Thus, additional research is warranted. Past studies have suggested that some older individuals experience additional barriers to technology adoption, such as vision, cognitive, and dexterity deficits [72]. Future studies should test whether modifications to patient portals, such as larger font size, increased contrast between the text and background, and voice-enabled applications, could increase comfort with patient portals among older frail adults. Some health care systems have also allowed patients to designate a caregiver to access the patient portal on their behalf, although uptake has been slow [73-75]. Implementation strategies that incorporate patient caregivers, such as proxy portal access, training for caregivers on the patient portal, and allowing patients to choose which information is shared with the caregiver, could help make the portal more accessible to patients who lack comfort with technology.

Limitations

This study has a number of limitations. First, the HINTS does not collect any information at the site of care where a patient is seen. It is possible that system- and provider-level factors influence which patients are most likely to experience certain barriers to patient portal adoption. For example, patients who receive care in a Veterans Affairs hospital may be more likely to report a need for the patient portal since the Veterans Affairs has been an early adopter of patient portals [57]. Additionally, there are other barriers that may affect patient portal adoption, such as lack of reimbursement for telemedicine. These factors were not captured in the HINTS. Second, the HINTS only asked about a limited set of barriers, and it is impossible to tell whether there are other barriers that may be more impactful in hindering patient portal adoption (eg, lack of Spanish language options for the patient portal). Further, the HINTS did not use an implementation framework to select questions related to patient portal barriers. It is possible that other important barriers may have been omitted. The survey questions do, however, capture many of the barriers reported in prior portal studies [38,46-49], suggesting that the questions align well with prior research. Third, the HINTS response rate was around 30%, so it is possible that the findings are not representative of the entire sample owing to nonresponse bias. Finally, the HINTS does not include a measure of multimorbidity (eg, the Charlson comorbidity index), which is positively associated with patient portal adoption. To address this limitation, we included a variable that indicated whether a patient had a chronic condition.

Conclusions

To our knowledge, this is the first study to assess the prevalence of barriers to patient portal adoption in a nationally representative sample and to discern which patient subgroups are most likely to experience certain barriers. Further research is needed to develop and test implementation strategies that target common barriers to patient portal adoption and tailor implementation and dissemination approaches based on patients' needs and preferences.

Authors' Contributions

KT developed the research question and study design, conducted the statistical analyses, and drafted the manuscript. AC reviewed the statistical analyses and the manuscript draft and provided feedback. YH reviewed the statistical analyses and the manuscript draft and provided feedback. AA reviewed the statistical analyses and the manuscript draft and provided feedback. CS helped refine aspects of the study design and methodology, reviewed the manuscript draft, and provided feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Nonadopter characteristics and no patient portal or internet access.
[DOCX File , 21 KB - jmir_v22i9e18870_app1.docx]

Multimedia Appendix 2
Nonadopter characteristics and privacy concerns or difficulty logging on.
[DOCX File , 21 KB - jmir_v22i9e18870_app2.docx]

References

1. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375 [FREE Full text] [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
2. Turner K, Klaman SL, Shea CM. Personal health records for people living with HIV: a review. AIDS Care 2016 Sep;28(9):1181-1187. [doi: 10.1080/09540121.2016.1153594] [Medline: 26917113]
3. McInnes DK, Shimada SL, Midboe AM, Nazi KM, Zhao S, Wu J, et al. Patient Use of Electronic Prescription Refill and Secure Messaging and Its Association With Undetectable HIV Viral Load: A Retrospective Cohort Study. J Med Internet Res 2017 Feb 15;19(2):e34 [FREE Full text] [doi: 10.2196/jmir.6932] [Medline: 28202428]
4. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]

5. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, Reed ME. Association of Mobile Patient Portal Access With Diabetes Medication Adherence and Glycemic Levels Among Adults With Diabetes. JAMA Netw Open 2020 Feb 05;3(2):e1921429 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21429] [Medline: 32074289]
6. Zhong X, Park J, Liang M, Shi F, Budd PR, Sprague JL, et al. Characteristics of Patients Using Different Patient Portal Functions and the Impact on Primary Care Service Utilization and Appointment Adherence: Retrospective Observational Study. J Med Internet Res 2020 Feb 25;22(2):e14410 [FREE Full text] [doi: 10.2196/14410] [Medline: 32130124]
7. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient Use and Experience With Online Access to Electronic Health Records in Norway: Results From an Online Survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
8. Anthony DL, Campos-Castillo C, Lim PS. Who Isn't Using Patient Portals And Why? Evidence And Implications From A National Sample Of US Adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
9. Jung C, Padman R, Shevchik G, Paone S. Who are portal users vs. early e-Visit adopters? A preliminary analysis. AMIA Annu Symp Proc 2011;2011:1070-1079 [FREE Full text] [Medline: 22195168]
10. Turner K, Hong Y, Yadav S, Huo J, Mainous AG. Patient portal utilization: before and after stage 2 electronic health record meaningful use. J Am Med Inform Assoc 2019 Oct 01;26(10):960-967. [doi: 10.1093/jamia/ocz030] [Medline: 30947331]
11. Graham TA, Ali S, Avdagovska M, Ballermann M. Effects of a Web-Based Patient Portal on Patient Satisfaction and Missed Appointment Rates: Survey Study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
12. Bailey SC, Wallia A, Wright S, Wismer GA, Infanzon AC, Curtis LM, et al. Electronic Health Record-Based Strategy to Promote Medication Adherence Among Patients With Diabetes: Longitudinal Observational Study. J Med Internet Res 2019 Oct 21;21(10):e13499 [FREE Full text] [doi: 10.2196/13499] [Medline: 31638592]
13. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the Users of a Patient Portal: Cross-Sectional Study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
14. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
15. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: A literature review. Int J Med Inform 2011 Jan;80(1):1-12. [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
16. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
17. 2016 Program Requirements. Centers for Medicare and Medicaid Services. 2016. URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/2016ProgramRequirements.html [accessed 2018-10-30]
18. Lye CT, Forman HP, Daniel JG, Krumholz HM. The 21st Century Cures Act and electronic health records one year later: will patients see the benefits? J Am Med Inform Assoc 2018 Sep 01;25(9):1218-1220. [doi: 10.1093/jamia/ocy065] [Medline: 30184156]
19. Promoting Interoperability. HealthIT. 2019. URL: https://www.healthit.gov/topic/meaningful-use-and-macra/promoting-interoperability [accessed 2020-09-10]
20. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in Electronic Health Record Patient Portal Use in Nephrology Clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
21. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth Dynamics of Patient-Provider Internet Communication: Trend Analysis Using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
22. Neinstein A, Thao C, Savage M, Adler-Milstein J. Deploying Patient-Facing Application Programming Interfaces: Thematic Analysis of Health System Experiences. J Med Internet Res 2020 Apr 3;22(4):e16813 [FREE Full text] [doi: 10.2196/16813] [Medline: 31983680]
23. Adler-Milstein J, Longhurst C. Assessment of Patient Use of a New Approach to Access Health Record Data Among 12 US Health Systems. JAMA Netw Open 2019 Aug 02;2(8):e199544 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9544] [Medline: 31411708]
24. Beckjord EB, Finney RL, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007 Jul;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
25. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
26. Spooner KK, Salemi JL, Salihu HM, Zoorob RJ. Disparities in perceived patient-provider communication quality in the United States: Trends and correlates. Patient Educ Couns 2016 May;99(5):844-854. [doi: 10.1016/j.pec.2015.12.007] [Medline: 26725930]

27. Patel V, Barker W, Siminerio E. Disparities in Individuals' Access and Use of Health Information Technology in 2014. HealthIT. 2016 Feb. URL: https://dashboard.healthit.gov/evaluations/data-briefs/disparities-individuals-access-use-health-information-technology.php [accessed 2020-09-10]
28. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
29. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
30. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in Access to and Use of Electronic Personal Health Information Between Rural and Urban Residents in the United States. J Rural Health 2018 Feb;34 Suppl 1:s30-s38. [doi: 10.1111/jrh.12228] [Medline: 28075508]
31. Bhandari N, Shi Y, Jung K. Seeking health information online: does limited healthcare access matter? J Am Med Inform Assoc 2014;21(6):1113-1117 [FREE Full text] [doi: 10.1136/amiajnl-2013-002350] [Medline: 24948558]
32. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. J Am Med Inform Assoc 2013;20(6):1128-1131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
33. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
34. Wildenbos GA, Peute L, Jaspers M. Facilitators and Barriers of Electronic Health Record Patient Portal Adoption by Older Adults: A Literature Study. Stud Health Technol Inform 2017;235:308-312. [Medline: 28423804]
35. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2014 May 12;18(6):2296-2305. [doi: 10.1111/hex.12199] [Medline: 24816246]
36. Powell KR. Patient-Perceived Facilitators of and Barriers to Electronic Portal Use: A Systematic Review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
37. Ronda MC, Dijkhorst-Oei L, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
38. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
39. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient Portal Use and Experience Among Older Adults: Systematic Review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
40. Sadasivaiah S, Lyles CR, Kiyoi S, Wong P, Ratanawongsa N. Disparities in Patient-Reported Interest in Web-Based Patient Portals: Survey at an Urban Academic Safety-Net Hospital. J Med Internet Res 2019 Mar 26;21(3):e11421 [FREE Full text] [doi: 10.2196/11421] [Medline: 30912747]
41. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
42. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito DA. Patient Portals as a Tool for Health Care Engagement: A Mixed-Method Study of Older Adults With Varying Levels of Health Literacy and Prior Patient Portal Use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
43. Kim S, Fadem S. Communication matters: Exploring older adults' current use of patient portals. Int J Med Inform 2018 Dec;120:126-136. [doi: 10.1016/j.ijmedinf.2018.10.004] [Medline: 30409337]
44. Health Information National Trends Survey 5 (HINTS 5) Cycle 3 Methodology Report. National Cancer Institute. 2019 Aug. URL: https://hints.cancer.gov/docs/Instruments/HINTS5_Cycle3_MethodologyReport.pdf [accessed 2020-09-10]
45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. International Journal of Surgery 2014 Dec;12(12):1495-1499 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
46. Eden KB, Totten AM, Kassakian SZ, Gorman PN, McDonagh MS, Devine B, et al. Barriers and facilitators to exchanging health information: a systematic review. Int J Med Inform 2016 Apr;88:44-51 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.01.004] [Medline: 26878761]
47. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
48. Lockwood MB, Dunn-Lopez K, Pauls H, Burke L, Shah SD, Saunders MA. If you build it, they may not come: modifiable barriers to patient portal use among pre- and post-kidney transplant patients. JAMIA Open 2018 Oct;1(2):255-264 [FREE Full text] [doi: 10.1093/jamiaopen/ooy024] [Medline: 31984337]

49. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": Understanding Barriers to Portal Use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
50. Lomas J. Diffusion, Dissemination, and Implementation: Who Should Do What? Ann NY Acad Sci 1993 Dec 31;703:226-235. [doi: 10.1111/j.1749-6632.1993.tb26351.x] [Medline: 8192299]
51. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
52. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting Lung Cancer Patients With an Interactive Patient Portal: Feasibility Study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
53. Li J, Yazdany J, Trupin L, Izadi Z, Gianfrancesco M, Goglin S, et al. Capturing a Patient-Reported Measure of Physical Function Through an Online Electronic Health Record Patient Portal in an Ambulatory Clinic: Implementation Study. JMIR Med Inform 2018 May 09;6(2):e31 [FREE Full text] [doi: 10.2196/medinform.8687] [Medline: 29743158]
54. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth Literacy: Patient Engagement in Identifying Strategies to Encourage Use of Patient Portals Among Older Adults. Population Health Management 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
55. Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA. A Patient-Facing Diabetes Dashboard Embedded in a Patient Web Portal: Design Sprint and Usability Testing. JMIR Hum Factors 2018 Sep 24;5(3):e26 [FREE Full text] [doi: 10.2196/humanfactors.9569] [Medline: 30249579]
56. Nazi KM, Turvey CL, Klein DM, Hogan TP. A Decade of Veteran Voices: Examining Patient Portal Enhancements Through the Lens of User-Centered Design. J Med Internet Res 2018 Jul 10;20(7):e10413 [FREE Full text] [doi: 10.2196/10413] [Medline: 29991468]
57. Maher M, Kaziunas E, Ackerman M, Derry H, Forringer R, Miller K, et al. User-Centered Design Groups to Engage Patients and Caregivers with a Personalized Health Information Technology Tool. Biol Blood Marrow Transplant 2016 Feb;22(2):349-358 [FREE Full text] [doi: 10.1016/j.bbmt.2015.08.032] [Medline: 26343948]
58. Garrido T, Kanter M, Meng D, Turley M, Wang J, Sue V, et al. Race/ethnicity, personal health record access, and quality of care. Am J Manag Care 2015 Feb;21(2):e103-e113 [FREE Full text] [Medline: 25880485]
59. Wallace LS, Angier H, Huguet N, Gaudino JA, Krist A, Dearing M, et al. Patterns of Electronic Portal Use among Vulnerable Patients in a Nationwide Practice-based Research Network: From the OCHIN Practice-based Research Network (PBRN). J Am Board Fam Med 2016 Oct;29(5):592-603 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160046] [Medline: 27613792]
60. Arcia A. Time to Push: Use of Gestational Age in the Electronic Health Record to Support Delivery of Relevant Prenatal Education Content. EGEMS (Wash DC) 2017 Apr 20;5(2):5 [FREE Full text] [doi: 10.13063/2327-9214.1281] [Medline: 29881749]
61. McAlearney AS, Sieck CJ, Hefner JL, Aldrich AM, Walker DM, Rizer MK, et al. High Touch and High Tech (HT2) Proposal: Transforming Patient Engagement Throughout the Continuum of Care by Engaging Patients with Portal Technology at the Bedside. JMIR Res Protoc 2016 Nov 29;5(4):e221 [FREE Full text] [doi: 10.2196/resprot.6355] [Medline: 27899338]
62. Ramsey A, Lanzo E, Huston-Paterson H, Tomaszewski K, Trent M. Increasing Patient Portal Usage: Preliminary Outcomes From the MyChart Genius Project. J Adolesc Health 2018 Jan;62(1):29-35 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.08.029] [Medline: 29169768]
63. McCleary NJ, Greenberg TL, Barysauskas CM, Guerette EJ, Hassan M, Jacobson JO, et al. Oncology Patient Portal Enrollment at a Comprehensive Cancer Center: A Quality Improvement Initiative. J Oncol Pract 2018 Aug;14(8):e451-e461. [doi: 10.1200/JOP.17.00008] [Medline: 30096276]
64. Ennis L, Robotham D, Denis M, Pandit N, Newton D, Rose D, et al. Collaborative development of an electronic Personal Health Record for people with severe and enduring mental health problems. BMC Psychiatry 2014;14(1):305 [FREE Full text] [doi: 10.1186/s12888-014-0305-9] [Medline: 25403285]
65. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van DG, Wei E. "You Get Reminded You're a Sick Person": Personal Data Tracking and Patients With Multiple Chronic Conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
66. Fix GM, Hogan TP, Amante DJ, McInnes DK, Nazi KM, Simon SR. Encouraging Patient Portal Use in the Patient-Centered Medical Home: Three Stakeholder Perspectives. J Med Internet Res 2016 Nov 22;18(11):e308 [FREE Full text] [doi: 10.2196/jmir.6488] [Medline: 27876686]
67. Gordon NP, Hornbrook MC. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
68. Zickuhr K, Smith A. Home Broadband 2013. Pew Research Center. 2013 Aug. URL: https://www.pewresearch.org/internet/2013/08/26/home-broadband-2013/ [accessed 2020-09-10]
69. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]

70. Xie B, Bugg JM. Public library computer training for older adults to access high-quality Internet health information. Libr Inf Sci Res 2009 Sep 1;31(3):155 [FREE Full text] [doi: 10.1016/j.lisr.2009.03.004] [Medline: 20161649]
71. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
72. Smith A. Older Adults and Technology Use. Pew Research Center. 2014 Apr. URL: https://www.pewresearch.org/internet/2014/04/03/older-adults-and-technology-use/ [accessed 2020-09-10]
73. Sarkar U, Bates DW. Care Partners and Online Patient Portals. JAMA 2014 Jan 22;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
74. Wolff JL, Berger A, Clarke D, Green JA, Stametz R, Yule C, et al. Patients, care partners, and shared access to the patient portal: online practices at an integrated health system. J Am Med Inform Assoc 2016 Nov;23(6):1150-1158. [doi: 10.1093/jamia/ocw025] [Medline: 27026614]
75. Weis A, Pohlmann S, Poss-Doering R, Strauss B, Ullrich C, Hofmann H, et al. Caregivers' role in using a personal electronic health record: a qualitative study of cancer patients and caregivers in Germany. BMC Med Inform Decis Mak 2020 Jul 13;20(1):158 [FREE Full text] [doi: 10.1186/s12911-020-01172-4] [Medline: 32660600]

Abbreviations

EHR: electronic health record
HINTS: Health Information National Trends Survey
NCI: National Cancer Institute
OR: odds ratio

Edited by G Eysenbach; submitted 24.03.20; peer-reviewed by T Garrido, M Barwick; comments to author 18.07.20; revised version received 22.07.20; accepted 16.08.20; published 17.09.20.

Please cite as:
Turner K, Clary A, Hong YR, Alishahi Tabriz A, Shea CM
Patient Portal Barriers and Group Differences: Cross-Sectional National Survey Study
J Med Internet Res 2020;22(9):e18870
URL: http://www.jmir.org/2020/9/e18870/
doi:10.2196/18870
PMID:32940620

Original Paper

Electronic Health Record Portal Messages and Interactive Voice Response Calls to Improve Rates of Early Season Influenza Vaccination: Randomized Controlled Trial

Jessica G Wijesundara[1], MPH; Mayuko Ito Fukunaga[1,2,3], MD; Jessica Ogarek[4], MS; Bruce Barton[1], PhD; Lloyd Fisher[5,6], MD; Peggy Preusse[6], RN; Devi Sundaresan[6], MS; Lawrence Garber[2,6], MD; Kathleen M Mazor[2,3], EdD; Sarah L Cutrona[1,7], MD, MPH

[1]Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[2]Meyers Primary Care Institute, Worcester, MA, United States

[3]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[4]Center for Gerontology and Healthcare Research, Brown University, Providence, MA, United States

[5]Department of Pediatrics, University of Massachusetts Medical School, Worcester, MA, United States

[6]Reliant Medical Group, Worcester, MA, United States

[7]Health Services Research & Development, Center of Innovation, Edith Nourse Rogers Memorial Hospital, Veterans Health Administration, Bedford, MA, United States

Corresponding Author:
Sarah L Cutrona, MD, MPH
Health Services Research & Development, Center of Innovation
Edith Nourse Rogers Memorial Hospital
Veterans Health Administration
200 Springs St
Building 70
Bedford, MA, 01730
United States
Phone: 1 508 856 4046
Email: Sarah.Cutrona@umassmed.edu

Abstract

Background: Patient reminders for influenza vaccination, delivered via an electronic health record patient portal and interactive voice response calls, offer an innovative approach to engaging patients and improving patient care.

Objective: The goal of this study was to test the effectiveness of portal and interactive voice response outreach in improving rates of influenza vaccination by targeting patients in early September, shortly after vaccinations became available.

Methods: Using electronic health record portal messages and interactive voice response calls promoting influenza vaccination, outreach was conducted in September 2015. Participants included adult patients within a large multispecialty group practice in central Massachusetts. Our main outcome was electronic health record–documented early influenza vaccination during the 2015-2016 influenza season, measured in November 2015. We randomly assigned all active portal users to 1 of 2 groups: (1) receiving a portal message promoting influenza vaccinations, listing upcoming clinics, and offering online scheduling of vaccination appointments (n=19,506) or (2) receiving usual care (n=19,505). We randomly assigned all portal nonusers to 1 of 2 groups: (1) receiving interactive voice response call (n=15,000) or (2) receiving usual care (n=43,596). The intervention also solicited patient self-reports on influenza vaccinations completed outside the clinic. Self-reported influenza vaccination data were uploaded into the electronic health records to increase the accuracy of existing provider-directed electronic health record clinical decision support (vaccination alerts) but were excluded from main analyses.

Results: Among portal users, 28.4% (5549/19,506) of those randomized to receive messages and 27.1% (5294/19,505) of the usual care group had influenza vaccinations documented by November 2015 (*P*=.004). In multivariate analysis of portal users, message recipients were slightly more likely to have documented vaccinations when compared to the usual care group (OR 1.07, 95% CI 1.02-1.12). Among portal nonusers, 8.4% (1262/15,000) of those randomized to receive calls and 8.2% (3586/43,596) of usual care had documented vaccinations (*P*=.47), and multivariate analysis showed nonsignificant differences. Over half of

portal messages sent were opened (10,112/19,479; 51.9%), and over half of interactive voice response calls placed (7599/14,984; 50.7%) reached their intended target, thus we attained similar levels of exposure to the messaging for both interventions. Among portal message recipients, 25.4% of message openers (2570/10,112) responded to a subsequent question on receipt of influenza vaccination; among interactive voice response recipients, 72.5% of those reached (5513/7599) responded to a similar question.

Conclusions: Portal message outreach to a general primary care population achieved a small but statistically significant improvement in rates of influenza vaccination (OR 1.07, 95% CI 1.02-1.12). Interactive voice response calls did not significantly improve vaccination rates among portal nonusers (OR 1.03, 95% CI 0.96-1.10). Rates of patient engagement with both modalities were favorable.

Trial Registration: ClinicalTrials.gov NCT02266277; https://clinicaltrials.gov/ct2/show/NCT02266277

(J Med Internet Res 2020;22(9):e16373) doi:10.2196/16373

KEYWORDS

electronic health records; influenza vaccination; patient care; patient engagement

Introduction

Influenza infections contribute to increased health care costs and loss of productivity and can lead to serious complications and even death [1]. Effective strategies for the prevention of influenza are of critical importance as we enter the 2020-2021 influenza season. The confluence of the upcoming influenza season and the ongoing coronavirus disease 2019 (COVID-19) pandemic is expected to place additional stress on our health care system, fueled in part by similarities in presenting complaints for these two illnesses, as well as by the potential for increased risk of poor outcomes in patients co-infected with COVID-19 and influenza [2-4].

An estimated 5%-20% of the US population contracts influenza every year, with several hundred thousand people hospitalized annually due to influenza-related complications [5]. Estimates of annual influenza and pneumonia-associated deaths over the past decade reached as high as 61,000 in the 2017-2018 season [6-12]. According to Centers for Disease Control and Prevention (CDC) estimates, during the 2018-2019 influenza season, vaccinations prevented approximately 4.4 million flu illnesses, 58,000 hospitalizations, and 3500 deaths [13,14]. Since it takes approximately 2 weeks for antibodies to develop in response to influenza vaccination, the CDC recommends getting vaccinated before flu begins to spread within a community, by the end of October [15].

Despite widespread publicity promoting influenza vaccination, vaccines are underutilized [16-19]. In 2017, national vaccination coverage among adults for influenza was 37.1% [20], while the Healthy People 2020 target was 70% [21]. The CDC estimated flu vaccination coverage among adults aged ≥18 years as of mid-November 2018 was 44.9% [22]; at the end of the 2018-2019 influenza season, national influenza vaccination coverage was 45.3%[23].

Clinical decision support has been shown to improve health outcomes by supporting the delivery of timely, evidence-based and guideline-concordant medical care, including annual vaccinations [24-26]. Clinical decision support includes computerized reminders both to providers and patients. Many health systems effectively use provider-directed clinical decision support, including noninterruptive or interruptive (pop-up) alerts, reminding providers of recommended prevention or screening measures [27,28]. While frequently effective, provider-directed clinical decision support is subject to important limitations. If alerts are triggered by erroneous or incomplete electronic health record data, or if providers experience alert fatigue from an overwhelming number of notifications, alerts may be ignored or overridden [29-32]. Considering these challenges, and in the setting of nationwide adoption of electronic patient portals, patient-directed clinical decision support delivered via a patient portal offers an innovative approach to the promotion of timely influenza vaccination.

Electronic patient portals are secure websites that provide patients with 24-hour online access to limited electronic health record information. A portal provides patients with a personal health record that is tethered to their electronic health record. Accessible information within a tethered portal varies by health system but may include vaccinations, laboratory results, problem lists, allergies, and information from recent doctor visits or hospitalizations [33,34]. A core function of portals is secure messaging—electronic communication with the physician or health care team [35]. Patient portals have the potential to improve patient-provider communication, improve medication adherence, decrease office visits, increase self-management of disease and disease awareness, increase use of preventative medicine, and increase inclusion of patients in medical decision making [36-38]. Previous patient outreach interventions have been shown to improve rates of vaccination completion and have been tested using multiple options including mailed letters, postcards, live phone calls, automated phone messages, and combination postcard/phone-based [39,40]. Few studies have tested the use of patient-directed vaccination reminders sent via patient portals [25,26,41].

We conducted a randomized controlled trial aimed at improving rates of influenza vaccination among eligible adult patients in a large multispecialty group practice in central Massachusetts. We used electronic health record patient portal messages as well as interactive voice recognition calls to (1) promote early season influenza vaccination completion and (2) solicit patient self-report on vaccinations completed outside the clinic.

Methods

Study Objectives

The overarching goal of this study was to improve rates of early influenza vaccination (by the end of October) among eligible adults in an outpatient population. Our primary objective was to determine whether our outreach increased completion of influenza vaccinations and, if so, whether one mode of outreach was most effective. Additional objectives were to improve documentation of influenza vaccinations administered outside the practice by inviting patient self-report (improving the accuracy of existing decision support tools) and to track process measures (eg, rates of portal message opening and interactive voice response call answering).

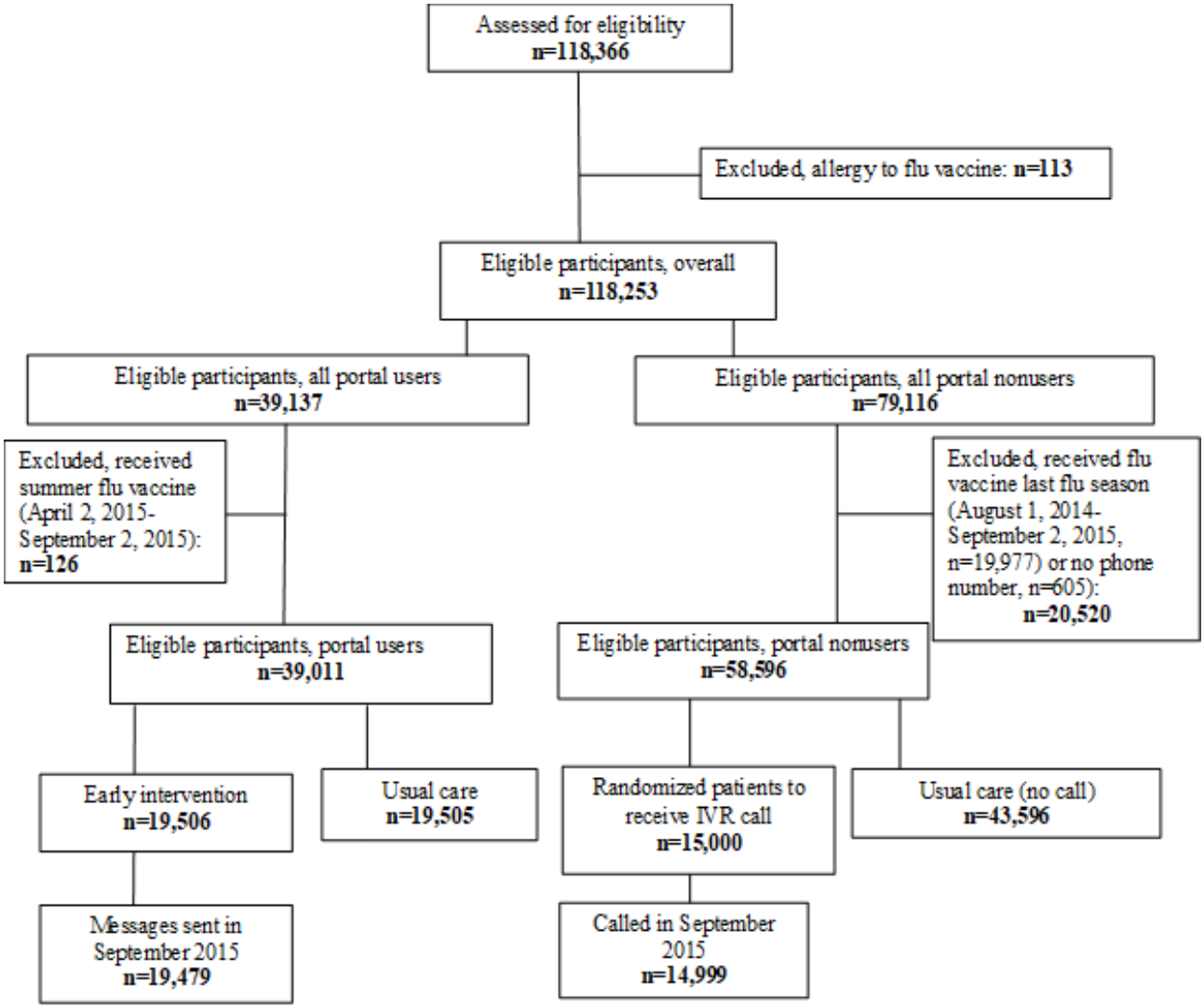
Study Design

We conducted a nonblinded randomized controlled intervention (NCT02266277) at a large multispecialty medical group in central Massachusetts. Previously, our team developed and tested interactive voice response and portal outreach which we targeted to patients who had no documented vaccination 2 months after the start of the season [25,26]; this study adapted our previous approach, targeting a broader population in early September in order to promote early vaccination and provide information on September and October flu clinic dates.

Using a computer-generated randomization table, we assigned all active portal users to receive either a portal message promoting influenza vaccination, listing upcoming clinics, and offering online scheduling of vaccination appointments (n=19,506) or usual care with no portal message (n=19,505). Separately, we randomized all portal nonusers to receive either an interactive voice response call (n=15,000) or usual care with no interactive voice response call (n=43,596) (Figure 1). For portal users only, after the conclusion of the study and assessment of outcomes on November 3, messages were sent to the usual care group if they still did not have an electronic health record–documented influenza vaccination for the 2015-2016 season (n=14,118). The cost of calls prevented us from being able to send interactive voice response messages to the usual care group in the portal nonusers.

The study was reviewed and approved in 2014 by the Reliant Medical Group institutional review board; due to administrative changes, oversight was transferred to the University of Massachusetts institutional review board in 2015. A waiver for informed consent for patient outreach was approved by these institutional review boards. Patients were not compensated for their participation. The principal investigator (SC) oversaw the trial and data analysis.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) Diagram describing randomized controlled trial to improve rates of early season influenza vaccinations using EHR portal messages and Interactive Voice Response calls.



Study Population

Eligibility Criteria

Patients were eligible for the study if they (1) had a primary care provider at the medical group during the 12 months prior to randomization; (2) were aged ≥18 on the date of randomization; (3) had a recent office visit or telephone encounter with an internal medicine practitioner or family practitioner (defined as having had an office visit, phone encounter, consult, or complete physical exam within the 12 months prior to randomization). This requirement was intended to minimize inclusion of patients who had moved to another practice but whose names were retained in the medical group records. To ensure capture of patients transitioning from pediatric to adult care, the recent office visit could also be with a pediatrician.

A patient was eligible for inclusion in the electronic patient portal portion of the trial if the patient was an active user, which was defined as having an activated portal with a log-in at least once in the year preceding randomization.

Exclusion Criteria

For both interactive voice response and portal outreach, patients were excluded if there was electronic health record documentation of an allergy to influenza vaccines. For interactive voice response calls only, exclusion criteria also included the presence of any of the following on the date of randomization: (1) electronic health record documentation of influenza vaccination completion in the 2014-2015 influenza season (or documented influenza vaccination after the end of the 2014-2015 influenza season but before the start of the 2015-2016 season); (2) no listed phone number.

Study Procedures

This study consisted of interactive voice response or portal-based outreach. Qualitative interviews conducted with patients, physicians, nurses, and staff informed the development of our outreach material [25,26].

Electronic Patient Portal Intervention

We designed an outgoing secure portal message to be sent via patient portal to patients randomized to the portal message arm (Multimedia Appendix 1). Portal message content appeared in letter format with the signature line reflecting the name of the

patient's primary care provider. Portal messages were delivered through standard channels used for portal-based correspondence between the medical group's health care providers and patients (ie, a generic message that contained neither personal health information nor any reference to vaccinations was delivered to patient's email account; the message prompted patients to log in to secure portal account via hyperlink). Once logged in to the portal accounts, patients clicked on a message labeled "A Message from Your Primary Care Provider" to view the outreach message. The outreach message included information about upcoming flu clinics for September and October 2015.

Unique to the portal message (compared to interactive voice response phone messages) was the option of direct online scheduling of appointments for influenza vaccination. Information about accessing the CDC vaccination website appeared within the body of the portal message as a hyperlink (and was conveyed verbally in the interactive voice response script). Opportunities to report external influenza vaccinations and to report intent to get vaccinated matched the interactive voice response call content.

Portal Message Delivery

Messages were sent out to 500 to 1500 patients daily over a period of 15 days in September 2015 in order to reduce the risk of being blocked by network bulk-spam filters.

Interactive Voice Response Call Intervention

Interactive voice response calls appeared on caller ID as originating from the medical group. This is consistent with identification of interactive voice response calls used for appointment reminders at the time of the study. Combining voice response with branching logic, calls elicited patient self-reports of influenza vaccinations completed outside the medical group (Multimedia Appendix 2). For patients reporting no influenza vaccination completed, calls included information about upcoming flu clinics for September and October 2015. Patients reporting no influenza vaccination completed were also asked whether they intended to get vaccinated. Patients reporting that they were unsure or did not intend to get vaccinated were asked further questions on specific reasons why they did not plan to get a flu vaccination.

Interactive Voice Response Call Delivery

Intervention interactive voice response calls, initiated on September 11, 2015, began by confirming that the person answering the phone was the intended patient recipient of the call. If voicemail was encountered or if the person reached was someone other than the patient, the interactive voice response system left a message asking patients to call back and provided an inbound call line number. The last interactive voice response outbound calls were placed on September 25, 2015. The inbound call line was maintained throughout the duration of outgoing calls and for 3 weeks after the final outgoing call was placed; patients who called this number from the phone number of record heard the interactive voice response call script in its entirety, beginning with questions confirming the identity of the caller.

Study Outcomes

Primary Outcome

Our primary outcome was percentage of eligible patients with influenza vaccinations documented in the electronic health record as of November 3, 2015. We pulled data on vaccination rates as of this date, chosen in order to assess the impact of early outreach on completion of early immunization. Immunizations captured solely through the patient portal questionnaire or through the interactive voice response were excluded from the primary analysis in order to enable comparison with the control groups. The origins of all entered influenza vaccinations were tracked to allow our team to distinguish between sources of information on completed vaccinations.

Process Measures and Additional Outcomes of Interest

For portal messages, we calculated (1) percentage of recipients who logged in to the patient portal during early flu season (through November 3, 2015), (2) percentage of recipients who opened messages during early flu season, and (3) percentage of recipients who completed questionnaires. We tracked self-reports (via the portal) of influenza vaccinations completed. We also tracked patient-reported intent to get an influenza vaccination during the 2015-2016 flu season.

For interactive voice response calls, we calculated (1) percentage of recipients reached and (2) percentage of recipients who completed the calls by responding to questions. We tracked self-report (via interactive voice response) of influenza vaccinations completed. We also tracked patient-reported intent to get an influenza vaccination during the 2015-2016 flu season.

Sample Size

With our proposed sample size, power calculations based on estimates of baseline vaccination rates indicated that 4286 participants per arm would give 80% power to detect a 3% improvement in influenza vaccination rates between groups (α=.05; 2-sided).

Statistical Methods

Primary Outcome

To determine the impact of our interventions on early vaccination rates for the 2015-2016 influenza season, we calculated frequencies and performed intention-to-treat bivariate analyses of randomized patients, assessing whether vaccination completion was associated with group assignment. Due to different rates of electronic health record–recorded vaccination measured at baseline (in 2014) between portal users (35.9%) and nonusers (25%), and due to the differences in intervention (portal versus interactive voice response call), analyses for these groups were conducted separately.

We then performed multivariate logistic regression analyses, adjusting for demographic and practice-level covariates, we modeled the odds of receiving an influenza vaccination in the 2015-2016 influenza season.

Results

Baseline Characteristics

Baseline characteristics of both patient portal users and portal nonusers are reported in Table 1. Baseline characteristics were similar among portal users and portal nonusers. However, compared to the portal nonusers, portal users were more likely to be women, older, and have a higher level of health care utilization.

Table 1. Baseline characteristics of participants.

Characteristics	All (N=97,607), n (%)	Portal users		Portal nonusers	
		Portal message (n=19,506), n (%)	Usual care (n=19,505), n (%)	Call (n=15,000), n (%)	Usual care (n=43,596), n (%)
Sex					
Female	53,250 (54.5)	12,230 (62.7)	12,249 (62.8)	7325 (48.8)	21,446 (49.2)
Male	44,349 (45.4)	7275 (37.3)	7256 (37.2)	7672 (51.1)	22,146 (50.8)
Missing	8 (0.01)	1 (0.0)	0 (0.0)	3 (0.0)	4 (0.0)
Age					
18-34	30,556 (31.3)	4508 (23.1)	4597 (23.6)	5615 (37.4)	15,836 (36.3)
35 - 49	25,579 (26.2)	5096 (26.1)	5062 (26.0)	3896 (26.0)	11,525 (26.4)
50 - 64	26,495 (27.1)	6275 (32.2)	6224 (31.9)	3547 (23.6)	10,449 (24.0)
65-74	8593 (8.8)	2419 (12.4)	2388 (12.2)	965 (6.4)	2821 (6.5)
75+	6384 (6.5)	1208 (6.2)	1234 (6.3)	977 (6.5)	2965 (6.8)
Race					
White	66,020 (67.6)	14,180 (72.7)	14,223 (72.9)	9608 (64.1)	28,009 (64.2)
Black	3417 (3.5)	456 (2.3)	440 (2.3)	630 (4.2)	1891 (4.3)
Asian	3502 (3.6)	737 (3.8)	715 (3.7)	528 (3.5)	1522 (3.5)
American Indian or Alaska Native	1021 (1.1)	140 (0.7)	144 (0.7)	194 (1.3)	543 (1.2)
Other	10,366 (10.6)	1672 (8.6)	1745 (8.9)	1769 (11.8)	5180 (11.9)
Missing	13,281 (13.6)	2321 (11.9)	2238 (11.5)	2271 (15.1)	6451 (14.8)
Health care utilization level					
Had office visit[a]	80,748 (82.7)	17,896 (91.7)	17,908 (91.8)	11,609 (77.4)	33,335 (76.5)
Did not have office visit	16,859 (17.3)	1610 (8.3)	1597 (8.2)	3391 (22.6)	10,261 (23.5)

[a]12 months prior to randomization.

Portal Users

Among portal users, 28.4% (5549/19,506) of message recipients and 27.1% (5294/19,505) of the usual care group had documentation in their electronic health records that they received that influenza vaccinations on or before November 3, 2015 (P=.004). Portal users who received the messages were significantly more likely to have received the influenza vaccination compared to the usual care group (odds ratio [OR] 1.07, 95% CI 1.02-1.12). This finding was consistent even after adjusting for age, race, sex, and health care utilization (OR 1.07, 95% CI 1.02-1.12) (Table 2).

Table 2. Likelihood of receiving an early-season influenza vaccination.

Recipients	n	Unadjusted		Adjusted[a]	
		OR[b]	95% CI	OR	95% CI
Portal message[c]	39,011	1.07	(1.02, 1.12)	1.07	(1.02, 1.12)
Interactive voice response call[c]	58,596	1.03	(0.96, 1.10)	1.03	(0.96, 1.11)

[a]Adjusted for age, sex, race, and health care utilization level (where utilization was defined as office visit, phone encounter, consult, or complete physical exam within the 12 months prior to randomization).

[b]OR: odds ratio.

[c]The reference is usual care.

Portal Nonusers

Among portal nonusers, 8.4% (1262/15,000) of call recipients and 8.2% (3586/43,596) of usual care recipients received vaccinations (*P*=.47). Bivariate and multivariate analysis showed nonsignificant differences in influenza vaccination rates between intervention and usual care groups (Table 2).

Process Measures

Portal Message

Among patient portal message recipients, 71.2% (13,862 recipients out of 19,479 to whom the message was sent) logged in to the patient portal on or after the date of message delivery through the end of early flu season (September 9, 2015 to November 3, 2015). Messages were opened by 51.9% (10,112/19,479) recipients; 13.2% (2570/19,479) responded to the first question asking if they received a flu vaccination on or after August 1, 2015, 2.0% (386/19,479) reported already receiving a vaccination and 11.2% (2176/19,479) responded to a second question assessing their intention to receive a flu vaccination during the flu season (asked only of those who were not already vaccinated) (Table 3).

Of those opening messages, 25.4% (2570/10,112) responded to our question on receipt of influenza vaccination, 3.8% (386/10,112) reported already receiving vaccinations and 21.5% (2176/10,112) responded to our question on whether they planned to get vaccinated.

Table 3. Process measures and self-reported influenza vaccinations for portal message recipients.

Action	Portal users, n (%)
Randomized	19,506
Message sent[a]	19,479 (100)
Logged in to patient portal	13,862 (71.2)
Opened message	10,112 (51.9)
Responded to “Have you received a flu vaccination on or after August 1, 2015?”	2570 (13.2)
Reported receiving a flu vaccination	386 (2.0)
Responded to “Do you plan to get a flu vaccination this flu season?”	2176 (11.2)
Reported that they planned to get a flu vaccination	1814 (9.3)

[a]27 patients were no longer portal users by the time we sent the messages (due to change in medical record numbers, invalid patient portal IDs, etc).

Interactive Voice Response Call

Among interactive voice response call recipients, 50.7% of patients were reached (7599/14,984); 36.8% (5513/14,984) responded to the first question asking if they received a flu vaccination on or after August 1, 2015, 3.8% (575/14,984) reported receiving their vaccination and 24.1% (3613/14,984) responded to a second question assessing their intention to receive a flu vaccination during the flu season (asked only of those who were not already vaccinated) (Table 4).

Table 4. Process measures and self-reported influenza vaccinations for interactive voice response call recipients.

Action	Call recipients, n (%)
Randomized	14,999
Call attempted[a]	14,984 (100)
Target reached (inbound and outbound calls)	7599 (50.7)
Responded to “Have you received a flu vaccination on or after August 1, 2015?”	5513 (36.8)
Reported receiving a flu vaccination	575 (3.8)
Responded to “Do you plan to get a flu vaccination this flu season?”	3613 (24.1)
Reported that they planned to get a flu vaccination	1415 (9.4)

[a]15 patients had invalid/blank phone numbers, or the patient was on a “Do Not Call” list, by the time we placed the calls.

Of those reached, 72.5% (5513/7599) responded to our question on receipt of influenza vaccination, 7.6% (575/7599) reported already receiving a vaccination, and 47.5% (3613/7599) responded to our question on whether they planned to get vaccinated.

Discussion

Principal Results

Our study showed a clinically small but statistically significant improvement in completion of early season influenza vaccinations among those randomized to receive outreach via patient portal, compared to a usual care control group (OR 1.07, 95% CI 1.02-1.12). There was no significant increase of early season influenza vaccinations among those randomized to

receive an interactive voice response call (OR 1.03, 95% CI 0.96-1.10).

This outreach was designed to deliver a relatively time-sensitive message (eg, reminding patients of the importance of influenza vaccination and alerting them of upcoming vaccination clinics in September and October), and as such, represents a successful, brief patient engagement effort. We attained greater than 50% opening rates for portal messages and reached the targeted patient on over 50% of interactive voice response calls, achieving similar levels of exposure to the messaging for the two interventions. Among portal message recipients, more than one quarter of those who opened the message responded to our subsequent question on whether they had received or intended to receive their influenza vaccination. Among interactive voice response call recipients, close to three-quarters of those reached responded to the question.

A small number of patients reported influenza vaccinations that had been completed in the community; data on patient-reported vaccinations performed in the community were then uploaded into the electronic health record in order to improve the accuracy of existing influenza vaccination alerts directed at primary care providers. The relatively small number of self-reported flu vaccinations (961/20,978 participants across the portal message intervention and the interactive voice response intervention arms combined) has several possible explanations. It is possible that participants who had already been vaccinated found the outreach less relevant; and therefore, chose not to engage (eg, chose not to open the message or chose to hang up the phone). Relatively low rates of self-reported vaccination completion may also be attributable to the timing of our outreach—intentionally positioned at the start of the flu season. In contrast, our past work [26] captured much higher proportions of self-reported vaccinations (2591/20,000 intervention patients) when participants were approached several months past the start of flu season, which highlights the importance of asking patients to self-report vaccinations administered outside the clinic.

Although these “discovered” immunizations were incorporated into the electronic health record, they were not counted in the primary analysis; this allowed comparison with the usual care groups who were not queried for outside immunizations. While recognizing that the actual immunization rate was higher than the documented rate, any differences in the documented rate between intervention and control groups can reasonably be attributed to increased vaccination rates. Data on intention to be vaccinated were also elicited. Studies have shown that stating intent to complete a behavior can enhance likelihood of follow-through [42,43], thus it is possible that engaging patients in this manner may have contributed to the modest success of our intervention.

Once designed, portal messages (sent out in several batches) required minimal staff time to deliver. Our messages were designed to be easily adapted for use in future years; our messages have already been adapted and implemented by our medical center to target high-risk populations (eg, children with asthma).

Comparison With Prior Work

There have been multiple studies [44-47] documenting success in use of patient-directed reminders for influenza vaccination. Effective reminders described include letters, telephone invitations, phone calls from a peer, tailored communications, customized letters/phone calls, and client-based appraisals [44]. Text message reminders have been shown to be successful for influenza vaccination reminders in pediatric and adolescent populations [45], in patients with rare diseases [46], and among high-risk patients [47].

While portal-based outreach directed at patients has been described for a variety of preventive measures [48], few studies have tested use of patient-directed influenza vaccination reminders sent via patient portals; one study [49] that has done so focused exclusively on untethered (ie, personally controlled) patient health records among a university student and staff population in Australia, yielding a 6.7% greater likelihood of influenza vaccination among users of the personal health records compared to those randomized to a 6-month waitlist for portal activation [49]. In our past work [25,26], we tested similar portal and interactive voice response outreach messages using a factorial design and targeting a population more likely to be nonadherent to vaccination guidelines (patients who had no documented vaccination 2 months after the start of the flu season). In that study [25,26], we found small but statistically significant improvements in influenza vaccination rates among recipients of either outreach method compared to usual care (portal message alone (OR 1.20, 95% CI 1.06-1.35); interactive voice response call alone (OR 1.15, 95 % CI 1.02-1.30); both messages (OR 1.29, 97.5% CI 1.13-1.48). In their recent randomized trial, Szilagyi et al [41] studied the effect of 1, 2, or 3 patient portal reminders on influenza vaccination rates among 164,205 patients in 52 primary care practices. They noted small but statistically significant improvements in vaccination rates with an attenuating effect of repeated messaging (37.5% for those receiving no reminders; 38.0% for 1 reminder; 38.2% for 2 reminders, and 38.2% for 3 reminders). Higher overall rates of vaccination were documented in their trial in comparison to ours; this is likely due in part to the study’s inclusion of children (who have a higher rate of influenza vaccination than that of the overall adult population) as well as their measurement of vaccination rates at the end of the influenza season (March 31).

The rates of patient engagement (measured by message opening and subsequent action) demonstrated by our patient portal recipients were comparable to those from previously published patient portal outreach studies. In our previous influenza outreach intervention [25,26] (in a population unvaccinated 2 months into the flu season), messages were opened by slightly more than half of message recipients and interactive voice response targets were reached in just over 60%. In that prior study, among those who opened portal messages, 28.6% responded to subsequent questions. Among those reached via interactive voice response call, 78.3% responded to questions. In their patient portal intervention aimed at improving influenza vaccination rates, Szilagyi and colleagues [41] found opening rates of 52.9% among patients receiving a single portal message

(increasing to 55.9% in a 2-reminder group and 58.8% in a 3-reminder group).

Fischer and colleagues [50] found that among patients randomized to active reminders about multiple health maintenance services (eg, hemoglobin A_{1c} testing, lipids, etc), nearly 65% of patients logged in to the portal after receiving the first of several messages. In a study [51] of colorectal cancer screening reminders delivered via patient portal, among the 552 patients randomized to receive messages, 54% viewed the message and 9% performed a suggested web-based risk assessment tool. In a study describing the reach and feasibility of an interactive lung cancer screening decision aid delivered by patient portal, Dharod and colleagues [52] found that 86% of lung cancer screening eligible patients identified by an electronic health record algorithm to receive a patient portal message read the message, 40% then visited a web-based decision aid for lung cancer screening, and 35% completed questionnaires to determine their eligibility for lung cancer screening.

Limitations

Our study had several limitations. Despite responding to inquiries on flu vaccinations receipt at higher rates, the interactive voice response group did not show a significant impact on vaccination rates (while the portal message group showed a small statistically significant improvement in vaccination rates). This discrepancy could be due in part to differences in the baseline characteristics of portal nonusers compared to portal users. For several reasons, interactive voice response call recipients may have been more resistant to completion of influenza vaccination than portal message recipients. Portal nonusers were less likely to have had an office visit in the previous 12 months and might have been less actively engaged in their health care overall. In addition, eligibility for the interactive voice response study was intentionally more stringent, with patients included only if they lacked documentation of an influenza vaccination in the previous flu season. This study design choice was made with guidance from the medical center where our study was implemented and was intended to make available the limited resource of interactive voice response calls to the broadest possible number of nonadherent patients (our study covered the cost for only 15,000 calls but we had no limit on the number of portal messages that we could send). It is possible that the interactive voice response calls might have yielded a greater impact if we had not opted to employ this more stringent eligibility criterion.

Conclusions

There are compelling reasons to use existing functionality within electronic health record–tethered portals to promote influenza vaccination. For vaccination outreach, data recorded in the electronic health record through routine care delivery can inform real-time identification of unvaccinated populations. Portal-based outreach can be more cost-effective than phone calls or mailings and easier to implement than a new app; simple messages can be sent out by office staff without additional informatics expertise. Studies [37,38,53,54] show that patient portals can enhance patient empowerment and sense of autonomy, enhance patient engagement, improve medication adherence, decrease office visits, increase self-management of disease and disease awareness, increase use of preventive medicine, and increase inclusion of patients in medical decision making.

Our parallel interventions to patient portal users and portal nonusers allowed us to assess the impact of outreach supporting influenza vaccination in both groups. Our study demonstrated a small improvement in influenza vaccination rates among portal message recipients, and successful patient engagement in both portal message recipients and interactive voice response call recipients. This method is readily applicable to current practice.

As we simultaneously face the upcoming influenza season and the ongoing COVID-19 pandemic, expanding influenza vaccination coverage in ambulatory populations can decrease the strain on our overtaxed health system and may help avert poor outcomes in patients at risk for co-infection with influenza and COVID-19.

Our findings of a small but significant improvement in influenza vaccination rates resulting from portal-based outreach represent an important contribution to the national conversation on caring for and protecting our patients in the upcoming months and for years to come.

Acknowledgments

This study was funded by Pfizer Independent Grants for Learning and Change. The authors designed and implemented the study with full oversight and responsibility for data collection, analysis, and manuscript preparation. Our study team acknowledges the contributions made by Eliza, an HMS Company, toward the design and implementation of the interactive voice response calls. A portion of MIF's work was supported by the National Heart, Lung, and Blood Institute through Grant 1K12HK138049-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of portal outreach message.
[PDF File (Adobe PDF File), 237 KB - jmir_v22i9e16373_app1.pdf]

Multimedia Appendix 2
Example of interactive voice response call script.
[PDF File (Adobe PDF File), 159 KB - jmir_v22i9e16373_app2.pdf]

References

1. Molinari NM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 2007 Jun 28;25(27):5086-5096. [doi: 10.1016/j.vaccine.2007.03.046] [Medline: 17544181]
2. Gostin LO, Salmon DA. The Dual Epidemics of COVID-19 and Influenza: Vaccine Acceptance, Coverage, and Mandates. JAMA 2020 Jul 28;324(4):335-336. [doi: 10.1001/jama.2020.10802] [Medline: 32525519]
3. Antony SJ, Almaghlouth NK, Heydemann EL. Are coinfections with COVID-19 and influenza low or underreported? An observational study examining current published literature including three new unpublished cases. J Med Virol 2020 Jun 12:26167. [doi: 10.1002/jmv.26167] [Medline: 32530531]
4. Yue H, Zhang M, Xing L, Wang K, Rao X, Liu H, et al. The epidemiology and clinical characteristics of co-infection of SARS-CoV-2 and influenza viruses in patients during COVID-19 outbreak. J Med Virol 2020 Jun 12:26163 [FREE Full text] [doi: 10.1002/jmv.26163] [Medline: 32530499]
5. NIH fact sheets - influenza. NIH. 2011. URL: https://archives.nih.gov/asites/report/09-09-2019/report.nih.gov/nihfactsheets/ViewFactSheet44bd.html?csid=133&key=I#I [accessed 2019-09-27]
6. National CFHS. Health, United States. Health, United States 2015;With special feature on racial and ethnic health disparities. Hyattsville (MD):104-112. [doi: 10.15620/cdc:82212]
7. Kochanek KD, Murphy SL, Xu J. Deaths: Final Data for 2011. Natl Vital Stat Rep 2015 Jul 27;63(3):1-120 [FREE Full text] [Medline: 26222597]
8. Kochanek KD, Xu J, Murphy SL, Miniño AM, Kung H. Deaths: final data for 2009. Natl Vital Stat Rep 2011 Dec 29;60(3):1-116 [FREE Full text] [Medline: 24974587]
9. Murphy SL, Kochanek KD, Xu J, Heron M. Deaths: Final Data for 2012. Natl Vital Stat Rep 2015 Aug 31;63(9):1-117 [FREE Full text] [Medline: 26759855]
10. Murphy SL, Xu J, Kochanek KD. Deaths: final data for 2010. Natl Vital Stat Rep 2013 May 8;61(4):1-117 [FREE Full text] [Medline: 24979972]
11. Xu J, Murphy SL, Kochanek KD, Bastian BA. Deaths: Final Data for 2013. Natl Vital Stat Rep 2016 Feb 16;64(2):1-119 [FREE Full text] [Medline: 26905861]
12. Prevention CDC. Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States - 2017-2018 influenza season. 2019. URL: https://www.cdc.gov/flu/about/burden/2017-2018.htm [accessed 2019-09-27]
13. Rolfes MA, Flannery B, Chung J, O'Halloran A, Garg S, Belongia EA, U.S. Flu VE Network, the Influenza Hospitalization Surveillance Network (FluSurv-NET),the Assessment Branch, Immunization Services Division, Centers for Disease ControlPrevention. Effects of Influenza Vaccination in the United States during the 2017-2018 Influenza Season. Clin Infect Dis 2019 Feb 02:1845-1853. [doi: 10.1093/cid/ciz075] [Medline: 30715278]
14. Prevention CDC. Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States - 2018-2019 influenza season. 2020. URL: https://www.cdc.gov/flu/about/burden/2018-2019.html [accessed 2019-09-27]
15. Frequently asked flu questions 2018-2019 influenza season. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/flu/about/season/flu-season-2018-2019.htm
16. National Vaccine Advisory Committee. A pathway to leadership for adult immunization: recommendations of the National Vaccine Advisory Committee: approved by the National Vaccine Advisory Committee on June 14, 2011. Public Health Rep 2012;127 Suppl 1:1-42 [FREE Full text] [Medline: 22210957]
17. Appel A. Improving adult immunization rates: overcoming barriers. Am Fam Physician 2011 Nov 01;84(9):977-978 [FREE Full text] [Medline: 22046935]
18. Bean-Mayberry B, Yano EM, Mor MK, Bayliss NK, Xu X, Fine MJ. Does sex influence immunization status for influenza and pneumonia in older veterans? J Am Geriatr Soc 2009 Aug;57(8):1427-1432 [FREE Full text] [doi: 10.1111/j.1532-5415.2009.02316.x] [Medline: 19515114]
19. Vaughn JA, Miller RA. Update on immunizations in adults. Am Fam Physician 2011 Nov 01;84(9):1015-1020 [FREE Full text] [Medline: 22046942]
20. Estimates of influenza vaccination coverage among adults - United States, 2017-18 flu season. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/flu/fluvaxview/coverage-1718estimates.htm [accessed 2019-06-14]
21. HealthyPeople.gov. Immunization and infectious diseases data details. URL: https://www.healthypeople.gov/node/3527/data-details [accessed 2019-10-17]
22. Early-season flu vaccination coverage - United States, November 2018. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/flu/fluvaxview/nifs-estimates-nov2018.htm [accessed 2019-06-14]
23. Prevention CDC. Flu Vaccination Coverage, United States, 2018-19 Influenza Season. 2019. URL: https://www.cdc.gov/flu/fluvaxview/coverage-1819estimates.htm [accessed 2019-09-27]

24. Clinical decision support: more than just 'alerts' tipsheet. Clinical decision support: more than just 'alerts' tipsheet. US Centers for Medicare & Medicaid Service. URL: https://www.cms.gov/regulations-and-guidance/legislation/EHRincentiveprograms/downloads/clinicaldecisionsupport_tipsheet-.pdf
25. Cutrona SL, Sreedhara M, Goff SL, Fisher LD, Preusse P, Jackson M, et al. Improving Rates of Influenza Vaccination Through Electronic Health Record Portal Messages, Interactive Voice Recognition Calls and Patient-Enabled Electronic Health Record Updates: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2016 May 06;5(2):e56 [FREE Full text] [doi: 10.2196/resprot.5478] [Medline: 27153752]
26. Cutrona SL, Golden JG, Goff SL, Ogarek J, Barton B, Fisher L, et al. Improving Rates of Outpatient Influenza Vaccination Through EHR Portal Messages and Interactive Automated Calls: A Randomized Controlled Trial. J Gen Intern Med 2018 May;33(5):659-667 [FREE Full text] [doi: 10.1007/s11606-017-4266-9] [Medline: 29383550]
27. Cutrona SL, Fouayzi H, Burns L, Sadasivam RS, Mazor KM, Gurwitz JH, et al. Primary Care Providers' Opening of Time-Sensitive Alerts Sent to Commercial Electronic Health Record InBaskets. J Gen Intern Med 2017 Nov;32(11):1210-1219. [doi: 10.1007/s11606-017-4146-3] [Medline: 28808942]
28. Amroze A, Field TS, Fouayzi H, Sundaresan D, Burns L, Garber L, et al. Use of Electronic Health Record Access and Audit Logs to Identify Physician Actions Following Noninterruptive Alert Opening: Descriptive Study. JMIR Med Inform 2019 Feb 07;7(1):e12650 [FREE Full text] [doi: 10.2196/12650] [Medline: 30730293]
29. Cash JJ. Alert fatigue. Am J Health Syst Pharm 2009 Dec 01;66(23):2098-2101. [doi: 10.2146/ajhp090181] [Medline: 19923309]
30. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
31. Ridgely M, Greenberg M. Too many alerts, too much liability: sorting through the malpractice implications of drug-drug interaction clinical decision support. J Health Law Policy 2012;5:257-296 [FREE Full text]
32. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. Health Aff (Millwood) 2011 Dec;30(12):2310-2317 [FREE Full text] [doi: 10.1377/hlthaff.2010.1111] [Medline: 22147858]
33. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
34. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
35. HealthIT.gov. What is a patient portal?. 2014. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2019-09-27]
36. Britto MT, Hesse EA, Kamdar OJ, Munafo JK. Parents' perceptions of a patient portal for managing their child's chronic illness. J Pediatr 2013 Jul;163(1):280-1.e1. [doi: 10.1016/j.jpeds.2013.02.041] [Medline: 23541773]
37. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
38. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
39. Odone A, Ferrari A, Spagnoli F, Visciarelli S, Shefer A, Pasquarella C, et al. Effectiveness of interventions that apply new media to improve vaccine uptake and vaccine coverage. Hum Vaccin Immunother 2015;11(1):72-82 [FREE Full text] [doi: 10.4161/hv.34313] [Medline: 25483518]
40. Jacobson VJC, Szilagyi P. Patient reminder and patient recall systems to improve immunization rates. Cochrane Database Syst Rev 2005(3):CD003941. [doi: 10.1002/14651858.CD003941.pub2] [Medline: 16034918]
41. Szilagyi PG, Albertin C, Casillas A, Valderrama R, Duru OK, Ong MK, et al. Effect of Patient Portal Reminders Sent by a Health Care System on Influenza Vaccination Rates: A Randomized Clinical Trial. JAMA Intern Med 2020 Jul 01;180(7):962-970. [doi: 10.1001/jamainternmed.2020.1602] [Medline: 32421168]
42. Orbell S, Hodgkins S, Sheeran P. Implementation Intentions and the Theory of Planned Behavior. Pers Soc Psychol Bull 1997 Sep;23(9):945-954. [doi: 10.1177/0146167297239004] [Medline: 29506445]
43. Pirolli P, Mohan S, Venkatakrishnan A, Nelson L, Silva M, Springer A. Implementation Intention and Reminder Effects on Behavior Change in a Mobile Health System: A Predictive Cognitive Model. J Med Internet Res 2017 Nov 30;19(11):e397. [doi: 10.2196/jmir.8217] [Medline: 29191800]
44. Thomas RE, Lorenzetti DL. Interventions to increase influenza vaccination rates of those 60 years and older in the community. Cochrane Database Syst Rev 2018 May 30;5:CD005188 [FREE Full text] [doi: 10.1002/14651858.CD005188.pub4] [Medline: 29845606]
45. Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. JAMA 2012 Apr 25;307(16):1702-1708. [doi: 10.1001/jama.2012.502] [Medline: 22535855]
46. Esteban-Vasallo MD, Domínguez-Berjón MF, García-Riolobos C, Zoni AC, Aréjula Torres JL, Sánchez-Perruca L, et al. Effect of mobile phone text messaging for improving the uptake of influenza vaccination in patients with rare diseases. Vaccine 2019 Aug 23;37(36):5257-5264. [doi: 10.1016/j.vaccine.2019.07.062] [Medline: 31353257]

47. Regan AK, Bloomfield L, Peters I, Effler PV. Randomized Controlled Trial of Text Message Reminders for Increasing Influenza Vaccination. Ann Fam Med 2017 Nov;15(6):507-514 [FREE Full text] [doi: 10.1370/afm.2120] [Medline: 29133488]
48. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
49. Lau AYS, Sintchenko V, Crimmins J, Magrabi F, Gallego B, Coiera E. Impact of a web-based personally controlled health management system on influenza vaccination and health services utilization rates: a randomized controlled trial. J Am Med Inform Assoc 2012;19(5):719-727 [FREE Full text] [doi: 10.1136/amiajnl-2011-000433] [Medline: 22582203]
50. Fischer GS, Hess R, Landeen BM, Weimer M, Zieth CR, Dong X, et al. Electronic reminders to patients within an interactive patient health record. Telemed J E Health 2013 Jun;19(6):497-500. [doi: 10.1089/tmj.2012.0116] [Medline: 23611639]
51. Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. Arch Intern Med 2011 Apr 11;171(7):636-641 [FREE Full text] [doi: 10.1001/archinternmed.2010.467] [Medline: 21149743]
52. Dharod A, Bellinger C, Foley K, Case LD, Miller D. The Reach and Feasibility of an Interactive Lung Cancer Screening Decision Aid Delivered by Patient Portal. Appl Clin Inform 2019 Jan;10(1):19-27. [doi: 10.1055/s-0038-1676807] [Medline: 30625501]
53. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A Web-Based Patient Portal for Mental Health Care: Benefits Evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
54. Sieck CJ, Hefner JL, McAlearney AS. Improving the patient experience through patient portals: Insights from experienced portal users. Patient Experience Journal 2018 Nov 06;5(3):47-54. [doi: 10.35680/2372-0247.1269]

Abbreviations

CDC: Centers for Disease Control and Prevention
OR: odds ratio

Edited by G Eysenbach; submitted 21.10.19; peer-reviewed by C Kruse, A Scherer; comments to author 04.12.19; revised version received 24.06.20; accepted 03.08.20; published 25.09.20.

Please cite as:
Wijesundara JG, Ito Fukunaga M, Ogarek J, Barton B, Fisher L, Preusse P, Sundaresan D, Garber L, Mazor KM, Cutrona SL
Electronic Health Record Portal Messages and Interactive Voice Response Calls to Improve Rates of Early Season Influenza Vaccination: Randomized Controlled Trial
J Med Internet Res 2020;22(9):e16373
URL: http://www.jmir.org/2020/9/e16373/
doi:10.2196/16373
PMID:32975529

Original Paper

DenTiUS Plaque, a Web-Based Application for the Quantification of Bacterial Plaque: Development and Usability Study

Nicolás Vila-Blanco[1,2], MEng; Vicente Freire[1], BEng; Carlos Balsa-Castro[2,3], BEng, MSc; Inmaculada Tomás[2,3], DDS, MSc, PhD; María J Carreira[1,2], MSc, PhD

[1]Centro Singular de Investigación en Tecnoloxías Intelixentes, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

[2]Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain

[3]Oral Sciences Research Group, Department of Surgery and Medical Surgical Specialities, School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Corresponding Author:
María J Carreira, MSc, PhD
Centro Singular de Investigación en Tecnoloxías Intelixentes
Universidade de Santiago de Compostela
C/Jenaro de la Fuente, s/n
Santiago de Compostela, E15782
Spain
Phone: 34 881816430
Email: mariajose.carreira@usc.es

Abstract

Background: In the dentistry field, the analysis of dental plaque is vital because it is the main etiological factor in the 2 most prevalent oral diseases: caries and periodontitis. In most of the papers published in the dental literature, the quantification of dental plaque is carried out using traditional, non-automated, and time-consuming indices. Therefore, the development of an automated plaque quantification tool would be of great value to clinicians and researchers.

Objective: This study aimed to develop a web-based tool called DenTiUS and various clinical indices to evaluate dental plaque levels using image analysis techniques.

Methods: The tool was executed as a web-based application to facilitate its use by researchers. Expert users are free to define experiments, including images from either a single patient (to observe an individual plaque growth pattern) or several patients (to perform a group characterization) at a particular moment or over time. A novel approach for detecting visible plaque has been developed as well as a new concept known as nonvisible plaque. This new term implies the classification of the remaining dental area into 3 subregions according to the risk of accumulating plaque in the near future. New metrics have also been created to describe visible and nonvisible plaque levels.

Results: The system generates results tables of the quantitative analysis with absolute averages obtained in each image (indices about visible plaque) and relative measurements (indices about visible and nonvisible plaque) relating to the reference moment. The clinical indices that can be calculated are the following: plaque index of an area per intensity (API index, a value between 0 and 100), area growth index (growth rate of plaque per unit of time in hours; percentage area/hour), and area time index (the time in days needed to achieve a plaque area of 100% concerning the initial area at the same moment). Images and graphics can be obtained for a moment from a patient in addition to a full report presenting all the processing data. Dentistry experts evaluated the DenTiUS Plaque software through a usability test, with the best-scoring questions those related to the workflow efficiency, value of the online help, attractiveness of the user interface, and overall satisfaction.

Conclusions: The DenTiUS Plaque software allows automatic, reliable, and repeatable quantification of dental plaque levels, providing information about area, intensity, and growth pattern. Dentistry experts recognized that this software is suitable for quantification of dental plaque levels. Consequently, its application in the analysis of plaque evolution patterns associated with different oral conditions, as well as to evaluate the effectiveness of various oral hygiene measures, can represent an improvement in the clinical setting and the methodological quality of research studies.

(J Med Internet Res 2020;22(9):e18570) doi:10.2196/18570

KEYWORDS

computer-aided diagnoses; computer-based biomedical applications; dental health; dental plaque quantification; web-based tools; medical informatics

Introduction

Dental plaque is a diverse community of microorganisms located on dental surfaces in the form of a biofilm embedded in an extracellular matrix of polymers from both the host and microbiota [1,2]. It is known that plaque is directly related to the appearance and progression of common oral pathologies like caries and periodontal diseases [3,4]. Monitoring of how it develops is, therefore, a topic of great clinical importance when it comes to establishing better strategies for the control of oral diseases caused by bacterial biofilms [5]. Several clinical indices for quantifying dental plaque have been developed over recent decades. These have been frequently used by the research community and in the clinical setting to evaluate the efficacy of different oral hygiene products. Some of these traditional plaque indices include those developed by Ramfjord [6], Greene and Vermillion [7], and Quigley and Hein [8] and later modified by Turesky et al [9], Löe [10], and O'Leary [11]. In general, most of these conventional indices employ an ordinal scale as part of a simple and semiquantitative method to evaluate surfaces covered by dental plaque. Their application, however, has major limitations, given the great subjectivity inherent in conducting visual examinations. Furthermore, the visual method is very imprecise when plaque levels are low or particularly high, and such clinical investigations are often laborious [12].

The planimetric method [13] was an improvement and involved taking a photograph of dyed plaque and determining its extent. This approach was much more accurate, as it employed a more objective measure to assess plaque levels (it produced a continuous, rather than ordinal, output). Additionally, as image sensors have improved year-on-year, the quality of the images generated is better than ever. Nevertheless, this is still a time-consuming process because the teeth and plaque regions in each image must be outlined by hand.

It was only in the 21st century that experts began to rely on techniques that employ analyses of digital images to quantify dental plaque. The main approach involved the planimetric method, with an imaging tool used to detect the dental plaque and tooth areas individually and calculate the ratio between them. An expert operator could outline both regions manually using a graphical interface [14,15]. Also possible were semiautomatic approaches, whereby the image-processing algorithm required intervention from the dental expert to work [16,17], or images could be segmented automatically using image-processing techniques [18-20].

Some researchers used image-processing software [16,17,21-23] or general-purpose data-processing tools [24,25], while others developed their own methods to process these images [24,26]. More recently, specific dental assessment software has been used to quantify plaque levels [18,27].

The computer vision techniques used previously vary from rudimentary to extremely complex. One of the simplest methods was image thresholding, which made it possible to isolate two or more different areas according to their color or light intensities. This technique was able to distinguish between disclosed plaque and nonplaque [24,28]; teeth, plaque, and gingiva pixels [23]; and isolated teeth, gums, plaque, and background areas [29,30]. More sophisticated machine learning algorithms were subsequently developed to enhance the results. Carter et al [20], for example, created a database of more than 600,000 pixels to analyze the relationship between pixel information (in both the RGB and HSI space) and pixel location (disclosed plaque, tooth, and gingivae). The information in this large table enabled a further step to be taken to create a classifier capable of labeling a pixel in a new image in its most probable location. Clustering methods have also been employed. An example is the approach adopted by Kang et al [31], who used a combination of fuzzy c-means and cellular neural network algorithms to classify image pixels into plaque, tooth surfaces, and backgrounds. A mean-shift based clustering algorithm was also used for plaque segmentation [19].

The different methods available today produce promising results, with many studies reporting the suitability of automatic or semiautomatic techniques for assessing dental plaque and, as a result, oral hygiene levels [16-18,24,27,29,32]. Specifically, quantitative light‐induced fluorescence digital (QLF-D) is an adaptation of QLF that employs a modified filter set (D007; Inspektor Research Systems BV, Amsterdam, The Netherlands), narrow‐band violet light (405 nm), and a high‐specification digital single‐lens reflex camera. This configuration has been specifically developed to enhance the visualization and quantification of dental plaque [33-35].

Following this trend of digital development [36], we present DenTiUS Plaque, which is a tool we have developed to enable the automatic assessment of dental plaque levels. DenTiUS Plaque was first envisaged as a standalone application. However, as we wanted to permit web access for expert users, a more general platform (DenTiUS Lab) was designed to separate the processing stages (DenTiUS Plaque proper module) and enable the execution of administrative tasks like user and patient management. The system also has a common interface where users can log in, include patients in the database, and interact with the DenTiUS Plaque module independently. The entire platform is web-based, so it can run on any web browser. It also has a friendly operator interface for nontechnical users.

The rest of the paper is structured as follows. The Methods section first describes the entire platform, including its main features and software architecture, and also contains a brief description of the patient database. Then, it explains DenTiUS Plaque, describing the kinds of images it can process, the experiment's design, and the processing algorithms. In the Results section, the different parameters of the quantitative analysis of dental plaque offered by the tool are presented, as well as several graphics, through a real case. Subsequently, the results of a questionnaire to individuals working in the field of dentistry on the usability of software are presented to evaluate

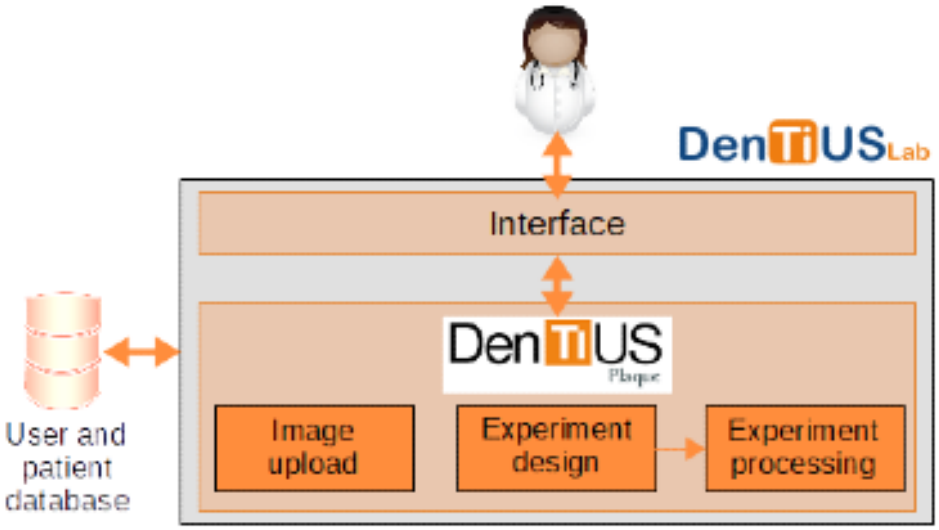
the ease of use of the instrument and its usefulness. Finally, the paper closes with a discussion and some closing remarks and identifies possible future improvements.

Methods

DenTiUS Structure

The DenTiUS platform emerged naturally as a way to manage users, images, and patients in dentistry research. It was initially developed as a standalone application for the quantification of dental plaque, with modules for managing the patients and experiments of a unique user, but not the different profiles and interactions of multiple users. As our research group is also working in other dentistry fields, all of which employ a shared patient and user database, the decision was made to develop an entire web-based system, DenTiUS Lab, for dental assessment experiments. DenTiUS Lab integrates DenTiUS Plaque's structure into a modular platform that isolates the user interface and the patient and user database from the main DenTiUS Plaque module. Figure 1 portrays a general block diagram of the complete platform, which was designed as a content management system. This allows clinicians and researchers to sign up, register, manage patients, and interact with DenTiUS Plaque by uploading plaque images and designing and processing experiments.

Figure 1. Block diagram of DenTiUS Plaque inside the DenTiUS global platform. The modular design facilitates the easy connection of the user and patient database to the DenTiUS Plaque tool.



DenTiUS was designed as a global platform to support clinical dental research through the inclusion of different modules that enable users to obtain a set of measures, graphs, and images with which to easily reach conclusions about their experiments. It was developed as a web application for several reasons: (1) most personal computers come with a pre-installed web browser, meaning that an installation process that could be complicated for nonspecialist users is avoided; (2) platform updates (including interface improvements, bug fixes, new features, and new modules) are deployed in a process that is completely transparent, but also unintrusive, which also ensures that all users are employing the most up-to-date application, removing any requirement to support old versions or manage compatibility issues between them; and (3) as having a web browser is the application's only requirement, it can be used on both computers and mobile devices.

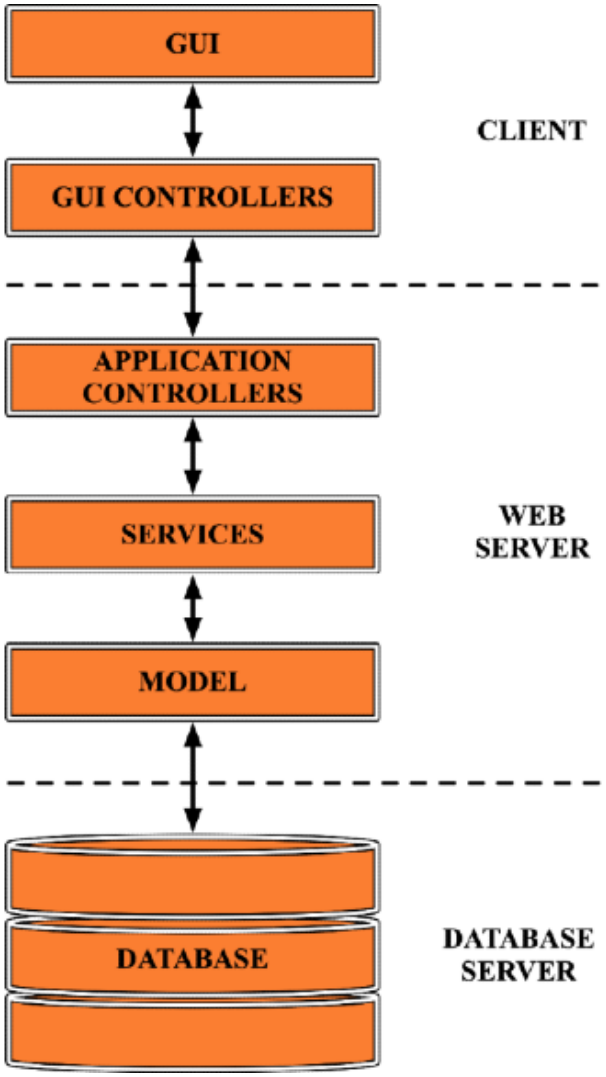
The platform was implemented as a client-server application, which runs on a central server where all the data are stored (Figure 2). All the processing tasks are carried out on the web server, so the client does not need a high-performance device. The application logic inside the server is organized into controllers (to process user requests), services (to perform operations), and models (to manage database operations), while the well-known Model-View-Controller pattern is used to manage entities, data access, dependency injections, and many other elements of the application.

The Java programming language was employed when developing the platform because of its popularity, power, verbosity, and ease of maintenance [37]. The Model-View-Controller pattern, meanwhile, was constructed with the Spring framework [38], and data storage was managed with both Hibernate [39], which is an object-relational mapping library, and a PostgreSQL database [40]. Image processing algorithms were executed in OpenCV [41], specifically its Java implementation [42]. As dental images are usually very large, a limit of 5 Mb was set in relation to the algorithm results to ensure accuracy, with the application resizing them automatically if required.

In summary, DenTiUS was developed with maintainability and extensibility as the main objectives. In this way, the application was implemented by following a modular structure, where the visual, logic, and data layers were isolated. The functionality in each layer was divided into classes, with many abstractions available to improve the extensibility. Consequently, the only

elements requiring implementation (if necessary) to develop a new module related to dental research are the data definition (eg, images, datasets), description of a new kind of experiment and its processing algorithms, and a customized report format (ie, another module with the same structure as DenTiUS Plaque, as seen in Figure 1).

Figure 2. Structural diagram of the DenTiUS platform, which is divided into a client side, web server, and database server. GUI: graphical user interface.

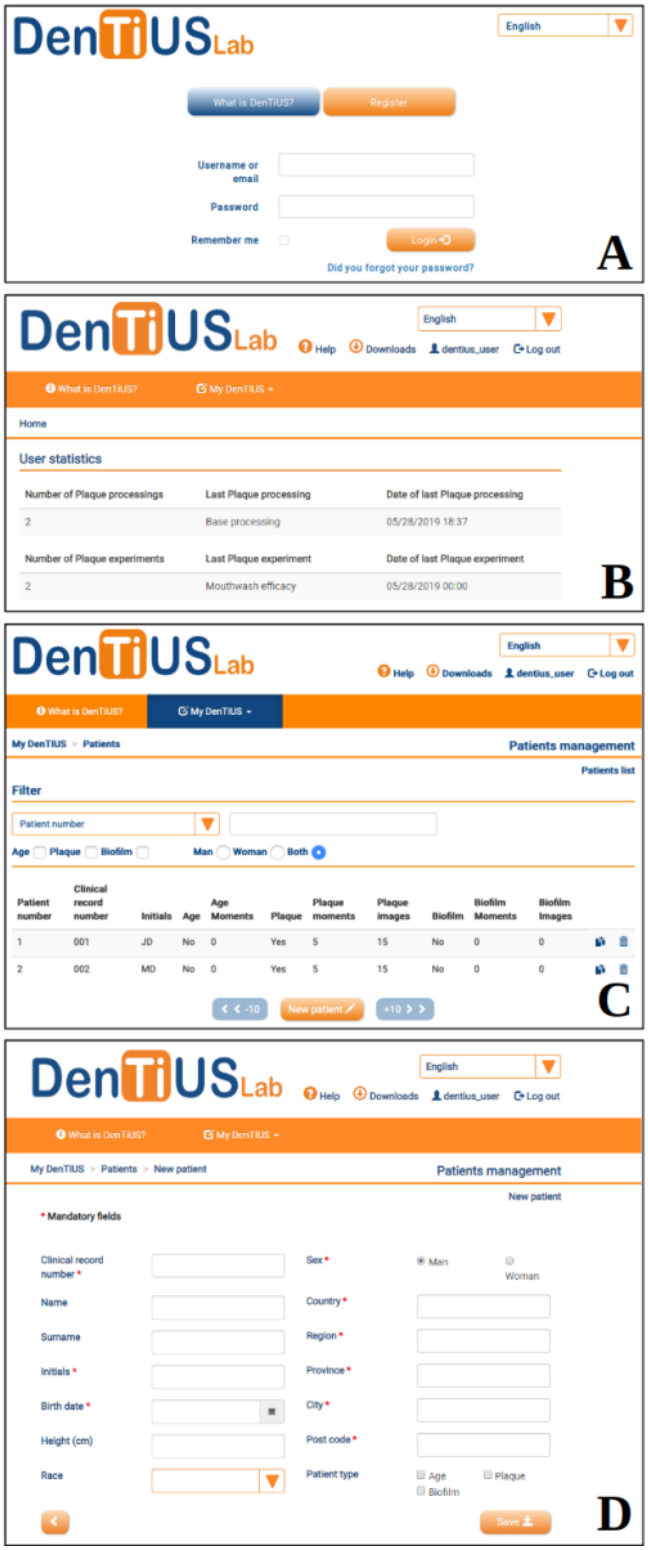


The DenTiUS Lab input screen (Figure 3A) asks the user to enter the application or register. If a user is new to the system, registration is a very straightforward process and takes only seconds to answer basic questions such as name, affiliation, and level of expertise. Once logged in, the system presents the statistics of use, namely finished processings, experiments, and patients (Figure 3B).

Every section of DenTiUS Lab is available from both the menu bar and side menu (visible when enough space is available). Access to other utilities is also possible through the common interface bar at the top of the screen (see Figure 3). These include a “Home” section, which is available by clicking on the DenTiUS Lab icon and is where the operator's statistics of use are displayed; “Help” page that explains the platform's features, details of the processing algorithms, and examples of use; “Downloads” section, where sample data can be downloaded for use in the application; “User” panel (accessible by clicking on the user name), where users can change their profile data and password; and “Language” section (English or Spanish). An “Admin” section is also available for users with administrative privileges.

A “Patient” database interface (Figure 3C) allows users to register and manage their patients, who can be filtered by patient number, clinical record number, patient initials, and sex. The user can also create new patient details by completing a form with data such as the patient's clinical record number, birth date, race, sex, city, and country (Figure 3D). The name and surname fields are optional to preserve patient anonymity.

Figure 3. Common modules of the DenTiUS structure relating to users: (A) login screen, (B) user statistics and patients, (C) patient database, (D) new patient definition.



DenTiUS Plaque

Patient Plaque Images

DenTiUS Plaque was the main objective behind our development of the DenTiUS platform. The goal was to enable DenTiUS Plaque to perform as an automated decision support system to help experts analyze and quantify the macroscopically observable dental plaque deposited on teeth. In the first step, the software requires reproducible images of the fluorescein-dyed plaque taken under ultraviolet light. Fluorescein is a well-known fluorochrome in the field of dentistry, and a patent for its use as a dental plaque marker was filed in the USA in 1967 by Herbert Brilliant (U.S. Patent 3-309-274; 1967). In these images, the bright blue region corresponds to the tooth area unaffected by plaque, while the

green region matches plaque deposits over both the tooth surfaces and gingiva.

To enable the assessment of dental plaque levels at different times, an entity called a "plaque moment" was defined for a single plaque image (usually the frontal view) or a set of images (front and lateral views) captured from a patient at a particular time. The possibility of including lateral views is a novelty of the system, as these provide a better view of the posterior teeth, where more plaque is usually accumulated.

Plaque moments can be attached to a patient's file via the Patients section (Figure 3C), which contains information about each patient concerning the number of image sets included for her or him and the number of plaque images distributed through all their moments. Another way to attach new plaque moments to a patient's record is by entering the Plaque section directly from the "MyDenTiUS" menu, where all the plaque moments are listed independently and can be modified or deleted. In any case, the user can modify, delete, or include a new moment where the images (at least one) have a date, time, brief description, and optional camera set-up parameters. At this point, it is crucial to define the so-called "reference moment" or "moment 0." This cannot be deleted, as it is used as a reference to compute the plaque-level growth over time and usually corresponds to circumstances where there has been "perfect dental cleaning."

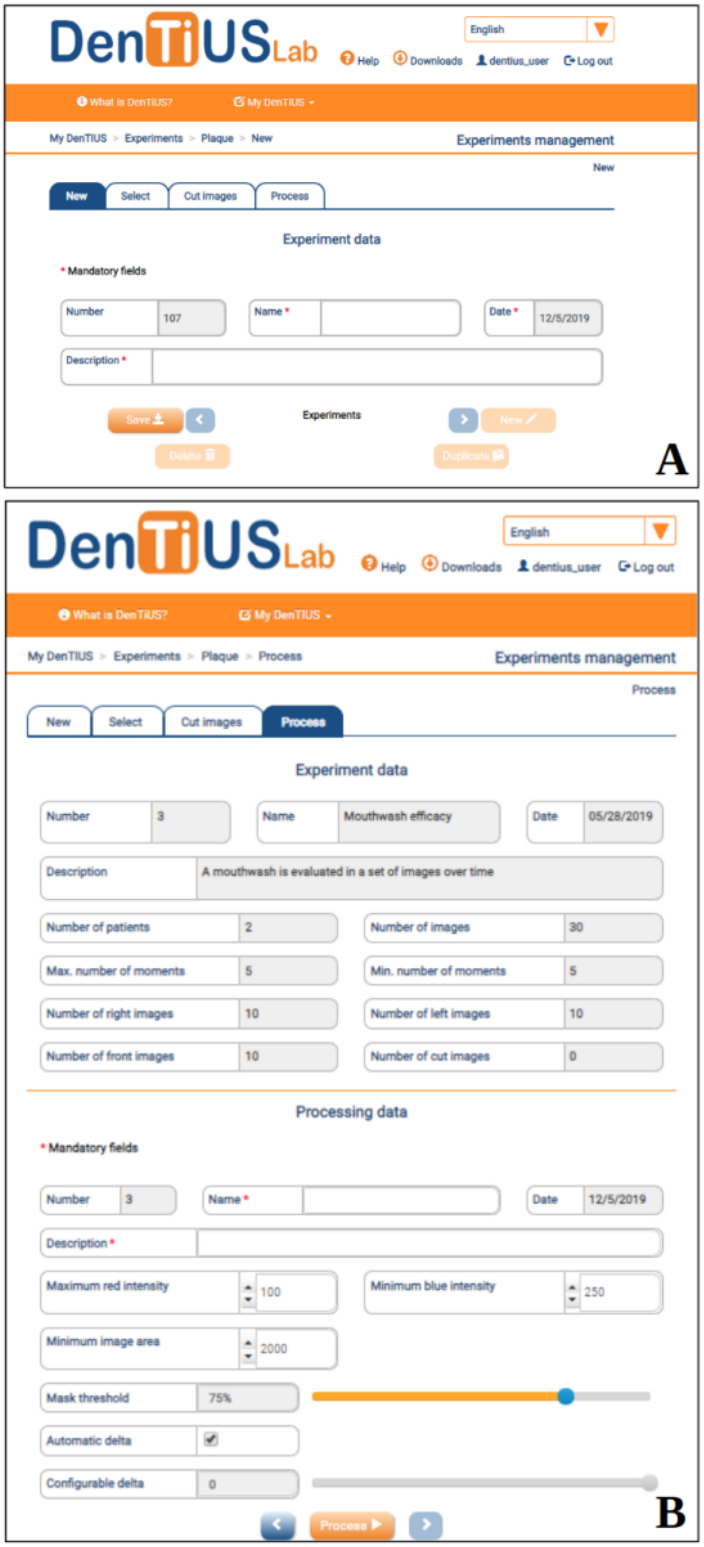
Plaque Experiment Design

The plaque experiment design module (see Figure 1) enables users to develop new experiments by selecting specific patient data and tuning the parameters of the image processing algorithms. The experiments are presented in a list view, with several shortcuts on the right side of the screen that trigger operations like duplicating, processing, deleting, or cropping the associated images. When the user creates a new experiment, a screen appears with the following tabs: New, Select, Cut Images, and Process (Figure 4). Figure 4A shows the "New" tab, on which the name and description fields are mandatory. On the "Select" tab, the user must choose the plaque moments to be processed by selecting them from the patient database. The system automatically includes the reference moment related to any selected moment. The "Cut Images" tab allows the user to optionally select part of the images or delete artifacts. Finally, the "Process" tab (Figure 4B) shows a summary of the "Experiment data" and, below that, the "Processing data" parameters, with a mandatory name and description. This permits users to try different configurations of the processing parameters for the same set of images. In this screen, the expert user can employ the default parameters or change these to different values. The meanings of these parameters will be explained later in this section.

The configuration of the experiment is decided by the user according to the specific circumstances (age groups, gender, single or multiple patients) or objectives (eg, effectiveness of mouthwash, effectiveness of brushing devices, smoking effects). The system performs the experiments defined by the user by processing the data of all the patients and all the patient moments included in the experiment. When the experiment involves multiple moments for the same patient, the images with the selected tooth areas are subjected to a process of normalization in the number of pixels.

DenTiUS Plaque does not permit comparisons between groups, as this was not the objective of the system, but users can export all the computed parameters to manage and prepare their own statistics with no limitations.

Figure 4. Plaque experiment design for the definition of a new experiment: (A) New tab and (B) Process tab, with a summary of the experiment and processing data showing the processing parameters that will be used when pushing the "Process" button at the bottom.



Processing Algorithms

After an exhaustive analysis of the color characteristics of each region of interest, a process is defined to segment the dental area, excluding the gingiva region: Initially, the blue and red channels are added together, and the result is thresholded according to a parameter fixed by the user (by default, the algorithm chooses 75% of the darkest pixels as the background and 25% of the lightest as the foreground). The isolated pixels are then removed, and a connected component analysis is performed to identify objects from that binary image. Thereafter, a set of conditions related to region size (>2000 pixels, by default) and solidity (>0.5, by default) are applied to remove nondental regions. Other specific rules, which were defined to eliminate bright artifacts, are included to make the process fully

automatic. All these parameters can be changed in the "Process" tab (see the "Processing data" parameters in Figure 4B).

Once the dental area is segmented, the system must perform an analysis of its values to extract and measure the dental plaque. As mentioned previously, plaque is characterized by a high green intensity, whereas the rest of the dental region has a high blue concentration. The first step is thus the detection of the (visible) plaque by analyzing the differences between the green and blue channels (Equation 1). The rest of the dental region is considered to be a first approximation of the nonvisible plaque area (Equation 2).

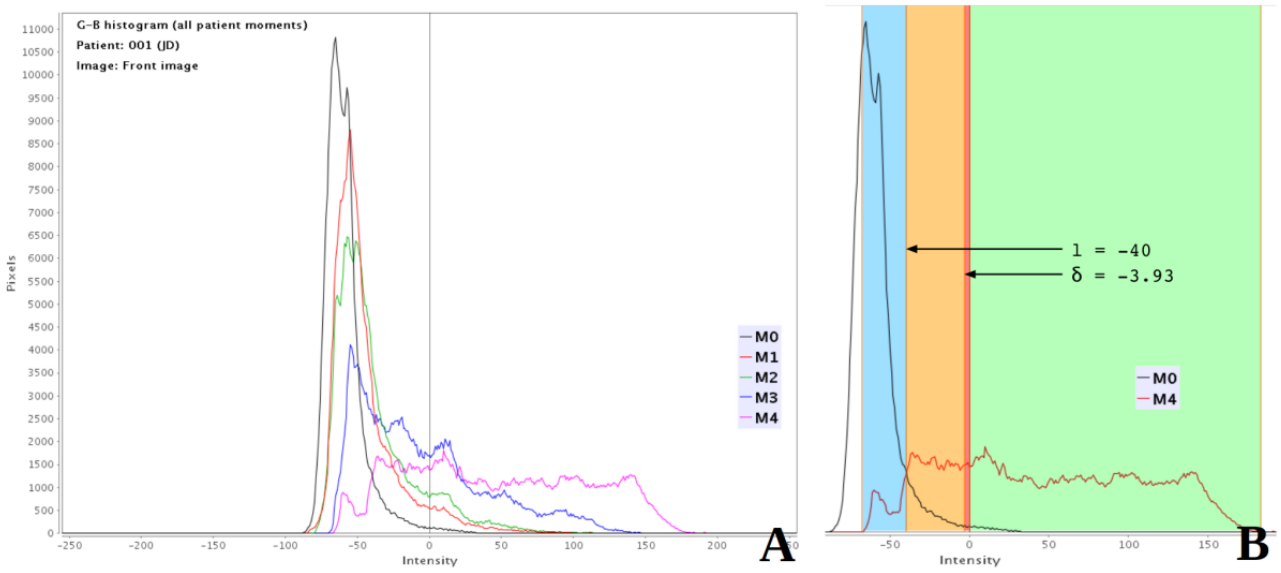
$P_{visible} = |G\text{-}B|$ for pixels with $[G\text{-}B > 0]$ (**1**)

$P_{nonvisible} = |G\text{-}B|$ for pixels with $[G\text{-}B \leq 0$ and $G>0$ and $B>0]$ (**2**)

where $P_{visible}$ is the visible plaque, $P_{nonvisible}$ is the initial approximation of the nonvisible plaque, and G and B are the green and blue channels of the dental area, respectively.

G-B dental area histograms are included to assess how levels of dental plaque increase over time when oral hygiene stops, with the positive and negative sides corresponding to the visible (G>B) and first approximation of the nonvisible plaque (G<B), respectively. Figure 5A shows an example of a 96-hour experiment, where moments were recorded every 24 hours, while Figure 6 presents the results of the different steps of the algorithm on a frontal image of a patient after 96 hours of perfect cleaning. After professional dental cleaning (moment M0), most of the histogram area corresponded to the first approximation of the nonvisible plaque area (Equation 2), as its values were highly concentrated in the negative part of the graph. In the moments that followed (M1 to M4), the histogram was flattened and displaced toward the positive side, thus increasing the visible plaque level. In the final moment (M4), 4 days after a dental cleaning, the largest part of the histogram was placed on the positive side of the graph, with the visible plaque area covering most of the dental region. The main finding behind this step was the discovery and measurement of the transition area between the visible plaque and the rest of the dental region, based on the risk of plaque developing in the near future. As a result, the system performs an extra segmentation step before displaying the nonvisible plaque area (as seen in Figure 6F).

Figure 5. G-B histogram where the vertical line represents G-B=0 and G and B are the green and blue channels, respectively. (A) G-B histograms for moments M0 to M4, showing the progression of the histogram values toward the G-B positive values (visible plaque); (B) Definition of l and δ for M4, showing plaque (green), nonplaque (blue), level 1 nonvisible plaque (red, defined by δ), and level 2 nonvisible plaque (orange, defined by l). The boundary between nonplaque and level 2 nonvisible plaque (l) is easily observable in this chart, as it corresponds to the intensity value where the 2 histograms cross.



According to 2 different parameters, δ and *l*, the blue/green differences are thresholded to produce 3 different region masks in the negative part of the graph, namely level 1 nonvisible plaque (risk of being plaque in the following hours), level 2 nonvisible plaque (risk of being plaque in the following days), and no plaque (no risk of being plaque in the medium term):

$P_{nonvisible}(level1) = |G\text{-}B|$ for pixels with $[G\text{-}B \leq 0$ and $G\text{-}B \geq \delta]$ (**3**)

$P_{nonvisible}(level2) = |G\text{-}B|$ for pixels with $[G\text{-}B < \delta$ and $G\text{-}B \geq l]$ (**4**)

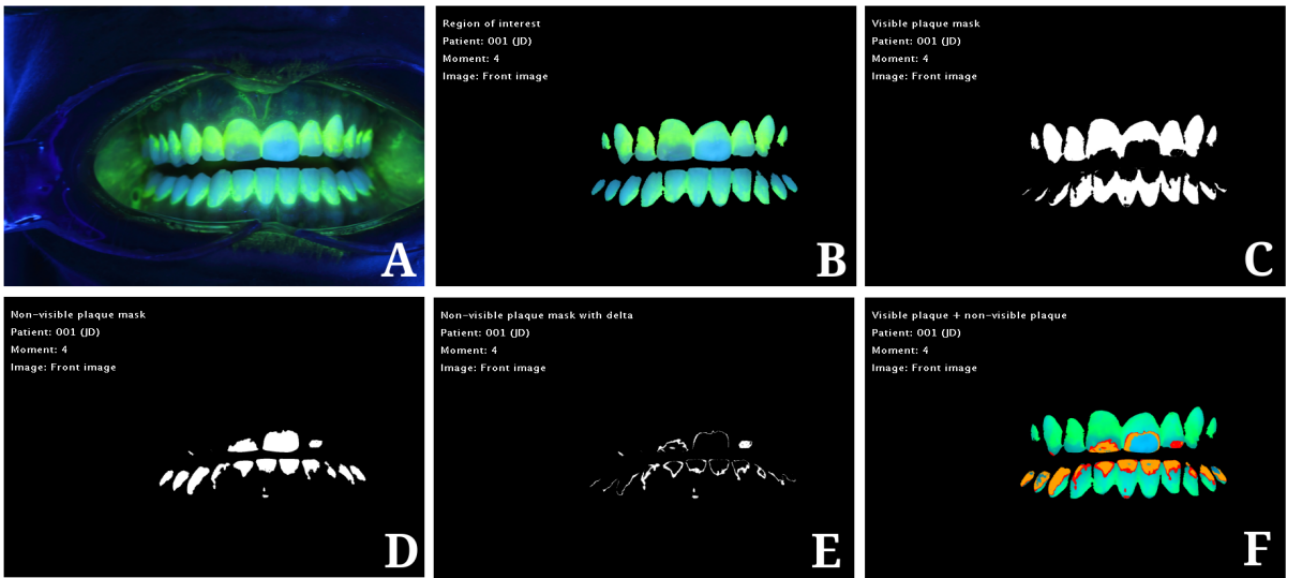
$NonPlaque = |G\text{-}B|$ for pixels with $[G\text{-}B < l]$ (**5**)

where the *l* parameter is automatically calculated from nonvisible plaque histograms of current and reference images (Figure 5B); it is defined as the first value where the number of nonvisible plaque pixels in the current image is greater than the number of nonvisible plaque pixels in the reference image (moment 0).

The δ parameter can be set by the user or is automatically resolved as the absolute difference between the average pixel values of the blue and green channels in the dental area. The modification of this parameter by the user in the "Process" tab (see Figure 4B) indicates an increase or decrease in the level 1 nonvisible plaque region (red area in Figure 6F). This means

that the pattern of plaque development can be simulated by starting with a minimum value and progressively increasing it.

Figure 6. Results of the different steps of the algorithm relating to the frontal image of a patient after 96 hours of perfect cleaning: (A) frontal ultraviolet image, (B) segmented dental area, (C) visible plaque mask, (D) nonvisible plaque mask, (E) nonplaque level 1 mask, (F) final labeled image (blue: nonplaque; green: visible plaque; red: nonvisible plaque level 1; orange: nonvisible plaque level 2).



Ethical Approval

Images from Spanish Caucasian subjects are part of a study protocol approved by the Galician Clinical Research Ethics Committee (registration number 2014/008). The image collection was performed following the ethical standards of our institution's research committee and the 1964 Declaration of Helsinki and its later amendments [43].

RESULTS

Plaque Processing Results

Figure 7 portrays the complete process for a single patient image set belonging to a particular experiment: Plaque images were uploaded to the database, the plaque experiment was designed, and the experiment processing was launched.

When the processing finishes, the user can view the results by clicking on the appropriate notification or the "Processings" submenu. The processing results are divided into 4 tabbed subsections (Figure 8): "Data and properties," "Single measurements," "Measurements download," and "Images and charts." The "Data and properties" tab presents a summary of statistics concerning the processing as well as the processing parameters. Figure 8B shows the "Single measurements" tab, which presents the results of the quantitative analysis for each image in table form, with absolute averages obtained in each image (indices about visible plaque) and relative measurements (indices about visible and nonvisible plaque) relating to the reference moment. As shown in this figure and just below the tabs with an information alert, the user must click on each image to display their processing results in the "Single measurements" and "Images and charts" tabs.

The clinical indices defined in this research provide information on the bacterial plaque on dental surfaces, specifically the area where it is present, its intensity, a combined value between the area and intensity, and different growth parameters. The system can produce dental plaque clinical indices in a computerized form, including in relation to the plaque index of an area per intensity (API index, a value between 0-100), area growth index (growth rate of plaque per unit of time in hours: percentage area/hour), and area time index (the time, in days, needed to achieve a plaque area of 100% concerning the initial area at the same moment).

In the example in Figure 8, after 96 hours of plaque accumulation, focusing on the visible plaque values, the patient presented an API index of 21.18 and a hypothetical plaque growth rate of 0.75% per hour. A particularly interesting parameter supplying the labeled image is the area time. In this case, the area-time relative value is 1.67 for visible plaque and 0.61 for total plaque (see Figure 8B). In other words, from a theoretical point of view and assuming a constant growth pattern for that moment, the patient would take 1.67 days to achieve a visible plaque level of 100%. This value is obviously lower, around 0.61 days, when it comes to realizing plaque levels of 100% for both nonvisible and visible plaque.

The measurement tables can be customized and downloaded in CSV and spreadsheet formats via the "Measurements download" tab, where users can also select specific parameters within the absolute and relative measurements.

Figure 7. Process of detecting and quantifying dental plaque using DenTiUS Plaque for one patient moment in an experiment.

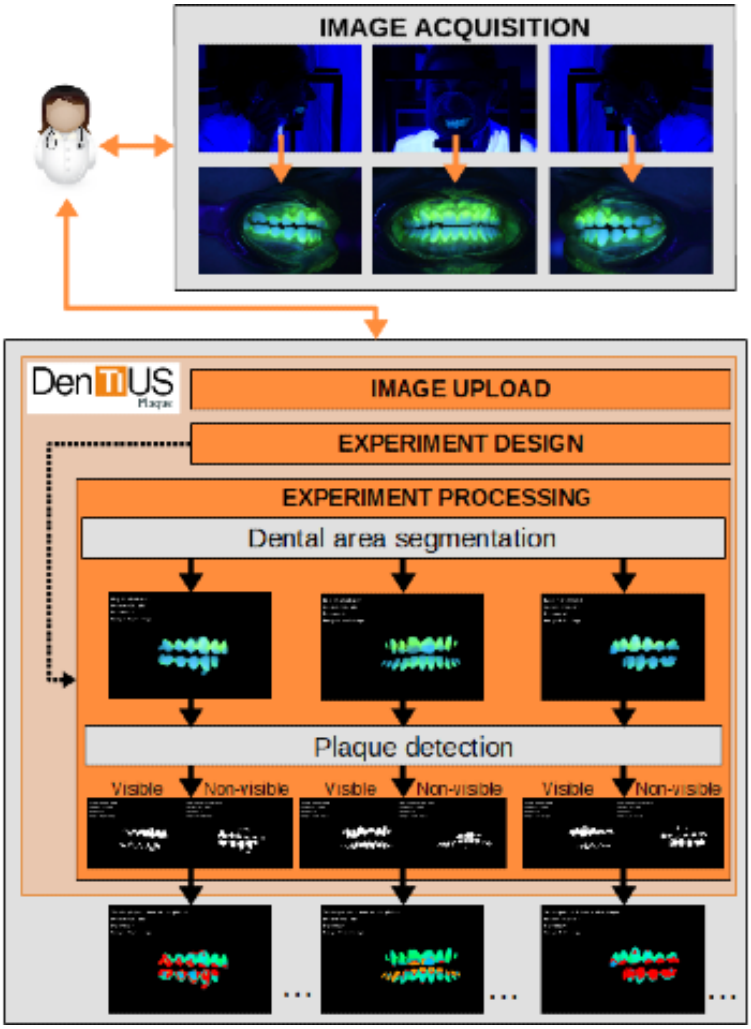


Figure 8C portrays an example of the images, tables, and charts obtained for one moment of one patient. The "Images and charts" tab provides not only the final view of the plaque regions in each plaque image but also the intermediate results of the computer vision algorithm, including the tooth mask, visible plaque mask, difference between the blue and green channels in both the teeth and visible plaque regions, and others. The final labeled image represents the areas with nonplaque (in blue) and visible plaque (in green). Also shown are the areas that will become plaque in the short term, graduating to near-plaque (red areas, nonvisible plaque level 1) and medium-term plaque (orange areas, nonvisible plaque level 2).

Pie and bar charts comparing the levels of visible plaque and nonvisible plaque, as well as the nonplaque areas, are also presented. Finally, 2 histograms are presented that compare the intensity distribution of the plaque in the selected image moment to the corresponding image of the reference moment and the intensity distributions of the plaque through all the moments. A vertical line at 0 can be seen in both histograms, separating values corresponding to visible plaque (Equation 1) from those relating to nonvisible plaque and nonplaque (Equations 3, 4, and 5). Finally, a full report presenting all the tables and graphs can be generated in PDF format via the main "Processings" menu. This will contain all the processing data, including all the information described previously for each processing tab.

Figure 8. Processing results. (A) and (B) The single measurements tab showing the absolute (first column, visible plaque) and relative measurements (relative to the reference image) for the frontal image; (A) and (C) Images and charts tab for the frontal image.

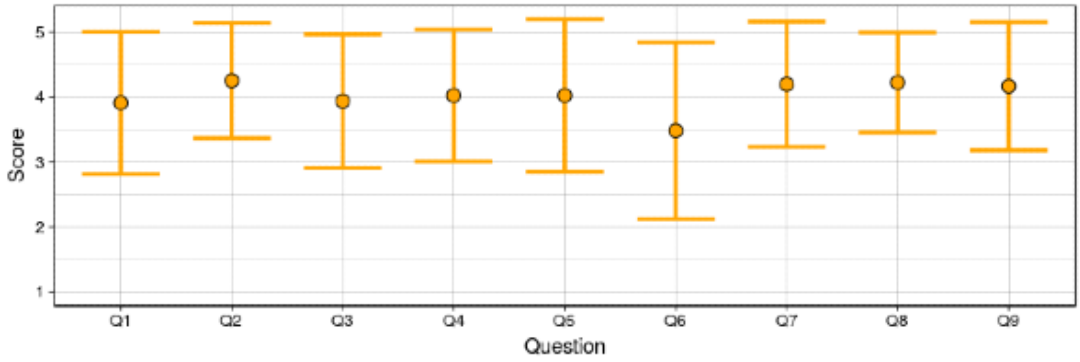
DenTiUS Graphical User Interface (GUI) Usability

As with most web applications, DenTiUS has a graphical user interface (GUI) that is designed to be friendly, simple, and intuitive, as well as adaptable to the various screen sizes on mobile phones, tablets, and computers. A usability test, based on the Computer System Usability Questionnaire [44], was proposed to assess the suitability of the GUI in real terms. This test was adapted to DenTiUS, resulting in 9 questions with scores ranging from 1 (strongly disagree) to 5 (strongly agree). The questions were designed to measure the following: (Q1) overall ease of use, (Q2) workflow efficiency, (Q3) user comfort, (Q4) learning process, (Q5) productivity, (Q6) value of the error messages, (Q7) online help, (Q8) attractiveness of the interface, and (Q9) overall satisfaction.

The questionnaire was sent to 34 dentists doing research in the field of dentistry, including mainly PhD researchers (28/34, 82%), undergraduate research fellows in their final year of career (3/34, 9%), and senior researchers (3/34, 9%). The testers were given instructions about how to enter DenTiUS Lab and proceed with a test case based on a patient with plaque growth over several days. In particular, they performed the following steps: signing up to the application, registering some patients and plaque moments, defining an experiment, continuing with the processing, and searching for various results.

Figure 9 demonstrates that all the questions produced fairly good results, with a mean score of around 4 in almost all cases. The best-scoring questions were Q2, Q7, Q8, and Q9, relating to the workflow efficiency, value of the online help, attractiveness of the user interface, and overall satisfaction, respectively. By contrast, some users stated in Q6 that the platform errors should be more informative and provide better solutions to fix any issues.

Figure 9. Score distribution of each usability test question. Each error bar represents the mean and standard deviation.



Availability of the Software

Access to the DenTiUS Lab web platform is available free of charge for non-commercial use for researchers and clinicians [45].

Discussion

Principal Findings

The importance of dental plaque in the etiopathogenesis of important oral diseases, such as caries and periodontitis, is well-known [5,46]; on the other hand, there are recognized limitations of conventional clinical indices of dental plaque quantification [20,23], which are widely used both in clinical and research settings. Therefore, to improve the diagnosis of dental plaque, it is essential to develop new computer systems that allow the objective quantification of dental plaque levels.

This paper introduces DenTiUS Plaque. This is a tool for the quantification of bacterial plaque integrated into a general web-based platform with a common management process for users and patients.

DenTiUS is the first dental research system to enable image collections and patient data to be managed, experiments to be designed, and images with customized configuration to be automatically processed. The developed tool produces accurate and repeatable results for the assessment of clinical indices of bacterial plaque levels relating to a patient or group of patients over time, ensuring the sustainability of the process in terms of the time and effort required by users of the system. Clinical users with no technical background can process the images in batches and obtain a table of measurements (most of which have been specifically produced for this platform) and explanatory graphs that can be exported in various formats (PDF report, spreadsheet, and JPEG images). Although a more complex final report could be designed, these different outputs were a requirement identified by dental researchers to enable them to have access to all the values needed to produce their own statistics.

Specifically, a novel algorithm was developed in DenTiUS Plaque to detect and quantify dental plaque levels from ultraviolet images. This approach first detects the dental region and then segments and quantifies visible plaque by analyzing the difference between green and blue channels. Initially, DenTiUS Plaque could represent a tool that offers the following advantages over the QFL-D system [33-35]. The first advantage is that DenTiUS Plaque allows the detection of different areas in the remaining dental regions according to the risk of developing plaque in the future; this finding was referred to as “nonvisible plaque.” DenTiUS Plaque provides the detection and quantification of nonvisible plaque at 2 levels: the probability of becoming visible plaque in the short term (level 1) or medium term (level 2). The second advantage is that DenTiUS Plaque not only provides the same indices of plaque quantification (the API index) as the QFL-D system but also indices to measure the plaque growth pattern over time for a given patient such as the area growth index and area time index. In addition, all the clinical indices developed for DenTiUS Plaque are applicable to quantifying both visible and nonvisible plaque levels.

Regarding the usability test conducted with people working in the dental field, overall, the users stated that they were likely to use the application in the future, with the results revealing that DenTiUS Plaque was suitable for its ultimate purpose.

Biological coherence between conventional clinical indices and new indices derived from image analysis is a study to be carried out to test the validity of the latter [24,47,48]. In this sense, our research group used an in situ 5-day bacterial plaque growth model to conduct some initial experiments on the validity of

the DenTiUS Plaque clinical indices, comparing the results with those obtained with a conventional clinical index; to obtain both types of indices, the plaque was stained with fluorescein and displayed using ultraviolet light [49,50]. Concerning the degree of correlation between the conventional and API indices, days 1, 2, and 3 of plaque formation revealed very high correlations between the two approaches (Spearman rho ≥0.770). However, in the days where there was little or excessive accumulation of bacterial plaque, days 0 and 4, respectively, the relationship between the 2 measurement systems was suboptimal (Spearman rho ≤0.540), highlighting the limitations of the conventional index and the convenience of applying the API index produced by DenTiUS Plaque for these clinical situations. An interesting objective of future research would be to apply both DenTiUS Plaque and the QFL-D system on the same group of patients to quantify the levels of dental plaque, verifying the correlation between both systems.

Limitations and Future Work

Further work will be conducted to improve the platform. Additional help will be included in the software based on the results of a GUI validation exercise.

The image cropping system will be improved to automatically exclude unwanted artifacts in the image. More research tools will be fully integrated into the DenTiUS platform. An example is DenTiUS Biofilm, which will enable the quantification of microscopic dental plaque as well as comparisons of the plaque growth pattern at the microscopic (biofilm) and macroscopic levels (plaque), taking advantage of the shared management of patients and users.

Conclusions

The DenTiUS Plaque software allows automatic, reliable, and repeatable quantification of dental plaque levels, providing information about area, intensity, and growth pattern. Dentistry experts recognized that this software is suitable for quantification of dental plaque levels. Consequently, its application in the analysis of plaque evolution patterns associated with different oral conditions as well as the evaluation of the effectiveness of various oral hygiene measures can represent an improvement in the clinical setting and methodological quality of research studies.

Acknowledgments

This work received financial support from Johnson & Johnson company (Grant 2016 - CE219), Consellería de Cultura, Educación e Ordenación Universitaria (accreditation 2019-2022 ED431G-2019/04, 2017-2020 Potential Growth Group ED431B 2017/029, 2017-2020 Competitive Reference Group ED431C 2017/69, and N. Vila-Blanco support ED481A-2017), and the European Regional Development Fund (ERDF), which acknowledges the CiTIUS-Research Center in Intelligent Technologies of the University of Santiago de Compostela as a Research Center of the Galician University System. Any requests to use the software should be made by sending an email to dentiuslab@gmail.com.

Authors' Contributions

The entire project was supervised by MJC and IT. NVB, VF, CBC, and MJC designed and implemented the entire platform. MJC, CBC, and IT supervised the implementation, designed the image processing algorithms, and defined the clinical indices.

Conflicts of Interest

None declared.

References

1. Marsh PD. Plaque as a biofilm: pharmacological principles of drug delivery and action in the sub- and supragingival environment. Oral Dis 2003;9 Suppl 1:16-22. [doi: 10.1034/j.1601-0825.9.s1.4.x] [Medline: 12974526]
2. Marsh PD. Dental plaque as a microbial biofilm. Caries Res 2004;38(3):204-211 [FREE Full text] [doi: 10.1159/000077756] [Medline: 15153690]
3. Socransky SS, Haffajee AD. Dental biofilms: difficult therapeutic targets. Periodontol 2000 2002;28:12-55. [doi: 10.1034/j.1600-0757.2002.280102.x] [Medline: 12013340]
4. Marsh PD. Dental plaque as a biofilm and a microbial community - implications for health and disease. BMC Oral Health 2006 Jun 15;6 Suppl 1:S14 [FREE Full text] [doi: 10.1186/1472-6831-6-S1-S14] [Medline: 16934115]
5. Jepsen S, Blanco J, Buchalla W, Carvalho JC, Dietrich T, Dörfer C, et al. Prevention and control of dental caries and periodontal diseases at individual and population level: consensus report of group 3 of joint EFP/ORCA workshop on the boundaries between caries and periodontal diseases. J Clin Periodontol 2017 Mar;44 Suppl 18:S85-S93. [doi: 10.1111/jcpe.12687] [Medline: 28266120]
6. Ramfjord SP. Indices for Prevalence and Incidence of Periodontal Disease. Journal of Periodontology 1959 Jan;30(1):51-59. [doi: 10.1902/jop.1959.30.1.51]
7. Greene JC, Vermillion JR. The oral hygiene index: a method for classifying oral hygiene status. The Journal of the American Dental Association 1960 Aug;61(2):172-179. [doi: 10.14219/jada.archive.1960.0177]
8. Quigley GA, Hein JW. Comparative cleansing efficiency of manual and power brushing. J Am Dent Assoc 1962 Jul;65:26-29. [doi: 10.14219/jada.archive.1962.0184] [Medline: 14489483]

9. Turesky S, Gilmore ND, Glickman I. Reduced plaque formation by the chloromethyl analogue of victamine C. J Periodontol 1970 Jan;41(1):41-43. [doi: 10.1902/jop.1970.41.41.41] [Medline: 5264376]
10. Löe H. The Gingival Index, the Plaque Index and the Retention Index Systems. J Periodontol 1967;38(6):Suppl:610-Suppl:616. [doi: 10.1902/jop.1967.38.6.610] [Medline: 5237684]
11. O'Leary T. The periodontal screening examination. J Periodontol 1967;38(6):Suppl:617-Suppl:624. [doi: 10.1902/jop.1967.38.6_part2.617] [Medline: 5237685]
12. Butler BL, Morejon O, Low SB. An accurate, time-efficient method to assess plaque accumulation. J Am Dent Assoc 1996 Dec;127(12):1763-6; quiz 1784. [doi: 10.14219/jada.archive.1996.0137] [Medline: 8990746]
13. Söder PO, Jin LJ, Söder B. Computerized planimetric method for clinical plaque measurement. Scand J Dent Res 1993 Feb;101(1):21-25. [doi: 10.1111/j.1600-0722.1993.tb01640.x] [Medline: 8441891]
14. Coulthwaite L, Pretty IA, Smith PW, Higham SM, Verran J. QLF is not readily suitable for in vivo denture plaque assessment. J Dent 2009 Nov;37(11):898-901. [doi: 10.1016/j.jdent.2009.07.002] [Medline: 19665277]
15. Raggio DP, Braga MM, Rodrigues JA, Freitas PM, Imparato JCP, Mendes FM. Reliability and discriminatory power of methods for dental plaque quantification. J Appl Oral Sci 2010;18(2):186-193 [FREE Full text] [doi: 10.1590/s1678-77572010000200014] [Medline: 20485931]
16. Rosa GM, Elizondo ML. New portable system for dental plaque measurement using a digital single-lens reflex camera and image analysis: Study of reliability and validation. J Indian Soc Periodontol 2015;19(3):279-284 [FREE Full text] [doi: 10.4103/0972-124X.152415] [Medline: 26229267]
17. Hope CK, Wang Q, Burnside G, Adeyemi AA, Quenby S, Smith PW, et al. Assessing the association between oral hygiene and preterm birth by quantitative light-induced fluorescence. Scientific World Journal 2014;2014:374694 [FREE Full text] [doi: 10.1155/2014/374694] [Medline: 24511282]
18. Han S, Kim B, Ko H, Kwon H, Kim B. Validity and reliability of autofluorescence-based quantification method of dental plaque. Photodiagnosis Photodyn Ther 2015 Dec;12(4):587-591. [doi: 10.1016/j.pdpdt.2015.10.003] [Medline: 26474950]
19. Kang J, Ji Z. Dental plaque quantification using mean-shift-based image segmentation. 2010 May 05 Presented at: International Symposium on Computer, Communication, Control and Automation (3CA); 2010; Tainan, Taiwan p. 470-473. [doi: 10.1109/3ca.2010.5533758]
20. Carter K, Landini G, Walmsley AD. Automated quantification of dental plaque accumulation using digital imaging. J Dent 2004 Nov;32(8):623-628. [doi: 10.1016/j.jdent.2004.06.006] [Medline: 15476956]
21. Sagel P, Lapujade P, Miller J, Sunberg R. Objective quantification of plaque using digital image analysis. Monogr Oral Sci 2000;17:130-143. [doi: 10.1159/000061638] [Medline: 10949837]
22. Smith RN, Lath DL, Rawlinson A, Karmo M, Brook AH. Gingival inflammation assessment by image analysis: measurement and validation. Int J Dent Hyg 2008 May;6(2):137-142. [doi: 10.1111/j.1601-5037.2008.00294.x] [Medline: 18412727]
23. Pretty IA, Edgar WM, Higham SM. A study to assess the efficacy of a new detergent free, whitening dentifrice in vivo using QLF planimetric analysis. Br Dent J 2004 Nov 13;197(9):561-6; discussion 551. [doi: 10.1038/sj.bdj.4811809] [Medline: 15543118]
24. Mohan N, Mahesh MR, Varghese VI, Pretty IA, Taylor AM, Ellwood RP. Evaluation of the sensitivity of a digital plaque imaging system on different tooth surfaces. J Clin Dent 2012;23(1):11-16. [Medline: 22435319]
25. Beerens MW, van der Veen MH, van Beek H, ten Cate JM. Effects of casein phosphopeptide amorphous calcium fluoride phosphate paste on white spot lesions and dental plaque after orthodontic treatment: a 3-month follow-up. Eur J Oral Sci 2010 Dec;118(6):610-617. [doi: 10.1111/j.1600-0722.2010.00780.x] [Medline: 21083623]
26. Yeganeh S, Lynch E, Jovanovski V, Zou L. Quantification of root surface plaque using a new 3-D laser scanning method. J Clin Periodontol 1999 Oct;26(10):692-697. [doi: 10.1034/j.1600-051x.1999.261009.x] [Medline: 10522781]
27. Alammari MR, Smith PW, de Josselin de Jong E, Higham SM. Quantitative light-induced fluorescence (QLF): a tool for early occlusal dental caries detection and supporting decision making in vivo. J Dent 2013 Feb;41(2):127-132. [doi: 10.1016/j.jdent.2012.08.013] [Medline: 22940557]
28. Smith RN, Rawlinson A, Lath DL, Brook AH. A digital SLR or intra-oral camera: preference for acquisition within an image analysis system for measurement of disclosed dental plaque area within clinical trials. J Periodontal Res 2006 Feb;41(1):55-61. [doi: 10.1111/j.1600-0765.2005.00841.x] [Medline: 16409256]
29. Bellamy PG, Jhaj R, Mussett AJ, Barker ML, Klukowska M, White DJ. Comparison of a stabilized stannous fluoride/sodium hexametaphosphate dentifrice and a zinc citrate dentifrice on plaque formation measured by digital plaque imaging (DPIA) with white light illumination. J Clin Dent 2008;19(2):48-54. [Medline: 18763686]
30. Bellamy PG, Prendergast M, Strand R, Yu Z, Day TN, Barker ML, et al. Can anti-erosion dentifrices also provide effective plaque control? Int J Dent Hyg 2011 Aug;9(3):223-228 [FREE Full text] [doi: 10.1111/j.1601-5037.2010.00480.x] [Medline: 21356021]
31. Kang J, Min L, Luan Q, Li X, Liu J. Novel modified fuzzy c-means algorithm with applications. Digital Signal Processing 2009 Mar;19(2):309-319. [doi: 10.1016/j.dsp.2007.11.005]
32. Klukowska M, Bader A, Erbe C, Bellamy P, White DJ, Anastasia MK, et al. Plaque levels of patients with fixed orthodontic appliances measured by digital plaque image analysis. Am J Orthod Dentofacial Orthop 2011 May;139(5):e463-e470. [doi: 10.1016/j.ajodo.2010.05.019] [Medline: 21536188]

33. Hope CK, de Josselin de Jong E, Field MRT, Valappil SP, Higham SM. Photobleaching of red fluorescence in oral biofilms. J Periodontal Res 2011 Apr;46(2):228-234. [doi: 10.1111/j.1600-0765.2010.01334.x] [Medline: 21198645]
34. Lee J, Choi D, Mah Y, Pang E. Validity assessment of quantitative light-induced fluorescence-digital (QLF-D) for the dental plaque scoring system: a cross-sectional study. BMC Oral Health 2018 Nov 20;18(1):187 [FREE Full text] [doi: 10.1186/s12903-018-0654-8] [Medline: 30458753]
35. Han S, Kim B, Ko H, Kwon H, Kim B. Validity and reliability of autofluorescence-based quantification method of dental plaque. Photodiagnosis Photodyn Ther 2015 Dec;12(4):587-591. [doi: 10.1016/j.pdpdt.2015.10.003] [Medline: 26474950]
36. Pretty IA, Edgar WM, Smith PW, Higham SM. Quantification of dental plaque in the research environment. J Dent 2005 Mar;33(3):193-207. [doi: 10.1016/j.jdent.2004.10.017] [Medline: 15725520]
37. Cass S, Bulusu P. Interactive: The top programming languages 2018. IEEE Spectrum. 2018. URL: https://spectrum.ieee.org/static/interactive-the-top-programming-languages-2018 [accessed 2020-03-05] [WebCite Cache ID 76J2flju1]
38. Spring. URL: http://spring.io/ [accessed 2020-03-05] [WebCite Cache ID 6LnpGE2En]
39. Hibernate. URL: http://hibernate.org/ [accessed 2020-03-05] [WebCite Cache ID 6NCWTqzvP]
40. PostgreSQL. URL: https://www.postgresql.org/ [accessed 2020-03-05] [WebCite Cache ID 5EOZqekgn]
41. OpenCV. URL: http://opencv.org/ [accessed 2020-03-05] [WebCite Cache ID 6GCEFB56n]
42. JavaCV: Java interface to OpenCV, FFmpeg, and more. Audet S. URL: https://github.com/bytedeco/javacv [accessed 2020-03-05] [WebCite Cache ID 6hZyxW85u]
43. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
44. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
45. Vila-Blanco N, Freire V, Balsa-Castro C, Tomás I, Carreira MJ. DenTIUS Lab. 2019. URL: http://tec.citius.usc.es/dentiuslab/login [accessed 2020-08-17]
46. Manji F, Dahlen G, Fejerskov O. Caries and Periodontitis: Contesting the Conventional Wisdom on Their Aetiology. Caries Res 2018;52(6):548-564 [FREE Full text] [doi: 10.1159/000488948] [Medline: 29694978]
47. Smith RN, Brook AH, Elcock C. The quantification of dental plaque using an image analysis system: reliability and validation. J Clin Periodontol 2001 Dec;28(12):1158-1162. [doi: 10.1034/j.1600-051x.2001.281211.x] [Medline: 11737514]
48. Smith RN, Rawlinson A, Lath D, Elcock C, Walsh TF, Brook AH. Quantification of dental plaque on lingual tooth surfaces using image analysis: reliability and validation. J Clin Periodontol 2004 Jul;31(7):569-573. [doi: 10.1111/j.1600-051X.2004.00523.x] [Medline: 15191594]
49. Regueira-Iglesias A, Blanco T, Freire V, Balsa-Castro C, Carreira MJ, Tomás I. High validity of dental plaque indices by imaging analysis. J Dent Res; 97(2941). 2018. URL: https://iadr.abstractarchives.com/abstract/18iags-2960545/high-validity-of-dental-plaque-indices-by-imaging-analysis [WebCite Cache ID 6y64ouws6]
50. Varas-Quintana P, Aneiros-Ardao A, Balsa-Castro C, Vila-Blanco N, Carreira MJ, Tomás I. In situ evolution pattern of dental plaque through plaque indices by imaging analysis. J Dent Res; 98 (0029). 2019. URL: https://tinyurl.com/yy8epdmt [WebCite Cache ID 6y64ouws6]

Abbreviations

API: area per intensity
GUI: graphical user interface
QLF-D: quantitative light - induced fluorescence digital

Edited by G Eysenbach; submitted 25.03.20; peer-reviewed by PH Cheng, E Muteba Ayumba; comments to author 15.06.20; revised version received 22.06.20; accepted 25.06.20; published 03.09.20.

Please cite as:
Vila-Blanco N, Freire V, Balsa-Castro C, Tomás I, Carreira MJ
DenTiUS Plaque, a Web-Based Application for the Quantification of Bacterial Plaque: Development and Usability Study
J Med Internet Res 2020;22(9):e18570
URL: https://www.jmir.org/2020/9/e18570
doi:10.2196/18570
PMID:32663148

Original Paper

Health Technology Readiness Profiles Among Danish Individuals With Type 2 Diabetes: Cross-Sectional Study

Ida Kær Thorsen[1], MSc; Sine Rossen[2], MSc, PhD; Charlotte Glümer[3], MD, PhD; Julie Midtgaard[4,5], MSc, PhD; Mathias Ried-Larsen[1], MSc, PhD; Lars Kayser[5], MD, PhD

[1]The Centre for Physical Activity Research, University of Copenhagen, Copenhagen, Denmark

[2]Copenhagen Centre for Cancer and Health, Municipality of Copenhagen, Copenhagen, Denmark

[3]Center for Diabetes, Municipality of Copenhagen, Copenhagen, Denmark

[4]The University Hospitals' Centre for Health Research, Rigshospitalet, Copenhagen, Denmark

[5]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Ida Kær Thorsen, MSc
The Centre for Physical Activity Research
University of Copenhagen
Rigshospitalet 7641
Blegdamsvej 9
Copenhagen, DK-2100
Denmark
Phone: 45 35 45 76 41
Fax: 35 45 76 44
Email: ida.kaer.thorsen@regionh.dk

Abstract

Background: Information technologies (IT) are increasingly implemented in type 2 diabetes (T2D) treatment as a resource for remotely supported health care. However, possible pitfalls of introducing IT in health care are generally overlooked. Specifically, the effectiveness of IT to improve health care may depend on the user's readiness for health technology.

Objective: We aim to investigate readiness for health technology in relation to mental well-being, sociodemographic, and disease-related characteristics among individuals with T2D.

Methods: Individuals with T2D (aged ≥18 years) who had been referred to self-management education, exercise, diet counseling, smoking cessation, or alcohol counseling completed a questionnaire survey covering (1) background information, (2) the 5-item World Health Organization Well-Being Index (WHO-5), (3) receptiveness to IT use in physical activity, and (4) the Readiness and Enablement Index for Health Technology (READHY), constituted by dimensions related to self-management, social support, and eHealth literacy. Individuals were divided into profiles using cluster analysis based on their READHY scores. Outcomes included differences across profiles in mental well-being, sociodemographic, and disease-related characteristics.

Results: Participants in the study were 155 individuals with T2D with a mean age of 60.2 (SD 10.7) years, 55.5% (86/155) of which were men and 44.5% (69/155) of which were women. Participants were stratified into 5 health technology readiness profiles based on the cluster analysis: Profile 1, high health technology readiness; Profile 2, medium health technology readiness; Profile 3, medium health technology readiness and high level of emotional distress; Profile 4, medium health technology readiness and low-to-medium eHealth literacy; Profile 5, low health technology readiness. No differences in sociodemographic and disease-related characteristics were observed across profiles; however, we identified 3 vulnerable subgroups of individuals: Profile 3 (21/155, 13.5%), younger individuals (mean age of 53.4 years, SD 8.9 years) with low mental well-being (mean 42.7, SD 14.7) and emotional distress (mean 1.69, SD 0.38); Profile 4 (20/155, 12.9%), older individuals (mean age 66.3 years, SD 9.0 years) with less IT use (50.0% used IT for communication) and low-to-medium eHealth literacy; and Profile 5 (36/155, 23.2%) with low mental well-being (mean 43.4, SD 20.1) and low readiness for health technology.

Conclusions: Implementation of IT in health care of individuals with T2D should be based on comprehensive consideration of mental well-being, emotional distress, and readiness for health technology rather than sociodemographic and disease-related characteristics to identify the individuals in need of social support, self-management education, and extensive IT support. A

one-size-fits-all approach to IT implementation in health care will potentially increase the risk of treatment failure among the most vulnerable individuals.

(J Med Internet Res 2020;22(9):e21195) doi:10.2196/21195

KEYWORDS

readiness for health technology; telemedicine; diabetes mellitus, type 2; socioeconomic factors; mental health; psychological distress; healthcare disparities; delivery of healthcare; exercise

Introduction

Information technologies (IT) are increasingly implemented in type 2 diabetes (T2D) treatment throughout the health care system [1-3]. This stems from a common consensus that the implementation of IT, such as telehealth, health apps, social media, and the use of computers, smartphones, smartwatches, and tablets, has great potential for improving health care and self-management. Self-management is a crucial and ubiquitous element of T2D treatment [1]. In this context, digitalization is expected to promote the individual's engagement in their own disease and health [2,3]. It is commonly held that digitalization strengthens the health professional–patient relationship, provides remote support, and increases time- and cost-efficiency [1-3]. Moreover, digitalization may facilitate the increasing person-centered and person-driven approach of health care, placing individuals in control of their own disease and treatment [2]. The use of IT is increasing among older populations (such as those who are ≥ 50 years), providing an easily accessible resource for remotely supported health care of individuals with T2D [4].

However, the possible pitfalls of introducing IT in health care have received little attention. Digitally supported weight loss or physical activity (PA) interventions have shown unexpected negative or lacking effects [5,6]. The effectiveness of IT implementation to improve health care and facilitate lifestyle change greatly depends on the user's competencies, motivation, and experience with IT solutions [7]. In addition, the phenomenon known as the digital divide may further affect the potential of IT implementation to improve health care universally [8]. As such, individuals with T2D with increased age, low education levels, and of certain ethnic minority groups are more likely to lack access to IT solutions [8]. Low health literacy is common among these individuals and is independently associated with poor glycemic control [9]. Poor glycemic control and low education levels are further associated with the increased occurrence of depression symptoms, indicating a negative influence on mental well-being [10]. Additionally, low socioeconomic status is associated with nonattainment of T2D treatment goals [11] and is a strong predictor of all-cause and cardiovascular mortality [12]. Altogether, this gap in health care related to sociodemographic characteristics is at risk of widening if IT solutions are introduced in the health care of individuals with T2D without considering the individual´s readiness for health technology.

Readiness for health technology, including the user's knowledge, skills, and attitudes towards health technology, their self-management of disease, and their social context, can be captured by the Readiness and Enablement Index for Health Technology (READHY) [13]. The READHY tool is based on a new understanding of eHealth literacy [14]. The eHealth Literacy Questionnaire (eHLQ) [15] constitutes the core of the tool, capturing the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem [15]. To further capture aspects of self-management and social support, this is supplemented by scales from 2 other validated tools [16,17].

The aims of the study are (1) to identify health technology readiness profiles among individuals with T2D using the READHY tool; (2) to investigate the differences between these profiles according to sociodemographic and disease-related characteristics, mental well-being, lifestyle factors, and IT use; and (3) to investigate the association of receptiveness to IT use in PA to sociodemographic and disease-related characteristics, mental well-being, smoking habits, and alcohol consumption.

Methods

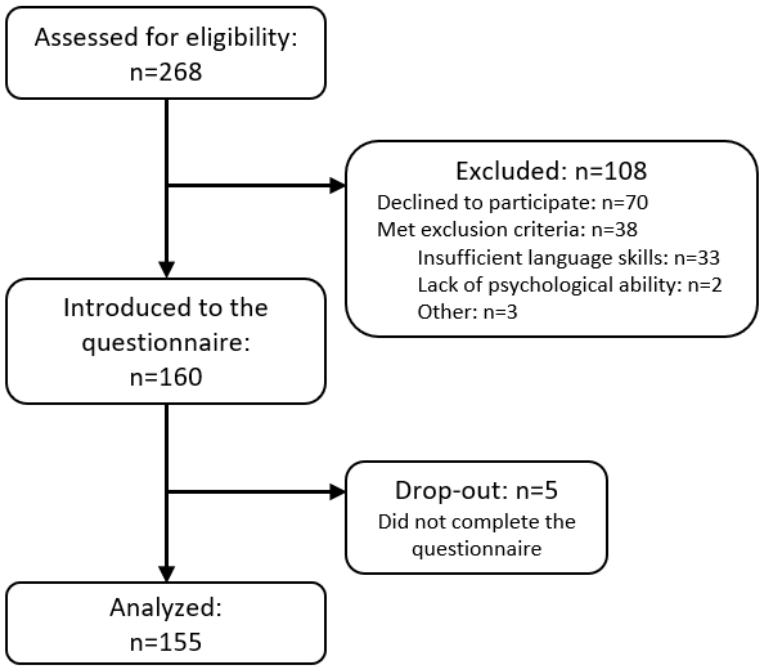
Study Design, Setting, and Participants

This is a cross-sectional study conducted as a questionnaire survey, including background information and 3 instruments: (1) the 5-item World Health Organization Well-Being Index (WHO-5) [18], (2) receptiveness to IT use in PA [19], and (3) the READHY tool [13]. The questionnaire was administered on-site using paper and pencil and was partly interviewer- and self-administered with the possibility of receiving assistance.

Participants for the study were recruited directly from the Center for Diabetes, Municipality of Copenhagen in Denmark, which provides lifestyle programs for individuals with T2D such as self-management education, exercise, diet counseling, smoking cessation, or alcohol counseling. The participant flow is depicted in Figure 1. All the individuals (N=268) who had an appointment at the center during the time period of February 5, 2018, to March 28, 2018, were approached by a project staff member. Eligibility for participation was based on the inclusion criteria of being ≥18 years of age and having a T2D diagnosis, and exclusion criteria of insufficient Danish language skills and a lack of a psychological ability to answer questions as per the evaluation of a health care professional or project staff member. A total of 155 individuals were included in the survey, resulting in a response rate of 57.8%. The individuals that declined or met exclusion criteria had a mean age of 58.8 (SD 12.0) years, 56.6% (64/113) of which were men and 43.4% (49/113) of which were women. According to these parameters, they did not differ markedly from the included individuals. Participants provided oral and written consent prior to participation. The

ethical committee of the Capital Region of Denmark confirmed that ethical approval was not required (18012824).

Figure 1. Flow of participants through the study (N=268).



Measures

Sociodemographic and Disease-Related Characteristics, Lifestyle Factors, and IT Use

Background information on sociodemographic and disease-related characteristics, lifestyle factors, and IT use were collected via self-report and include sex, age, education level, cohabitation status, source of income, time since diagnosis, type of medication, T2D complications, additional chronic conditions, smoking habits, alcohol consumption, daily PA, participation in lifestyle programs, and ownership and purpose of IT use. Education level was categorized according to the International Standard Classification of Education 2011 (ISCED-2011) [20] as follows: comprehensive school (ISCED-2011 levels 1–2), short education (ISCED-2011 levels 3–5), medium education (ISCED-2011 level 6), and long education (ISCED-2011 levels 7–8). Alcohol consumption was evaluated according to recommendations for men and women, respectively [21].

Mental Well-Being

Mental well-being was assessed using the WHO-5 [18]. Risk of depression was defined as scores of <50 [18].

The Readiness and Enablement Index for Health Technology (READHY)

The READHY tool was used to assess readiness for health technology. The READHY tool consists of the eHealth Literacy Questionnaire (eHLQ) [15] that includes 7 scales, supplemented with 4 scales from the Health Education Impact Questionnaire (heiQ) [17] and 2 scales from the Health Literacy Questionnaire (HLQ) [16]. Together, these scales capture eHealth literacy, self-management, and social context. The READHY tool has been validated [13]. The 13 scales were assessed using 65 items, each item presented to the participant as a statement and scored on a 4-point rating, from 1=strongly disagree to 4=strongly agree. The overall score of each scale was calculated as the mean score of the 4–6 items (ie, statements) that constitute the scale. If <50% of items in a scale were answered, the scale was regarded as missing and the questionnaire survey was considered incomplete for the respondent.

Receptiveness to IT Use in Physical Activity

Receptiveness to IT use in PA was categorized into 2 groups, receptive and nonreceptive, according to self-reported answers (“yes” or “no,” respectively) to the question, “Can you imagine supplementing exercise with the use of IT solutions?”

Statistical Methods

The participants for this study constituted a convenience sample obtained during the period of February 5, 2018, to March 28, 2018.

Using a data driven approach, a combination of hierarchical and K-means cluster analysis was applied to the READHY data to divide participants into clusters (hereafter referred to as profiles) according to their level of readiness for health technology. Hierarchical cluster analysis with Ward’s method for linkage (L2 squared measure) was used to determine the optimal number of profiles. Evaluation of the dendrogram, elbow method, and silhouette coefficients assessed 4 profiles as the best fit and 5 profiles as the second-best fit.

Thereafter, K-means cluster analysis was conducted with both the 4- and 5-profile solution in 8 iterations. For both solutions, the difference between profiles for each READHY scale was assessed using a one-way analysis of variance (ANOVA). The magnitude of the F value indicates how well the respective scale discriminates between profiles. Pairwise comparisons were

performed if indicative by the one-way ANOVA (P<.05). The READHY scale scores for each of the profiles are reported as a mean and standard deviation (SD). For both the 4- and 5-profile solution, differences between the identified profiles in sociodemographic and disease-related characteristics, mental well-being, lifestyle factors, and IT use were tested using the Fisher exact test for frequencies and using the one-way ANOVA for continuous variables. Pairwise comparisons were performed if indicative by the Fisher exact test or one-way ANOVA (P<.05). Frequencies are reported as numbers and proportions, and continuous variables are reported as a mean and SD. We found that the 5-profile solution contained the information from the 4-profile solution and further added information to the analyses. Therefore, we chose to only report the results from the 5-profile solution. The correlation between mental well-being and heiQ8-emotional distress was tested using the Pearson's correlation to evaluate the association between these 2 parameters.

The association of sociodemographic and disease-related characteristics, mental well-being, smoking habits, and alcohol consumption to the receptiveness to IT use in PA were tested using logistic regression. Receptiveness to IT use in PA was defined as the binary outcome variable. All exposure variables were included in the model individually. For each of the exposure variables, the odds ratio and a 95% confidence interval for receptiveness to IT use in PA are reported. P values and 95% confidence intervals were calculated using exact statistics.

Model assumptions were investigated prior to analyses by investigating predicted values and standardized residuals. Data were analyzed as observed—no imputations were used to replace missing data. All statistical analyses were performed using Stata IC 13 (StataCorp). The significance level was set to P<.05 (2-tailed).

Role of the Funding Source

The funding source was not involved in the study design, in the collection, analysis, or interpretation of data, in writing the report; or in the decision to submit the paper for publication. IKT had full access to all the data in the study. All authors had final responsibility for the decision to submit for publication.

Results

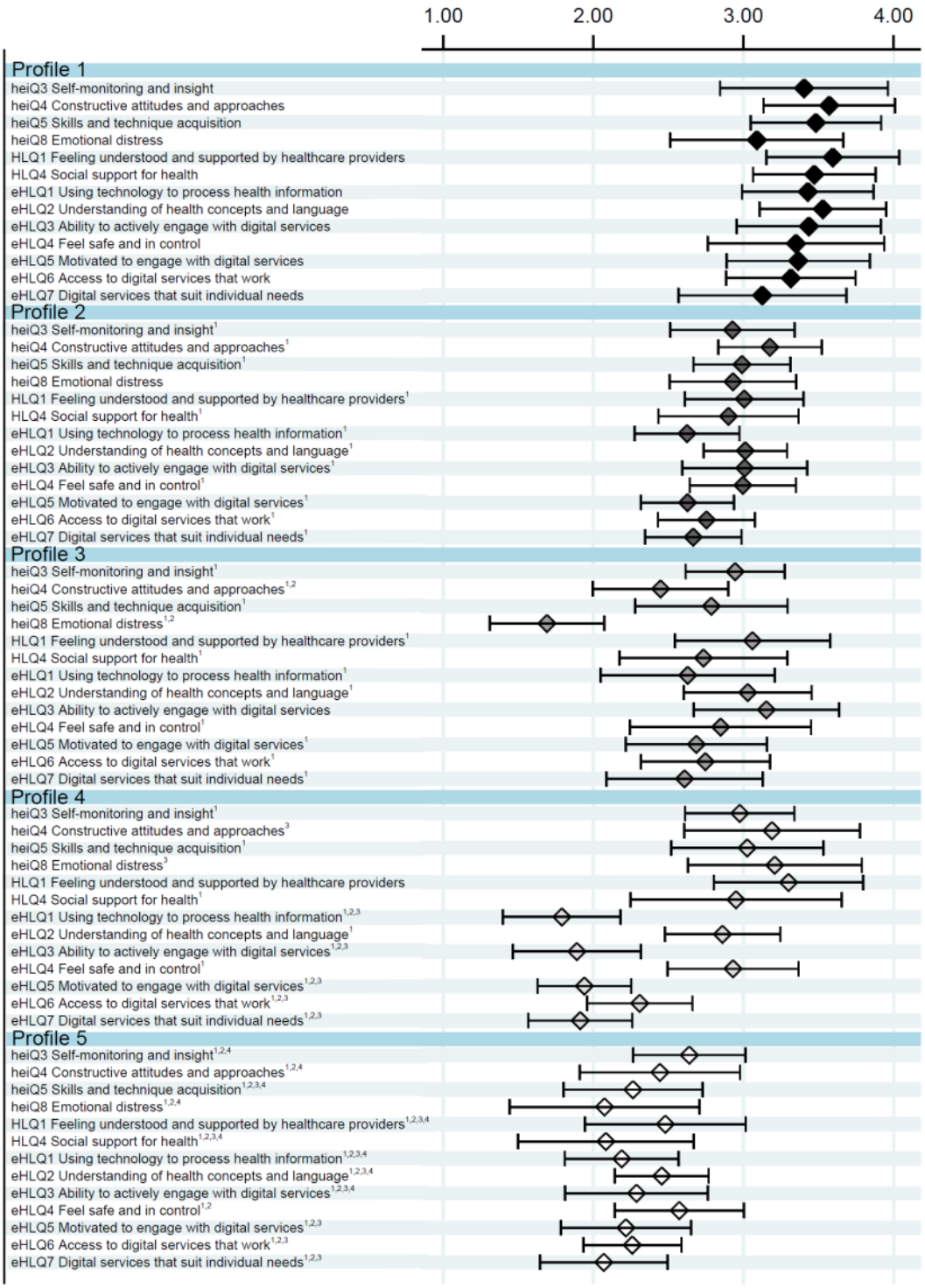
Participant Characteristics

Participants had a mean age of 60.2 (SD 10.7) years; 55.5% (86/155) were men and 44.5% (69/155) were women. Most participants were diagnosed less than 5 years ago (124/155, 80.0%), were prescribed peroral medication (102/155, 65.8%), and experienced T2D complications (90/122, 73.8%). Moreover, one-third of the participants were at risk of depression (49/154, 31.8%), and most participants had 2 or more additional chronic conditions (78/155, 50.3%), owned a smartphone (111/155, 71.6%), and reported a wish to be more physically active (125/155, 80.7%).

Readiness for Health Technology

The combined data-driven cluster analyses resulted in 5 distinct profiles within which participants were similar regarding readiness for health technology, with a large variability between profiles and a small variability within profiles (Figure 2). Profiles 1 to 5 are presented in ascending order according to their overall mean READHY score. Of the 5 profiles, 3 profiles consistently scored high, medium, and low, respectively, whereas 2 profiles had varying scores on some scales. Profile 1 (n=28) consistently scored high on all scales. Profile 2 (n=50) consistently scored medium on all scales. Profile 3 (n=21) scored medium on all scales, except for low scores on heiQ4-constructive attitudes and approaches and heiQ8-emotional distress. Profile 4 (n=20) generally scored high on scales related to self-management and social support and low-to-medium on scales related to eHealth literacy; Profile 5 (n=36) consistently scored low on all scales. Overall between-group differences were observed across profiles (P=.001). Pairwise comparisons of each scale are presented in Figure 2.

Figure 2. Readiness and Enablement Index for Health Technology (READHY) scale scores for the 5 identified profiles based on cluster analysis. heiQ: Health Education Impact Questionnaire; HLQ: Health Literacy Questionnaire; eHLQ: eHealth Literacy Questionnaire. Data are presented as mean (SD). heiQ8 was reversed (ie, a high score indicates a low level of emotional distress). heiQ3, F=13.16; heiQ4, F=34.38; heiQ5, F=33.81; heiQ8, F=42.87; HLQ1, F=24.24; HLQ4, F=28.12; eHLQ1, F=55.49; eHLQ2, F=37.90; eHLQ3, F=51.19; eHLQ4, F=11.31; eHLQ5, F=48.17; eHLQ6, F=39.12; eHLQ7, F=35.52.



Sociodemographic Characteristics

Sociodemographic characteristics are presented in Table 1. Participants in Profile 3 were younger than participants in Profiles 1 and 4 (-9.0 years, *P*=.03; and -12.8 years, *P*=.001, respectively). More participants in Profile 3 received public income support or had no income compared with participants in Profiles 1, 2, and 4 (+39.2 percent points (pp), *P*=.02, +45.1 pp, *P*=.001, and +47.1 pp, *P*=.004, respectively) of which more received retirement pension (+25.0 pp, *P*=.02, +23.4 pp, *P*=.001, and +36.4 pp, *P*=.004, respectively). More participants in Profile 2 received salary compared with participants in Profile 3 (+21.7

pp, *P*=.001). Moreover, more participants in Profile 5 received public income support or had no income compared with participants in Profiles 2 and 4 (+21.3 pp, *P*=.02 and +23.3 pp, *P*=.04, respectively) of which more received retirement pension (+15.9 pp *P*=.02, and +28.9 pp, *P*=.04, respectively). There were no differences across profiles in sex, highest level of education, or cohabitation status (*P*>.05).

Table 1. Sociodemographic characteristics of participants (N=155) across profiles [data are presented as mean (SD) for continuous variables and numbers (proportions) for frequencies].

Characteristics	All (N=155)	Profile 1 (n=28, 18.1%)	Profile 2 (n=50, 32.2%)	Profile 3 (n=21, 13.5%)	Profile 4 (n=20, 12.9%)	Profile 5 (n=36, 23.2%)	*P* value
Sex, n (%)							.46
Women	69 (44.5)	13 (46.4)	21 (42.0)	9 (42.9)	6 (30.0)	20 (55.6)	
Men	86 (55.5)	15 (53.6)	29 (58.0)	12 (57.1)	14 (70.0)	16 (44.4)	
Age, mean (SD)	60.2 (10.7)	62.4 (10.6)	59.8 (11.3)	53.4 (8.9)[a]	66.3 (9.0)[b]	59.8 (9.6)	.002
Highest attained level of education, n (%)							.15
Comprehensive school	24 (15.5)	2 (7.2)	7 (14.0)	7 (33.3)	1 (5.0)	7 (19.4)	
Short education	81 (52.2)	13 (46.4)	28 (56.0)	8 (38.1)	16 (80.0)	16 (44.5)	
Medium education	33 (21.3)	10 (35.7)	9 (18.0)	4 (19.1)	1 (5.0)	9 (25.0)	
Long education	17 (11.0)	3 (10.7)	6 (12.0)	2 (9.5)	2 (10.0)	4 (11.1)	
Cohabitation status, n (%)							.69
Living alone	78 (50.3)	13 (46.4)	22 (44.0)	12 (57.1)	10 (50.0)	21 (58.3)	
Living with spouse and/or children	77 (49.7)	15 (53.6)	28 (56.0)	9 (42.9)	10 (50.0)	15 (41.7)	
Source of income, n (%)							.007
Salary	45 (29.0)	8 (28.6)	18 (36.0)	3 (14.3)[c]	5 (25.0)	11 (50.6)	
Retirement pension	73 (47.1)	15 (53.6)	26 (52.0)	6 (28.6)[a,c]	13 (65.0)[b]	13 (36.1)[c,d]	
Public income support/no incomes	37 (23.9)	5 (17.8)	6 (12.0)	12 (57.1)[a,c]	2 (10.0)[b]	12 (33.3)[c,d]	

[a]Different from Profile 1, *P*<.05.

[b]Different from Profile 3, *P*<.05.

[c]Different from Profile 2, *P*<.05.

[d]Different from Profile 4, *P*<.05.

Disease-Related Characteristics and Mental Well-being

Disease-related characteristics and mental well-being are presented in Table 2. Participants in Profiles 3 and 5 scored lower on mental well-being compared to participants in Profiles 1, 2, and 4 (Profile 3, -29.0, -24.6, and -24.5, respectively, *P*=.001; Profile 5, -23.8, -23.8, and -23.8, respectively, *P*=.001). Further, more participants in Profile 3 and 5 were at risk of depression compared with participants in Profiles 1, 2, and 4 (Profile 3, +56.0 pp, +54.5 pp, and +56.7 pp, respectively, *P*=.001; Profile 5, +56.0 pp, +54.5 pp, and +56.7 pp, respectively, *P*=.001). The correlation between mental well-being and heiQ8-emotional distress was strong (r=0.645, *P*=.001). There were no differences across profiles in time since diagnosis, type of medication, T2D complications, and additional chronic conditions (*P*>.05).

Table 2. Disease-related characteristics and mental well-being of participants (N=155) across profiles [data are presented as mean (SD) for continuous variables and numbers (proportions) for frequencies].

Characteristics	All (N=155)	Profile 1 (n=28, 18.1%)	Profile 2 (n=50, 32.3%)	Profile 3 (n=21, 13.5%)	Profile 4 (n=20, 12.9%)	Profile 5 (n=36, 23.2%)	*P* value
Time since diabetes diagnosis, n (%)							.87
≤5 years	124 (80.0)	22 (78.6)	42 (82.0)	16 (76.2)	15 (75.0)	29 (80.6)	
≥6 years	31 (20.0)	6 (21.4)	8 (16.0)	5 (23.8)	5 (25.0)	7 (19.4)	
Type of medication, n (%)							.58
None	19 (12.3)	4 (14.3)	8 (16.0)	3 (14.3)	1 (5.0)	3 (8.3)	
Peroral	102 (65.8)	19 (67.9)	32 (64.0)	10 (47.6)	16 (80.0)	25 (69.5)	
Injection	34 (21.9)	5 (17.8)	10 (20.0)	8 (38.1)	3 (15.0)	8 (22.2)	
T2D complications, n (%)[a]							.37
Yes	90 (73.8)	5 (22.7)	7 (18.0)	6 (35.3)	7 (41.2)	7 (25.9)	
No	32 (26.2)	17 (77.3)	32 (82.0)	11 (64.7)	10 (58.8)	20 (74.1)	
Additional chronic conditions, n (%)							.05
No additional conditions	17 (11.0)	3 (10.7)	10 (20.0)	1 (4.8)	2 (10.0)	1 (2.8)	
1 additional condition	60 (38.7)	13 (46.4)	23 (46.0)	5 (23.8)	7 (35.0)	12 (33.3)	
2+ additional conditions	78 (50.3)	12 (42.9)	17 (34.0)	15 (71.4)	11 (55.0)	23 (63.9)	
Mental well-being, mean (SD)[b]	59.1 (20.2)	71.7 (16.3)	67.3 (14.6)	42.7 (14.7)[c,d]	67.2 (12.7)[e]	43.4 (20.1)[c,d,f]	<.001
Risk of depression (score <50), n (%)[b]	49 (31.8)	3 (10.7)	6 (12.2)	14 (66.7)[c,d]	2 (10.0) [e]	24 (66.7)[c,d,f]	.001

[a]n=122.

[b]n=154.

[c]Different from Profile 1, *P*<.05.

[d]Different from Profile 2, *P*<.05.

[e]Different from Profile 3, *P*<.05.

[f]Different from Profile 4, *P*<.05.

Lifestyle Factors

Lifestyle factors are presented in Table 3. There were no differences across profiles in smoking habits, alcohol consumption, daily PA, wish to be more physically active, change in exercise habits with T2D, or participation in lifestyle or exercise programs (*P*>.05).

Table 3. Lifestyle factors of participants (N=155) across profiles [data are presented as numbers (proportions)].

Lifestyle factors	All (N=155)	Profile 1 (n=28, 18.1%)	Profile 2 (n=50, 32.3%)	Profile 3 (n=21, 13.5%)	Profile 4 (n=20, 12.9%)	Profile 5 (n=36, 23.2%)	*P* value
Smoking habits, n (%)[a]							.82
Current	26 (16.9)	2 (7.1)	8 (16.3)	6 (28.6)	3 (15.0)	7 (19.4)	
Earlier	76 (49.3)	16 (57.2)	23 (47.0)	9 (42.8)	10 (50.0)	18 (50.0)	
Never	52 (33.8)	10 (35.7)	18 (36.7)	6 (28.6)	7 (35.0)	11 (30.6)	
Alcohol consumption, n (%)[a]							.08
No alcohol	55 (35.7)	9 (32.1)	11 (22.4)	11 (52.4)	5 (25.0)	19 (52.8)	
According to recommendations	85 (55.2)	16 (57.2)	34 (69.4)	9 (42.8)	13 (65.0)	13 (36.1)	
Above recommendations	14 (9.1)	3 (10.7)	4 (8.2)	1 (4.8)	2 (10.0)	4 (11.1)	
Daily physical activity, n (%)[a]							.77
<30 min/day	38 (24.7)	5 (17.9)	14 (28.6)	3 (14.3)	6 (30.0)	10 (27.8)	
30-60 min/day	69 (44.8)	14 (50.0)	19 (38.8)	13 (61.9)	7 (35.0)	16 (44.4)	
>60 min/day	47 (30.5)	9 (32.1)	16 (32.6)	5 (23.8)	7 (35.0)	10 (27.8)	
Wish to be more physically active, n (%)							.42
Yes	125 (80.7)	23 (82.2)	37 (74.0)	21 (100.0)	16 (80.0)	28 (77.8)	
No	12 (7.7)	2 (7.1)	6 (12.0)	0 (0.0)	2 (10.0)	2 (5.5)	
Maybe	18 (11.6)	3 (10.7)	7 (14.0)	0 (0.0)	2 (10.0)	6 (16.7)	
Change in exercise habits with T2D, n (%)							.14
Increased	75 (48.4)	17 (60.7)	25 (50.0)	12 (57.1)	7 (35.0)	14 (38.9)	
Unchanged	73 (47.1)	10 (35.7)	24 (48.0)	6 (28.6)	12 (60.0)	21 (58.3)	
Decreased	7 (4.5)	1 (3.6)	1 (2.0)	3 (14.3)	1 (5.0)	1 (2.8)	
Lifestyle intervention, n (%)							.28
No lifestyle courses	41 (26.4)	10 (35.7)	12 (24.0)	6 (28.6)	2 (10.0)	11 (30.6)	
1 lifestyle course	35 (22.6)	3 (10.7)	10 (20.0)	5 (23.8)	9 (45.0)	8 (22.2)	
2+ lifestyle courses	79 (51.0)	15 (53.6)	28 (56.0)	10 (47.6)	9 (45.0)	17 (47.2)	
Exercise intervention, n (%)							.94
Yes	73 (47.1)	13 (46.4)	25 (50.0)	9 (42.9)	8 (40.0)	18 (50.0)	
No	82 (52.9)	15 (53.6)	25 (50.0)	12 (57.1)	12 (60.0)	18 (50.0)	

[a]n=154

IT Use

Factors related to IT use are presented in Table 4. Fewer participants in Profile 4 owned a smartphone compared to participants in Profiles 1 and 2 (-33.6 pp, *P*=.03; and -37.0 pp, *P*=.003, respectively). Compared with participants in Profiles 1, 2, and 3, fewer participants in Profile 4 used IT for information seeking (-27.9 pp, *P*=.02, -29.0 pp, *P*=.004, and -30.2 pp, *P*=.02, respectively), communication with family and friends (-42.9 pp, *P*=.002, -42.0 pp, *P*=.001, and -45.2 pp, *P*=.001, respectively), and entertainment (-39.3 pp, *P*=.004, -26.0 pp, *P*=.05, and -45.2 pp, *P*=.001, respectively). Moreover, fewer participants in Profile 5 used IT for entertainment compared to participants in Profiles 1 and 3 (-23.6 pp, *P*=.04, and -29.5 pp, *P*=.02, respectively). Finally, fewer participants in Profile 4 were receptive to IT use in PA compared to participants in Profiles 1, 2, and 3 (-38.6 pp, *P*=.01, -34.0 pp, *P*=.01, and -41.0 pp, *P*=.01, respectively). There were no differences across profiles in smartwatch, tablet, and computer ownership, or exercise and work purposes of IT use (*P*>.05).

Table 4. Information technologies (IT) use of participants (N=155) across profiles [data are presented as numbers (proportions)].

IT use	All (N=155)	Profile 1 (n=28, 18.1%)	Profile 2 (n=50, 32.3%)	Profile 3 (n=21, 13.5%)	Profile 4 (n=20, 12.9%)	Profile 5 (n=36, 23.2%)	*P* value
Owns smartphone, n (%)							.03
Yes	111 (71.6)	22 (78.6)	41 (82.0)	16 (76.2)	9 (45.0)[a,b]	23 (63.9)	
No	44 (28.4)	6 (21.4)	9 (18.0)	5 (23.8)	11 (55.0)[a,b]	13 (36.1)	
Owns smartwatch, n (%)							>.99
Yes	8 (5.2)	1 (3.6)	3 (6.0)	1 (4.8)	1 (5.0)	2 (5.6)	
No	147 (94.8)	27 (96.4)	47 (94.0)	20 (95.2)	19 (95.0)	34 (94.4)	
Owns tablet, n (%)							.12
Yes	82 (52.9)	9 (32.1)	30 (60.0)	10 (47.6)	7 (35.0)	16 (44.4)	
No	73 (47.1)	19 (67.9)	20 (40.0)	11 (52.4)	13 (65.0)	20 (55.6)	
Owns computer (NOT smartphone, smartwatch, or tablet), n (%)							.08
Yes	24 (15.5)	2 (7.1)	5 (10.0)	3 (14.3)	7 (35.0)	7 (19.4)	
No	131 (84.5)	26 (92.9)	45 (90.0)	18 (85.7)	13 (65.0)	29 (80.6)	
Purpose of using IT, n (%)[c]							
Exercise	27 (17.5)	7 (25.0)	9 (18.0)	2 (9.5)	2 (10.0)	7 (20.0)	.61
Work	55 (35.7)	10 (35.7)	21 (42.0)	5 (23.8)	6 (30.0)	13 (37.1)	.67
Information seeking	135 (87.7)	26 (92.9)	47 (94.0)	20 (95.2)	13 (65.0)[a,b,d]	29 (82.9)	.01
Communication with family/friends	129 (83.8)	26 (92.9)	46 (92.0)	20 (95.2)	10 (50.0)[a,b,d]	27 (77.1)	.001
Entertainment	116 (75.3)	25 (89.3)	38 (76.0)	20 (95.2)	10 (50.0)[a,b,d]	23 (65.7)[a,d]	.003
Receptiveness to IT use in PA, n (%)							.03
Receptive	107 (69.0)	22 (78.6)	37 (74.0)	17 (81.0)	8 (40.0)[a,b,d]	23 (63.9)	
Nonreceptive	48 (31.0)	6 (21.4)	13 (26.0)	4 (19.0)	12 (60.0)[a,b,d]	13 (36.1)	

[a]Different from Profile 1, *P*<.05.

[b]Different from Profile 2, *P*<.05.

[c]n=154.

[d]Different from Profile 3, *P*<.05.

Receptiveness to IT Use in Physical Activity

Of the 155 participants, a total of 107 (69.0%) responded that they could imagine supplementing exercise with the use of IT solutions. Sociodemographic and disease-related characteristics, mental well-being, smoking habits, and alcohol consumption are presented in Multimedia Appendix 1 for participants that were receptive and nonreceptive to IT use in PA, respectively. Increasing age decreased the odds of being receptive (OR=0.94, 95% CI 0.90-0.97; *P*=.001). There were no significant associations between receptive and nonreceptive participants regarding the remaining exposure variables. Nonreceptive participants scored lower on eHLQ1-using technology to process health information (*P*=.001), eHLQ3-ability to actively engage with digital services (*P*=.001), and eHLQ5-motivated to engage with digital services (*P*=.001) compared to receptive participants (Multimedia Appendix 2).

Discussion

The main finding of this study is the identification of vulnerable subgroups of individuals with T2D characterized by low mental well-being, emotional distress, and low readiness for health technology. Notably, the findings indicate that these vulnerable subgroups could not be identified by their disease-related and sociodemographic characteristics, including ambiguous findings according to age. Thus, it is crucial that IT-supported T2D health care is individually tailored based on an evaluation of mental well-being, emotional distress, and readiness for health technology rather than sociodemographic characteristics, including age and the severity of T2D.

This is the first study to add measures of self-management and social support to eHealth literacy in a profound understanding of readiness for health technology among individuals with T2D referred to a lifestyle program. The stratification of individuals into profiles based on their level of readiness for health technology is supported by previous findings among individuals

with cancer referred to a rehabilitation program [19]. In this study, we identified a subgroup (Profile 3) of relatively younger individuals outside the labor market with a particularly high level of emotional distress, which we do not see among cancer survivors. In the context of diabetes, emotional distress has previously been described as diabetes distress [10]. Diabetes distress may affect up to 45% of individuals with T2D, of which 70% do not meet the criteria for major depressive disorder (MDD) [10]. In line with MDD, individuals with diabetes distress are less likely to engage in self-managing behaviors, which negatively affects health outcomes (for example, leading to poor glycemic control) [10]. The medium level of eHealth literacy along with the high receptiveness to IT use in PA in this subgroup indicates that under the right conditions, IT implementation may be a beneficial alternative or supplement to center-based exercise programs. We report a strong correlation between low mental well-being (an indicator of risk of depression [18]) and emotional distress (an indicator of diabetes distress [10]). Individuals with MDD may benefit from pharmacotherapy, whereas individuals with diabetes distress are not likely to [10]. Diabetes self-management education is an effective treatment for diabetes distress [10]. Therefore, by distinguishing between depression and diabetes distress, self-management education could be a specific focus when implementing IT in the health care of this subgroup.

We identified 2 additional vulnerable subgroups. One of these (Profile 5) was characterized by low mental well-being and low readiness for health technology. Self-reported diagnosis of depression has previously been negatively associated with eHealth literacy [22]. This indicates that if IT is implemented in health care, it should include extensive support, covering social, self-management, and IT-related aspects. The other subgroup (Profile 4) was characterized by older age, higher mental well-being, as well as a lower level of IT ownership, use, and receptiveness. This subgroup had low eHealth literacy and a medium level of self-management and social support. Rossen and colleagues [19] suggest that IT implementation among such subgroups should be based on a dialogue with the individual about the potential benefits of using IT along with a thorough introduction and IT support. For some individuals in both of these subgroups, this support may not be sufficient to prevent treatment failure, and IT support should be implemented with caution.

In contrast to individuals with cancer and a previous South Korean study in T2D, individuals with low versus high readiness for health technology in this study were not characterized by a sociodemographic gradient [19,23]. The South Korean study used a Korean version of the eHealth Literacy Scale (eHEALS) tool [24], which assesses individuals' combined knowledge, comfort, and perceived skills at finding, evaluating, and applying electronic health information to health problems [24], while in our study, we include measures of self-management and social support in a new understanding of eHealth literacy. This variation in the understanding of eHealth literacy potentially explains the discrepancies between the studies. Moreover, discrepancies may be attributed, in part, to cultural differences in eHealth literacy between Denmark and South Korea. Specifically, smartphone ownership and internet use among the Danish and South Korean populations is ≥86% and ≥90%, respectively, regardless of sociodemographic characteristics [4,25], indicating that differences may primarily be attributed to cultural differences in health literacy [26]. However, none of the disease-related characteristics reported by Kim and colleagues [23] were associated with eHealth literacy, which agrees with the present findings.

Previous studies investigating the association of sociodemographic and disease-related characteristics to eHealth literacy among healthy individuals and individuals with chronic conditions report rather ambiguous findings [22,27-29]. However, these studies generally agree that age is negatively associated with eHealth literacy [22,27,28]. Our findings indicate that younger age may generally be associated with higher odds of being receptive to IT use for physical activity purposes; however, previous findings do not directly support our identification of a relatively younger vulnerable subgroup (Profile 3). This indicates that a thorough assessment of mental well-being and diabetes distress among relatively younger individuals with T2D is warranted before delivering IT-supported health care.

Limitations of this study include the cross-sectional design, which precludes causal inferences regarding the effects of targeting readiness for health technology in health care and the potential mediating effects of socioeconomic status and mental well-being. Longitudinal designs should be implemented to investigate how this stratification of individuals with T2D reflects interindividual health effects of IT-supported health care. This will clarify the need for social support, self-management education, and IT support among different subgroups, and elucidate whether IT-supported health care induces changes in individual levels of readiness for health technology. Further, the sample representativeness may be suboptimal. First, referral to lifestyle programs is potentially limited among individuals with low socioeconomic status [30]. Second, at the time of data collection, the READHY tool was only validated in Danish, precluding participation from non-Danish speaking individuals, such as individuals from ethnic minority groups. As such, the most vulnerable individuals with T2D may not be fully represented in this convenience sample, indicating an underrepresentation of the identified vulnerable subgroups according to magnitude and diversity. This emphasizes the importance of comprehensively considering the need for social support, self-management education, and extensive IT support when implementing IT in health care. Moreover, as participants constituted a convenience sample, no a priori sample size calculation was performed, increasing the risk of false-negative findings (type 2 errors). Finally, with the strong correlation between mental well-being and heiQ8-emotional distress, it is not surprising that profiles with low levels of emotional distress score high on mental well-being, and vice versa. Two profiles scored similarly low on mental well-being; however, interestingly, one of these (Profile 3) was characterized by a particularly high level of emotional distress.

In this study, we identified vulnerable subgroups of individuals with T2D characterized by low mental well-being, emotional distress, and low readiness for health technology, who could not be identified by their sociodemographic and disease-related

characteristics. Based on this investigation, we suggest that implementation of IT in the health care of individuals with T2D should be based on a comprehensive consideration of mental well-being, emotional distress, and readiness for health technology to identify the individuals in need of social support, self-management education, and extensive IT support. IT solutions should possibly be tailored to accommodate these needs and should not stand alone. Overall, a one-size-fits-all approach to IT implementation in health care will potentially increase the risk of treatment failure among the most vulnerable individuals with T2D.

Acknowledgments

Cecilie Fau Brinkløv (CFAS), Trine C. Møller (Center for Diabetes), and Line Autrup (Center for Diabetes) are acknowledged for their contributions to preparing, organizing, and administering the questionnaire survey. Emilia Lüders (CFAS) and other staff members at the Center for Diabetes are acknowledged for their contribution to the data collection. The Centre for Physical Activity Research (CFAS) is supported by TrygFonden (grants ID 101390 and ID 20045). CFAS is a member of the Danish Center for Strategic Research in Type 2 Diabetes (DD2; the Danish Council for Strategic Research, grant no. 09-067009 and 09-075724). The study was further supported by TrygFonden (grant 124708).

Authors' Contributions

IKT, SR, CG, MR-L, and LK designed the trial. IKT collected the data with the staff members named in the Acknowledgments section. IKT planned and conducted the statistical analyses and wrote the first draft of the manuscript under the supervision of SR, JM, MR-L, and LK. All authors reviewed and revised the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Odds ratio for being receptive to IT use in physical activity according to sociodemographic and disease-related characteristics, mental well-being, smoking habits, and alcohol consumption.
[DOCX File , 38 KB - jmir_v22i9e21195_app1.docx]

Multimedia Appendix 2
Readiness and Enablement Index for Health Technology (READHY) scale scores for participants that are receptive vs. nonreceptive to IT use in physical activity.
[DOCX File , 32 KB - jmir_v22i9e21195_app2.docx]

References

1. American Diabetes Association. 5. Facilitating Behavior Change and Well-being to Improve Health Outcomes: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S48-S65. [doi: 10.2337/dc20-S005] [Medline: 31862748]
2. Ashrafzadeh S, Hamdy O. Patient-Driven Diabetes Care of the Future in the Technology Era. Cell Metab 2019 Mar 05;29(3):564-575 [FREE Full text] [doi: 10.1016/j.cmet.2018.09.005] [Medline: 30269984]
3. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
4. Taylor K, Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-05-06]
5. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial. JAMA 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
6. McDermott MM, Spring B, Berger JS, Treat-Jacobson D, Conte MS, Creager MA, et al. Effect of a Home-Based Exercise Intervention of Wearable Technology and Telephone Coaching on Walking Performance in Peripheral Artery Disease: The HONOR Randomized Clinical Trial. JAMA 2018 Dec 24;319(16):1665-1676 [FREE Full text] [doi: 10.1001/jama.2018.3275] [Medline: 29710165]
7. Monkman H, Kushniruk AW. The Consumer Health Information System Adoption Model. Stud Health Technol Inform 2015;218:40592. [Medline: 26262522]
8. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

9. Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, et al. Association of health literacy with diabetes outcomes. JAMA 2002;288(4):475-482. [Medline: 12132978]
10. Kreider KE. Diabetes Distress or Major Depressive Disorder? A Practical Approach to Diagnosing and Treating Psychological Comorbidities of Diabetes. Diabetes Ther 2017 Mar;8(1):1-7 [FREE Full text] [doi: 10.1007/s13300-017-0231-1] [Medline: 28160185]
11. Heltberg A, Andersen JS, Kragstrup J, Siersma V, Sandholdt H, Ellervik C. Social disparities in diabetes care: a general population study in Denmark. Scand J Prim Health Care 2017 Mar;35(1):54-63 [FREE Full text] [doi: 10.1080/02813432.2017.1288702] [Medline: 28277046]
12. Rawshani A, Svensson A, Zethelius B, Eliasson B, Rosengren A, Gudbjörnsdottir S. Association Between Socioeconomic Status and Mortality, Cardiovascular Disease, and Cancer in Patients With Type 2 Diabetes. JAMA Intern Med 2016 Aug 01;176(8):1146-1154. [doi: 10.1001/jamainternmed.2016.2940] [Medline: 27367969]
13. Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, et al. Development of the Multidimensional Readiness and Enablement Index for Health Technology (READHY) Tool to Measure Individuals' Health Technology Readiness: Initial Testing in a Cancer Rehabilitation Setting. J Med Internet Res 2019 Feb 12;21(2):e10377 [FREE Full text] [doi: 10.2196/10377] [Medline: 30747717]
14. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowledge Management & E-Learning 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
15. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A Multidimensional Tool Based on the eHealth Literacy Framework: Development and Initial Validity Testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
16. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
17. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
18. World Health Organization. Well-being measures in primary health care: the DepCare project. In: Report on a WHO Meeting.: WHO Regional Office for Europe; 1998 Presented at: 3rd Consensus Meeting; 12-13 February 1998; Stockholm, Sweden p. 1-39.
19. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. Acta Oncol 2019 May;58(5):610-618. [doi: 10.1080/0284186X.2018.1562213] [Medline: 30698060]
20. OECD, European Union, UNESCO Institute for Statistics. ISCED 2011 Operational Manual:Guidelines for classifying national education programmes and related qualifications. In: http://dx.doi.org/10.1787/9789264228368-en. Paris: OECD Publishing; 2015.
21. U.S. Department of Health and Human Services, U.S. Department of Agriculture. 2015-2020 Dietary Guidelines for Americans. health.gov. 2015 Dec. URL: https://health.gov/our-work/food-nutrition/2015-2020-dietary-guidelines/guidelines/ [accessed 2020-09-03]
22. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
23. Kim KA, Kim YJ, Choi M. Association of Electronic Health Literacy With Health-Promoting Behaviors in Patients With Type 2 Diabetes: A Cross-sectional Study. Comput Inform Nurs 2018 Sep;36(9):438-447. [doi: 10.1097/CIN.0000000000000438] [Medline: 29742548]
24. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
25. Statistics Denmark. BEBRIT09: Use of internet for private purposes - per cent of the population (16-74 years) by type and internet activity (2008-2020). StatBank Denmark. 2020. URL: https://www.statbank.dk/BEBRIT09 [accessed 2020-05-27]
26. Levin-Zamir D, Leung AYM, Dodson S, Rowlands G. Health Literacy in Selected Populations: Individuals, Families, and Communities from the International and Cultural Perspective. Stud Health Technol Inform 2017;240:392-414. [Medline: 28972530]
27. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth Literacy: Predictors in a Population With Moderate-to-High Cardiovascular Risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
28. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]

29. Milne RA, Puts MTE, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of High eHealth Literacy in Primary Lung Cancer Survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
30. Sortsø C, Lauridsen J, Emneus M, Green A, Jensen PB. Socioeconomic inequality of diabetes patients' health care utilization in Denmark. Health Econ Rev 2017 Dec;7(1):21 [FREE Full text] [doi: 10.1186/s13561-017-0155-5] [Medline: 28550486]

Abbreviations

ANOVA: analysis of variance
eHLQ: eHealth Literacy Questionnaire
heiQ: Health Education Impact Questionnaire
HLQ: Health Literacy Questionnaire
IT: information technology
MDD: major depressive disorder
PA: physical activity
pp: percent points
READHY: Readiness and Enablement Index for Health Technology
T2D: type 2 diabetes
WHO-5: 5-item World Health Organization Well-Being Index

Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by J Lauridsen, J Lander; comments to author 26.06.20; revised version received 07.07.20; accepted 14.07.20; published 15.09.20.

Please cite as:
Thorsen IK, Rossen S, Glümer C, Midtgaard J, Ried-Larsen M, Kayser L
Health Technology Readiness Profiles Among Danish Individuals With Type 2 Diabetes: Cross-Sectional Study
J Med Internet Res 2020;22(9):e21195
URL: http://www.jmir.org/2020/9/e21195/
doi:10.2196/21195
PMID:32930669

Original Paper

Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: Pooled Analysis Study Using Mixed Models

Sadiq Said[1*]; Malgorzata Gozdzik[1*]; Tadzio Raoul Roche[1]; Julia Braun[2], PhD; Julian Rössler[1]; Alexander Kaserer[1], MD; Donat R Spahn[1], MD; Christoph B Nöthiger[1], MD; David Werner Tscholl[1], MD

[1]Department of Anesthesiology, University Hospital Zurich, Zurich, Switzerland

[2]Department of Epidemiology and Biostatistics, University of Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Sadiq Said
Department of Anesthesiology
University Hospital Zurich
Raemistrasse 100
Zurich, 8091
Switzerland
Phone: 41 786718911
Email: sadiq.said@usz.ch

Abstract

Background: Patient monitoring is indispensable in any operating room to follow the patient's current health state based on measured physiological parameters. Reducing workload helps to free cognitive resources and thus influences human performance, which ultimately improves the quality of care. Among the many methods available to assess perceived workload, the National Aeronautics and Space Administration Task Load Index (NASA-TLX) provides the most widely accepted tool. However, only few studies have investigated the validity of the NASA-TLX in the health care sector.

Objective: This study aimed to validate a modified version of the raw NASA-TLX in patient monitoring tasks by investigating its correspondence with expected lower and higher workload situations and its robustness against nonworkload-related covariates. This defines criterion validity.

Methods: In this pooled analysis, we evaluated raw NASA-TLX scores collected after performing patient monitoring tasks in four different investigator-initiated, computer-based, prospective, multicenter studies. All of them were conducted in three hospitals with a high standard of care in central Europe. In these already published studies, we compared conventional patient monitoring with two newly developed situation awareness–oriented monitoring technologies called Visual Patient and Visual Clot. The participants were resident and staff anesthesia and intensive care physicians, and nurse anesthetists with completed specialization qualification. We analyzed the raw NASA-TLX scores by fitting mixed linear regression models and univariate models with different covariates.

Results: We assessed a total of 1160 raw NASA-TLX questionnaires after performing specific patient monitoring tasks. Good test performance and higher self-rated diagnostic confidence correlated significantly with lower raw NASA-TLX scores and the subscores (all $P<.001$). Staff physicians rated significantly lower workload scores than residents ($P=.001$), whereas nurse anesthetists did not show any difference in the same comparison ($P=.83$). Standardized distraction resulted in higher rated total raw NASA-TLX scores ($P<.001$) and subscores. There was no gender difference regarding perceived workload ($P=.26$). The new visualization technologies Visual Patient and Visual Clot resulted in significantly lower total raw NASA-TLX scores and all subscores, including high self-rated performance, when compared with conventional monitoring (all $P<.001$).

Conclusions: This study validated a modified raw NASA-TLX questionnaire for patient monitoring tasks. The scores obtained correctly represented the assumed influences of the examined covariates on the perceived workload. We reported high criterion validity. The NASA-TLX questionnaire appears to be a reliable tool for measuring subjective workload. Further research should focus on its applicability in a clinical setting.

(J Med Internet Res 2020;22(9):e19472) doi:10.2196/19472

KEYWORDS

workload; questionnaires; National Aeronautics and Space Administration Task Load Index; awareness; situation awareness; patient monitoring; thromboelastometry

Introduction

Workload

The World Health Organization considers attentive anesthesia providers to be essential to prevent perioperative disability and death [1]. To maintain high quality of care, all factors negatively affecting human performance should be minimized. Various subjective factors, such as high complexity of tasks, stressful personal factors, high-pressure working environment, lack of situation awareness, fatigue, and increased workload, all impair human performance, the quality of care, and thus patient safety [2-4]. The International Organization for Standardization defines workload as the totality of external conditions and requirements in a work system, which affects the physiological and/or psychological state of a person [5]. The perceived workload and a person's ability to create and maintain adequate situation awareness are interconnected [6]. Situation awareness incorporates the perception of the current status of a situation's critical elements, with understanding of their meaning, and the projection of this knowledge into the near future [2,7]. For physicians and nurses working inside the operating theatre or intensive care unit, it is crucial to keep situation awareness at a high level through constant mental reassessment. This process however requires substantial cognitive effort. A high workload is a psychological stress factor that takes up part of a person's naturally limited working memory and ultimately leads to fewer cognitive resources being available.

Hence, accurate assessment of workload is of great importance to manage stressors. Various methods for quantifying perceived workload have been described, which can be divided into the following two large groups: subjective assessment through questionnaires and objective physiological assessment of variables such as heart rate, galvanic skin resistance, breathing rate, pupil diameter, and blinking frequency [8,9].

National Aeronautics and Space Administration Task Load Index

The National Aeronautics and Space Administration Task Load Index (NASA-TLX) provides the most widely accepted and validated tool to measure overall workload after completing a task [10-12]. It was initially created by the Human Performance Research Group at NASA Ames Research Center for the aviation industry. Since then, its use has expanded to many other fields such as computer science [13], psychophysiology [14], and transportation [15]. The NASA-TLX is a multidimensional tool that contains six predefined dimensions. Three dimensions measure the demands imposed on the subject (mental, physical, and temporal demands), and three dimensions focus on how the subject deals with the task at hand (self-rated performance, effort, and frustration level). Dividing the workload into six subcategories intends to reduce variability among subjects and show the source of the workload. There are two different methods of using the NASA-TLX tool. The weighted NASA-TLX score is a two-step process, in which the user first rates all six subcategories after completing a specific task and then weights the contribution of each factor in a predefined manner. This aims to further understand which potential source accounts mostly for the perceived workload. On the other hand, in the raw NASA-TLX score, the user rates all six subcategories after completing a specific task, without weighing them. The result is the arithmetic mean of all subscales. Research has shown that the raw NASA-TLX has a high correlation with the weighted one [11], but is more time efficient and simpler to apply [16,17]. We used the raw NASA-TLX questionnaire in several studies within the scope of our research activities in the field of patient monitoring and situation awareness–oriented visualization technologies [18].

Patient Monitoring Tasks

Patient monitoring can be generalized as continuous observation of a condition or certain parameters, regardless of the method used [19,20]. In previous studies, we simplified the presentation of information in different patient monitoring devices by creating two new situation awareness–oriented information transfer technologies called Visual Patient [21,22] and Visual Clot [23]. We further discuss the functionality and applicability of these technologies in the methods section. In tested scenarios, both Visual Patient and Visual Clot helped the health care provider to make correct diagnoses faster and with less perceived cognitive workload compared with standard presentation only.

Study Aims and Hypotheses

Only few studies have investigated the validity of the raw NASA-TLX score in the health care sector [24-27]. There is no existing gold standard in measuring workload. The primary objective of this study was to validate the raw NASA-TLX questionnaire in patient monitoring tasks by investigating a broad set of over 1000 NASA-TLX scores. We based our evaluation of the robustness of the questionnaire against covariates not associated with workload and on the relation of the NASA-TLX scores with covariates that are known to influence workload. Therefore, we investigated criterion validity. We expect that the scores will increase with the difficulty of the task and distraction, and decrease with the work experience and self-confidence of the user. Further, we assume that the new visualization technologies will reduce all dimensions of the workload approximately uniformly and that not one workload aspect alone will be responsible for observed reductions.

Methods

Ethical Permission

The leading ethics committee (Cantonal Ethics Committee of Zurich, Switzerland) reviewed the study protocols for all four

included studies and issued a declaration of no objection for each one of them (reference numbers: 2016-00103, 2017-00795, and 2018-00933). Reference number 2017-00795 covers two of the studies, as they were performed simultaneously and included the same participants. Additionally, before participation in this study, we obtained written informed consent from all participants to collect data for scientific purposes and publications.

Study Design and Data Collection

To validate the NASA-TLX, we analyzed data from four different studies all performed by anesthesia and intensive care providers at the University Hospital Zurich and Cantonal Hospital Winterthur in Switzerland, and University Hospital Frankfurt in Germany [21,23,28,29]. Table 1 provides an overview of the included studies. All of them were investigator-initiated, computer-based, prospective, dual-center studies, and have been published in the past 2 years [21,23,28,29]. In all studies, we based the recruitment of participants on their clinical availability. The day before data collection, we contacted available individuals by institutional email or telephone to plan participation in the study. We then carried out data collection during regular working hours. During this time, we released them from all other duties, including telephone availability.

Table 1. Description of the four studies used with the respective participant numbers and completed National Aeronautics and Space Administration Task Load Index questionnaires.

Study title[a]	Location	Number of participants	NASA-TLX[b] questionnaires, n
Using an Animated Patient Avatar to Improve Perception of Vital Sign Information by Anesthesia Professionals	USZ[c] and KSW[d]	32	128
Avatar-Based Versus Conventional Vital Sign Display in a Central Monitor for Monitoring Multiple Patients: A Multicenter Computer-Based Laboratory Study	USZ and KSW	38	312
Effects of a Standardized Distraction on Caregivers' Perceptive Performance with Avatar-Based and Conventional Patient Monitoring: A Multicenter Comparative Study	USZ and KSW	38	312
Improving Decision Making Through Presentation of Viscoelastic Tests as 3D Animated Blood Clot: the Visual Clot	USZ and UKF[e]	60	720

[a]The second and third studies included the same participants.

[b]NASA-TLX: National Aeronautics and Space Administration-Task Load Index.

[c]USZ: University Hospital Zurich.

[d]KSW: Cantonal Hospital of Winterthur.

[e]UKF: University Hospital Frankfurt.

In all included studies, we used the original formulation of the NASA-TLX questionnaire provided by the official NASA website. However, we modified the raw NASA-TLX surveys from six to only five dimensions by removing the physical demand question as our tasks did not require any physical effort. Table 2 shows the modified raw NASA-TLX questionnaire we used in our studies. Participants were staff or resident anesthesiologists and nurse anesthetists with completed specialization qualification. They were all employed in the abovementioned centers during the course of the trials. We selected the participants randomly, regardless of sex, age, job description, staff position, or education level. Participation was voluntary, and none of the subjects received compensation in any form.

Table 2. Description of the modified raw National Aeronautics and Space Administration Task Load Index questions and rating scale.

Workload	Descriptive question	Endpoint[a]
Mental demand	Was the task easy or demanding, simple or complex?	0 to 100
Temporal demand	How much time pressure did you feel performing the task?	0 to 100
Self-rated performance[b]	How successful or satisfied did you feel upon the performance or completion of the given task?	0 to 100
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?	0 to 100
Frustration level	How insecure, discouraged, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?	0 to 100

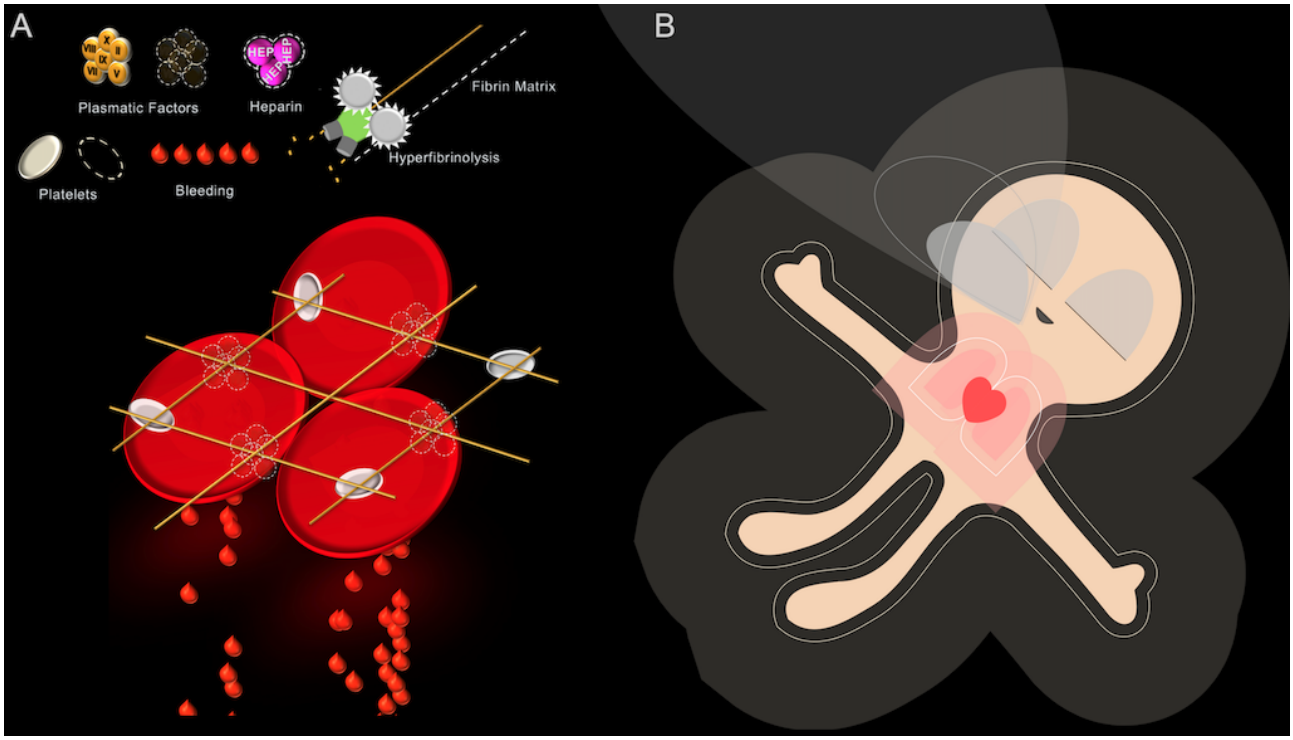
[a]Subjects rated the subscores numerically from 0 (very low) to 100 (very high). The endpoints regarding performance are inverted with 0 indicating very good performance and 100 indicating very poor performance.

[b]The term self-rated performance indicates the performance dimension of the National Aeronautics and Space Administration Task Load Index score.

Visual Patient and Visual Clot Technologies

Visual Patient technology is a situation awareness–oriented visualization tool that displays up to 11 of the most frequently used patient vital signs in the form of a patient avatar in addition to the standard monitoring screen [21,22]. Visual Patient transforms the conventional numerical and waveform display of the vital signs in real time into a patient avatar with the ability to adjust its color, shapes, and rhythmic movements depending on the patient's current situation. For example, if a patient shows an elevated body temperature (ie, more than 37.5°C), the avatar will display heat radiation rising from the patient model. Another example is the patient's neuromuscular state of relaxation, which is best described by the train-of-four ratio. If the mentioned ratio drops below 20, the patient avatar changes its posture and goes into a floppy state. In situations where several vital signs are out of range, Visual Patient technology is able to illustrate them simultaneously. Visual Clot technology [23] on the other hand illustrates abstract conventional viscoelastic rotational thromboelastometry (ROTEM) readings in form of a three-dimensional animated blood clot. This aims to help the user create a simpler mental model of the current coagulation disorder. The image consists of a blood clot model with its various components such as platelets, coagulation factors, and enzymes. Based on established conventional monitoring values, the visualization shows these coagulation components as either present or not. Figure 1 illustrates both the Visual Clot and Visual Patient technologies. We have provided instructional videos in Multimedia Appendix 1 and Multimedia Appendix 2 explaining how both technologies work.

Figure 1. Graphic showing the Visual Clot and Visual Patient technologies. (A) Bleeding Visual Clot illustrated with its different coagulation components. These are either present or absent, depending on the coagulation status. (B) Visual Patient with its different visualizations of vital parameters.



A relevant limitation for both visualization technologies is the absence of quantification values. Both avatars only display the following three states: too low, within range, and too high. This limitation defines the intended use of the technologies, which is the situation awareness–oriented supplementation of conventional monitoring methods. They were not designed to replace numerical values. Both visualization technologies were invented, prototyped, and patented by our research group at the Institute for Anesthesiology of the University Hospital Zurich. We are developing Visual Patient into a product under a Joint Development and Licensing Agreement with Royal Philips NV and Philips Medizin-Systeme Böblingen GmbH. Regarding Visual Clot, we have signed a letter of intent with Instrumentation Laboratory. Both technologies are in the prototype stage of development and neither Visual Patient nor Visual Clot is currently CE certified as a medical device.

Patient Monitoring Tasks

We assessed the perceived workload using the raw NASA-TLX questionnaire after performing patient monitoring tasks using either one of the newly developed visualization tools (Visual Patient or Visual Clot) or respective conventional monitoring alone. The monitoring scenarios appeared in randomized order for predefined relatively short periods. Furthermore, the participants had to rate their diagnostic confidence level on a four-point Likert scale ranging from "very unconfident" to "very confident" to further assess uncertainty as a psychological stress factor. In the four mentioned studies, we generated workload through specific tasks as follows:

- In the primary Visual Patient study [21], we displayed four patient monitoring scenarios on a computer for 3 or 10 seconds, portraying either the conventional monitoring screen or the animated patient avatar in randomized order.

Subsequently, the participants had to recall the patient's condition.

- The Visual Patient central monitoring study [29] included four different scenarios in which we showed two critical and four healthy patients simultaneously on a central monitor (as in the intensive care unit or operating room) for 10 or 30 seconds. Afterwards, the participants had to recall the patient's condition.
- The same participants (as in the central monitoring study [29]) also took part in the Visual Patient distraction study [28]. In this study, we distracted the participants with standardized simple calculation tasks. Simultaneously, we showed them different monitoring scenarios either with the conventional monitoring screen alone or with the help of Visual Patient. The workload task also demanded recall of the patient's condition.
- Finally, in the Visual Clot study [23], we showed the participants 12 scenarios representing six different hemostatic conditions, which they had to solve using either standard viscoelastic ROTEM results alone or using the matching animated blood clot.

Validation Method and Studied Covariates

We investigated the robustness of the NASA-TLX questionnaire based on criterion validity, which can be further divided into predictive and concurrent validity. Predictive validity indicates the extent to which an assumption under investigation can be predicted [30]. Concurrent validity compares the result in question with an already known relationship of the same variable [30].

In all included studies, we had consistently recorded 11 different covariates. In this pooled analysis, we determined the expected impact on workload from these measured covariates, using both literature research and logical deductions. It was described that more professional experience should result in lower cognitive workload [31,32]. Since we had recorded the educational stage of the participants through their job descriptions, we regarded this equal to experience. Further, we correlated the measured covariate self-rated confidence with experience and test performance through both literature research and logical reasoning. In order to not confuse the terms, in this manuscript, we defined test performance as the actual testing outcome of the participants and self-rated performance as the subscores of the NASA-TLX. We expected a decreased NASA-TLX score in participants with high self-confidence. Regarding the actual test performance, we investigated predictive validity. We expected good performing participants to perceive less workload. As far as the task's difficulty is concerned, it cannot be assessed directly as various factors influence it. We defined scenarios in which distraction occurred as more difficult. They divide one's attention and thus affect the work-related receptiveness [33]. Therefore, we expected the results to show an increased perceived workload and thus correspond to concurrent validity. We did not find any evidence in a literature search and everyday clinical practice of a different perceived workload between the sexes. Therefore, we regarded this as a covariate without an expected influence on workload.

Statistical Analysis

We validated the raw NASA-TLX score and the different subscores by fitting mixed linear regression models with a random intercept per person (to cover repeated measurements) and a random intercept per study. We fitted univariate models with binary test performance (task correct or incorrect), binary confidence (unconfident or confident), distraction, center, gender, profession, binary daytime (above or below the median), binary playback sequence (first or second half of the tasks), and central monitor performance as covariates.

We analyzed interactions of some covariates with the technology variable to explore more in depth why the NASA-TLX and its subscores improved in the case of the new visual technologies. To explore which subscore benefited the most from the new visualization technologies, we calculated univariate models for each subscore that included only the technology variable, and we compared the size of the estimated coefficients. To characterize the individuals who benefited the most from the introduction of the new technologies, we fitted a joint model for the total NASA-TLX with the technology variable and several other covariates. In one additional model per variable, we included an interaction term between technology and the respective covariate to see if the impact of certain variables was particularly strong in the case of the new technologies.

All analyses were performed using R version 3.6.2 (R Foundation for Statistical Computing). We considered a P value <.05 to indicate statistical significance.

Data Sharing Statement

We provide the complete data used for this study in Multimedia Appendix 3.

Results

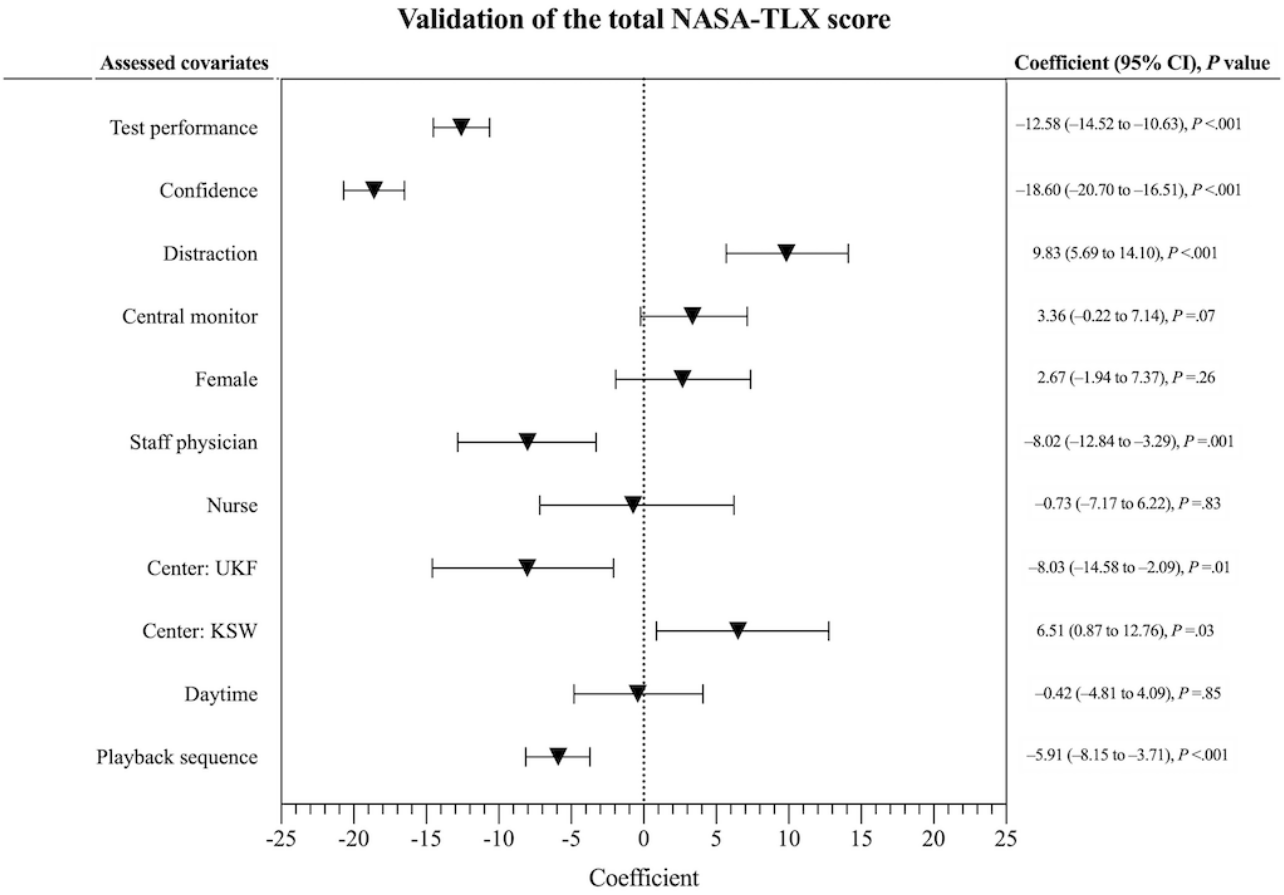
Study and Participant Characteristics

In all four evaluated studies [21,23,28,29], 128 anesthesia providers participated and rated a total of 1160 NASA-TLX questionnaires. Overall, 552 of 1160 (47.6%) NASA-TLX surveys were collected at the University Hospital Zurich, 360 of 1160 (31.0%) at the University Hospital Frankfurt, and 248 of 1160 (21.4%) at the Cantonal Hospital Winterthur. Further, 648 of 1160 (55.9%) ratings were provided by male participants, and 512 of 1160 (44.1%) by female participants. According to job description, 556 of 1160 (47.9%) ratings were provided by staff physicians, 432 of 1160 (37.2%) by resident physicians, and 172 of 1160 (14.8%) by nurse anesthetists. Comparing the technologies used, 62.1% (720/1160) of all data originated from the three Visual Patient projects, and 37.9% (440/1160) from the Visual Clot study.

Quantitative Analyses of the NASA-TLX Questionnaire

In Figure 2 and Figure 3, we illustrate the correlation of the total NASA-TLX workload score and its subscores after performing univariate analysis with test performance, confidence, distraction, center, gender, job description, daytime, and playback sequence as different covariates. The playback sequence was described as the first or second half of the task. We provide the NASA-TLX coefficient and the 95% CI.

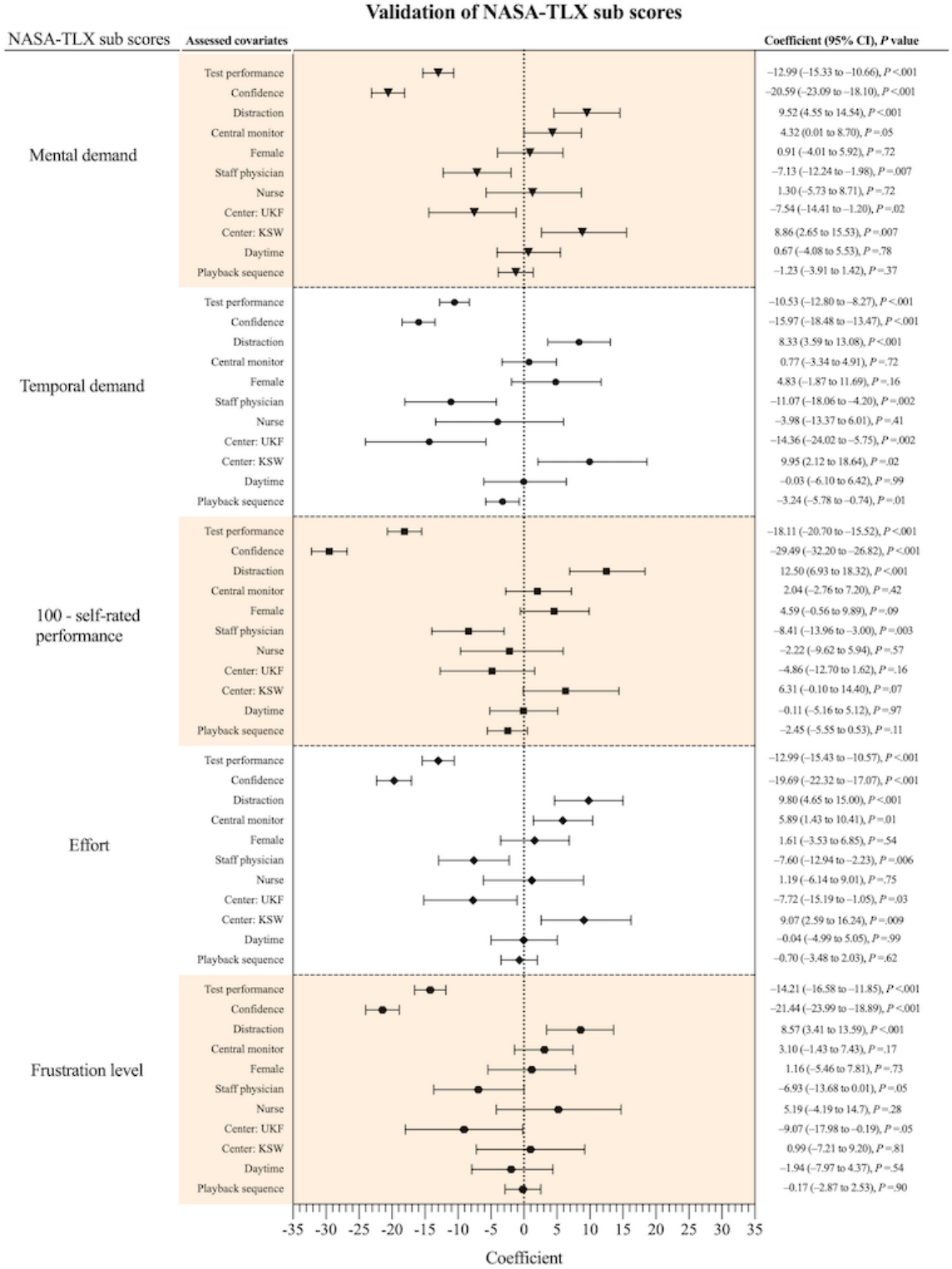
Figure 2. Correlation of different covariates with the total score of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) workload assessment tool. Left and right of the dashed line indicate lower and higher perceived workload, respectively. KSW: Cantonal Hospital Winterthur; UKF: University Hospital Frankfurt.



The total workload score analysis showed that participants' test performance and higher confidence correlated significantly with lower NASA-TLX scores (both *P*<.001). Compared with the University Hospital Zurich, participants from the University Hospital Frankfurt had significantly lower total workload scores (−8.03, 95% CI −14.58 to −2.09; *P*=.01). Further, the second half of the playback sequence had a significant positive effect on lowering the perceived workload (−5.91, 95% CI −8.15 to −3.71; *P*<.001). Regarding the job position, resident physicians served as the comparison. Staff physicians (−8.02, 95% CI −12.84 to −3.29; *P*=.001) rated significantly lower workload than residents, whereas nurses did not show any significant difference compared with residents (*P*=.83). Distraction and the Cantonal Hospital Winterthur compared with the University Hospital Zurich correlated significantly with higher rated NASA-TLX scores (*P*<.001 and *P*=.03, respectively). The other listed covariates did not show any relevant difference.

Figure 3 illustrates the entire evaluation of the examined covariates for the subscores of the NASA-TLX. Good test performance and high confidence level after performing the task correlated significantly with lower workload scores in every subcategory of the NASA-TLX (all *P*<.001). Staff physicians also differed significantly from resident physicians with lower workload scores in every subcategory, except the frustration level (−6.93, 95% CI −13.68 to 0.01; *P*=.05). Distraction was the only covariate to show a relevant effect on increasing the workload in every subscore of the NASA-TLX. Comparing the participants at the different centers with those at the University Hospital Zurich, the participants at the University Hospital Frankfurt had less mental (*P*=.02) and temporal (*P*=.002) demands, and less perceived effort (*P*=.003). The participants at the Cantonal Hospital Winterthur had more mental (*P*=.007) and temporal demands (*P*=.02), and more required effort (*P*=.03). We observed no gender difference in all subscores. Showing several patients simultaneously on a central monitor correlated significantly with increased mental demand (*P*=.05) and increased effort (*P*=.01) to fulfill the task. We observed no effect for temporal demand (0.77, 95% CI −3.34 to 4.91; *P*=.72), frustration level (3.10, 95% CI −1.43 to 7.43; *P*=.17), and self-rated performance (2.04, 95% CI −2.76 to 7.20; *P*=.42) using a central monitor. Playback sequence showed its only significant effect on the temporal demand subscore with lower perceived workload in the second half of the task (−3.24, 95% CI −5.78 to −0.74; *P*=.01).

Figure 3. Correlation of different covariates with subscores of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) workload assessment tool. The original NASA-TLX questionnaire evaluated performance on an inverted X-axis from perfect to failure, with a low raw score corresponding to good self-rated performance. Therefore, to ensure that the X-axis in this figure is the same, the self-rated performance is displayed inverted. KSW: Cantonal Hospital Winterthur; UKF: University Hospital Frankfurt.



Performing the same tasks, we compared the conventional monitoring devices with the respective visualization technologies. The total NASA-TLX score and all listed subscores correlated significantly with lower workload scores when using Visual Patient and Visual Clot (all P<.001). Table 3 demonstrates the comparison of the new visualization technologies with respective conventional monitoring. The self-rated performance dimension showed the most relevant change in the perceived workload assessment (coefficient −18.28).

Table 3. Overall comparison of the new visual technologies Visual Patient and Visual Clot with conventional monitoring.

Variable	Coefficient	CI lower	CI upper	*P* value
Total	−14.36	−16.06	−12.65	<.001
Mental demand	−15.79	−17.85	−13.72	<.001
Temporal demand	−10.53	−12.59	−8.46	<.001
Self-rated performance[a]	−18.28	−20.63	−15.93	<.001
Effort	−16.80	−18.92	−14.68	<.001
Frustration level	−15.97	−18.06	−13.87	<.001

[a]The term self-rated performance indicates the performance dimension of the National Aeronautics and Space Administration Task Load Index score.

Data Sharing Statement

We provide the complete analysis of used data for this study in Multimedia Appendix 4.

Discussion

Principal Findings

This pooled analysis examined a broad set of raw NASA-TLX scores obtained after performing patient monitoring tasks. Good test performance, high self-confidence in successfully completing a task, high hierarchical job position, and training correlated with decreased raw NASA-TLX workload scores, whereas distraction scenarios increased perceived workload. The questionnaire was robust against nonworkload-related factors such as gender. The new visualization technologies Visual Patient and Visual Clot both decreased all workload dimensions compared with conventional monitoring alone and patient monitoring reference values provided by the literature [11]. Reducing workload helps to free cognitive resources, which can be used to understand the provided monitoring information crucial to maintain a high quality of care [7].

The analysis of 1160 raw NASA-TLX questionnaires showed that self-rated high confidence correlated significantly with lower overall raw NASA-TLX scores and its subscores (all $P<.001$). Furthermore, staff physicians had significantly lower total workload scores compared with residents. During clinical residency training, inexperienced physicians go through a period of professional and personal growth. They acquire knowledge and skills [6], and experience many challenging clinical situations that build professional competency and thereby increase their confidence at work [34]. Generally, staff physicians are more certain of their clinical abilities and thus exude more confidence than residents. The participants at the University Hospital Frankfurt also rated lower total workload scores compared with those at the University Hospital Zurich. This is in line with our abovementioned train of thought, as the participating physicians in Frankfurt had more professional experience overall than those in Zurich [23]. When comparing nurses with residents, there was no difference in the perceived total workload after performing the same monitoring task. We explain this interesting finding on the basis of existing practical experience. All nurses who took part in this study had completed their specialty qualification and thus often had more professional experience than the residents who were still in training. Their self-confidence may have increased owing to their completed education. Our results confirm our hypothesis that confidence as a trait, which is compatible with a higher job position, lowers perceived workload. It is known from other domains that individuals with higher confidence make better decisions [35]. The World Health Organization aims to continuously improve, protect, and promote the health, safety, and well-being of all workers [36]. Analyzing workload in the perioperative setting can lead directly to practical work-based implications, such as tailored task assignment and improved training plans, for residents or other personnel. This efficient identification and management of workload influences employee well-being [37] and lowers the source of fatigue from work overload, which is an independent risk factor for exhaustion and burnout [38]. Moreover, workload has been shown to be associated with adverse patient outcomes [39-41].

Good test performance in conducting patient monitoring tasks correlated with lower total raw NASA-TLX scores, as well as all subscores, including high self-rated performance. We propose that this connection resulted from both training and experience. When a person has received a lot of training and thus experience in performing a task, the perceived workload decreases and the actual performance increases. Nevertheless, this finding allows not drawing a linear relationship between task performance and workload. This relation is complex when investigated in more detail. Both excessive workload and low demand situations can degrade performance [42], and additional factors, such as personal resilience, influence the work capacity to a great extent. This shows that there are other factors apart from workload that influence task performance.

Further, our results showed that standardized distraction negatively affected the raw NASA-TLX scores with all subcategories. This is in line with our hypothesis that distractions increase perceived workload. They take up part of the already limited mind while coping with several things simultaneously and reduce the work-related memory capacity of humans. It was shown that distractions impair situational awareness and thus affect clinical decision making [33,43]. The source of most anesthesia adverse events lies in reduced situational awareness [4,44]. This further demonstrates the importance of minimizing working environment distractions in areas that involve workload-sensitive tasks such as patient monitoring inside the operating theatre [45]. In one study, 22 of 25 (88%) anesthesia providers agreed to the statement that human factor problems do lead to critical information not being received [46].

The newly developed visualization technologies Visual Patient and Visual Clot were associated with decreased perceived workload as expected. These technologies have been developed to link several sources of information together and create an avatar-based visualization aimed to facilitate the mental model of the current situation [21-23]. Endsley defined the goal of optimal situation awareness–oriented design to transfer information as quickly as possible and with the least cognitive effort [7].

Decreasing workload while monitoring a patient reduces psychological stress by saving cognitive resources, which are especially needed in critical situations. In 2015, Grier et al [11] examined a vast amount of published NASA-TLX scores in a meta-analysis showing the distribution frequency of the scores by task type. They analyzed 174 monitoring-type tasks involving change detection, speech detection, and vigilance tasks. A mean NASA-TLX global workload score of 52.24 with an IQR of 22.66 (39.97-62.63) was reported [11]. This fits the mean NASA-TLX global workload score of 54.60 with an IQR of 27.0 (42.0-69.0), which we observed when evaluating our conventional monitoring tasks. Using the situation awareness–oriented visualization monitoring technologies, we found a mean NASA-TLX global workload score of 40.2 with an IQR of 31.0 (25.0-56.0). These technologies lowered the mean perceived workload compared with the literature value by 23.0%. This is comparable to a workload that occurs when driving a car (mean 41.52, IQR 23.7 [28.05-51.73]) [11].

This study contributes to the ongoing validation of the NASA-TLX in the medical field. The results are consistent with those of other studies that have found high construct validity of the NASA-TLX score [24,25]. These studies described an increase in the workload score due to distraction [47], case complexity [24], and low task performance [48], and a reduction due to training [49]. Other studies found good correlation with other questionnaires for workload assessment [50,51] and with physiological stress measurements [52,53].

Strengths and Limitations

Our study had several limitations. All tasks used to validate the raw NASA-TLX questionnaire took place in a testing environment with clearly defined monitoring limits within scenarios. Perceived workload and therefore its assessment might differ in a clinical setting, where each situation must be interpreted independently. Future studies should test the applicability of the raw NASA-TLX in the clinical setting. However, it is plausible that this effect is marginal as the score reflects a subjective evaluation while information intake and thus perception of workload remain similar. Further, we assessed and validated a modified version of the questionnaire. Since our monitoring tasks did not require any relevant physical effort, we removed this dimension from the scale in order not to distort the value of the total workload assessed. Future studies are required to investigate whether such a modification affects the internal consistency of the NASA-TLX questionnaire. Moreover, a true validation would correlate obtained scores with objectively measurable stress characteristics such as heart rate and pupil diameter. Another limitation is that all patient monitoring scenarios took place in central Europe in high quality of care hospitals. Perceived workload can differ in other parts of the world and might influence the reproducibility of the assessment. Finally, interpretation of the raw NASA-TLX scores requires comparative values after performing similar tasks. Therefore, more available data using the same questionnaire in patient monitoring tasks would reduce this limitation. Nevertheless, the results of this study support the use of the raw NASA-TLX score in patient monitoring tasks.

Among the particular strengths of this study are the multicenter design, the large data set, and the consistent recording of identical covariates in all included studies. We examined more than 1000 raw NASA-TLX questionnaires, which were completed by 130 participants, with individual selection solely based on daily clinical availability. This large proportion of staff from the respective institutions constitutes a representative sample. Further, in all included studies, intraparticipant comparisons took place as there was an evaluation of the same monitoring task with both examined interface designs (ie, conventional versus visualization). This greatly reduces the influence of confounding variables if the main factor responsible for the difference remains the interface modality shown, and it further increases the quality of the study.

Conclusions

For patient monitoring, this study validated a modified version of the raw NASA-TLX questionnaire, in which the physical dimension had been removed from the scale owing to the nature of the given tasks. The obtained scores correctly depicted the assumed influences of the covariables that affect perceived workload. This provided a high extent of criterion validity. The modified raw NASA-TLX questionnaire appears to be a reliable tool for measuring the subjective workload of anesthesia providers who monitor patients inside the operating room. Further research is needed to investigate the applicability of the NASA-TLX questionnaire in the clinical setting and its transferability to personnel working in intensive care units. Moreover, a true validation study for the subjective workload assessment should correlate the NASA-TLX scores with objectively measurable stress characteristics such as heart rate and pupil diameter.

Acknowledgments

The authors are thankful to the study participants for their time and effort. The Institute of Anesthesiology of the University Hospital Zurich, Zurich, Switzerland, funded this study, and DWT received a career development grant from the University of Zurich.

Authors' Contributions

SS, GM, TR, DS, CN, and DT helped to design the study; JR, AK, CN, and DT helped to collect the data; SS, JB, and DT helped to analyze the data; and SS, MG, TR, AK, JB, JR, DS, CN, and DT helped to write the manuscript and approved the final version.

Conflicts of Interest

The University of Zurich owns the intellectual property rights to the technologies described in this manuscript and registered "Visual Clot" and "Visual Patient" as trademarks. The University of Zurich and Instrumentation Laboratory Company/Werfen Corporation, Bedford, MA, USA, signed a letter of intent regarding a proposed joint development and licensing agreement to develop a product based on the concept of Visual Clot. As designated inventors DS, CN, and DT may receive royalties in the event of commercialization. The authors DT, DS, and CN are in a joint development agreement with the monitoring manufacturer Philips Healthcare (Koninklijke Philips NV) for Visual Patient. Within the framework of this cooperation, a monitoring system based on an avatar will be developed. Within the framework of licensing the technology via the University, the authors DT and CN might receive royalties as designated inventors in the event of successful product release. DS academic department is receiving grant support from the Swiss National Science Foundation, Berne, Switzerland; the Swiss Society of Anesthesiology and Reanimation (SGAR), Berne, Switzerland; the Swiss Foundation for Anesthesia Research, Zurich, Switzerland; and Vifor SA, Villars-sur-Glâne, Switzerland. DS is the co-chair of the ABC-Trauma Faculty, sponsored by unrestricted educational grants from Novo Nordisk Health Care AG, Zurich, Switzerland; CSL Behring GmbH, Marburg, Germany; LFB Biomédicaments, Courtaboeuf Cedex, France; and Octapharma AG, Lachen, Switzerland. Dr Spahn received honoraria/travel support for consulting or lecturing from Danube University of Krems, Austria; US Department of Defense, Washington, USA; European Society of Anesthesiology, Brussels, BE; Korean Society for Patient Blood Management, Seoul, Korea; Korean Society of Anesthesiologists, Seoul, Korea; Network for the Advancement of Patient Blood Management, Haemostasis and Thrombosis, Paris, France; Baxalta Switzerland AG, Volketswil, Switzerland; Bayer AG, Zürich, Switzerland; B. Braun Melsungen AG, Melsungen, Germany; Boehringer Ingelheim GmbH, Basel, Switzerland; Bristol-Myers-Squibb, Rueil-Malmaison Cedex, France and Baar, Switzerland; CSL Behring GmbH, Hattersheim am Main, Germany and Berne, Switzerland; Celgene International II Sàrl, Couvet, Switzerland; Daiichi Sankyo AG, Thalwil, Switzerland; Ethicon Sàrl, Neuchâtel, Switzerland; Haemonetics, Braintree, MA, USA; Instrumentation Laboratory (Werfen), Bedford, MA, USA; LFB Biomédicaments, Courtaboeuf Cedex, France; Merck Sharp & Dohme, Kenilworth, New Jersey, USA; PAION Deutschland GmbH, Aachen, Germany; Pharmacosmos A/S, Holbaek, Denmark; Photonics Healthcare BV, Utrecht, Netherlands; Pfizer AG, Zürich, Switzerland; Pierre Fabre Pharma, Alschwil, Switzerland; Roche Diagnostics International Ltd, Reinach, Switzerland; Sarstedt AG & Co, Sevelen, Switzerland and Nümbrecht, Germany; Shire Switzerland GmbH, Zug, Switzerland; Tem International GmbH, Munich, Germany; Vifor Pharma, Munich, Germany, Neuilly sur Seine, France, and Villars-sur-Glâne, Switzerland; Vifor (International) AG, St. Gallen, Switzerland; and Zuellig Pharma Holdings, Singapore, Singapore. CN and DT received travel support for consulting and lecturing from Instrumentation Laboratory (Werfen), Bedford, MA, USA. CN and DT received proof-of-concept funding from the University of Zurich to prototype Visual Patient. The University of Zurich and Koninklijke Philips NV, Amsterdam, Netherlands entered a joint development and licensing agreement to develop a product based on Visual Patient. As inventors, CN and DT may receive royalty payments in the event of commercialization. AK received honoraria for lecturing from Bayer AG (Switzerland). The other authors do not have any conflicts of interest.

Multimedia Appendix 1
Instructional video explaining the Visual Clot technology.
[MOV File , 13733 KB - jmir_v22i9e19472_app1.mov]

Multimedia Appendix 2
Instructional video explaining the Visual Patient technology.
[MOV File , 14845 KB - jmir_v22i9e19472_app2.mov]

Multimedia Appendix 3
Complete data for this study.
[XLSX File (Microsoft Excel File), 5971 KB - jmir_v22i9e19472_app3.xlsx]

Multimedia Appendix 4
Complete analysis of used data for this study.
[PDF File (Adobe PDF File), 145 KB - jmir_v22i9e19472_app4.pdf]

References

1. World Health Organization. Safe Surgery Saves Lives. PACEsetterS 2008;5(3):21. [doi: 10.1097/01.jbi.0000393253.18990.c6]

2. Schulz CM, Endsley MR, Kochs EF, Gelb AW, Wagner KJ. Situation awareness in anesthesia: concept and research. Anesthesiology 2013 Mar;118(3):729-742. [doi: 10.1097/ALN.0b013e318280a40f] [Medline: 23291626]
3. Reason J. Human error: models and management. West J Med 2000 Jun;172(6):393-396 [FREE Full text] [doi: 10.1136/ewjm.172.6.393] [Medline: 10854390]
4. Schulz CM, Krautheim V, Hackemann A, Kreuzer M, Kochs EF, Wagner KJ. Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. BMC Anesthesiol 2016 Jan 16;16:4. [doi: 10.1186/s12871-016-0172-7] [Medline: 26772179]
5. Nachreiner F. International standards on mental work-load--the ISO 10,075 series. Ind Health 1999 Apr;37(2):125-133 [FREE Full text] [doi: 10.2486/indhealth.37.125] [Medline: 10319562]
6. Goldman J, Wong BM. Nothing soft about 'soft skills': core competencies in quality improvement and patient safety education and practice. BMJ Qual Saf 2020 Aug;29(8):619-622. [doi: 10.1136/bmjqs-2019-010512] [Medline: 32066572]
7. Endsley MR. Designing for Situation Awareness An Approach to User-Centered Design, Second Edition. Boca Raton: CRC Press, Inc; Apr 19, 2016.
8. Kremer L, Leeser L, Breil B. Mental Workload Relating Health Information System - A Literature Review. Stud Health Technol Inform 2019 Sep 03;267:289-296. [doi: 10.3233/SHTI190840] [Medline: 31483284]
9. Al Abdi RM, Alhitary AE, Abdul Hay EW, Al-Bashir AK. Objective detection of chronic stress using physiological parameters. Med Biol Eng Comput 2018 Dec;56(12):2273-2286. [doi: 10.1007/s11517-018-1854-8] [Medline: 29911251]
10. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2016 Nov 05;50(9):904-908. [doi: 10.1177/154193120605000909]
11. Grier RA. How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2016 Dec 20;59(1):1727-1731. [doi: 10.1177/1541931215591373]
12. Dias RD, Ngo-Howard MC, Boskovski MT, Zenati MA, Yule SJ. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. Br J Surg 2018 Apr;105(5):491-501 [FREE Full text] [doi: 10.1002/bjs.10795] [Medline: 29465749]
13. Peng X, Xu Z, Peng X. [Comparison of medical student's mental workload between VDT and paper-based reading]. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi 2008 Dec;26(12):738-740. [Medline: 19257934]
14. Fournier LR, Wilson GF, Swain CR. Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. Int J Psychophysiol 1999 Jan;31(2):129-145. [doi: 10.1016/s0167-8760(98)00049-x] [Medline: 9987059]
15. Qiao F, Zhang R, Yu L. Using NASA-Task Load Index to Assess Drivers' Workload on Freeway Guide Sign Structures. In: ICCTP 2011: Towards Sustainable Transportation Systems. Nanjing, China: ICCTP; 2011:4342-4353.
16. Hendy KC, Hamilton KM, Landry LN. Measuring Subjective Workload: When Is One Scale Better Than Many? Hum Factors 2016 Nov 23;35(4):579-601. [doi: 10.1177/001872089303500401]
17. Nygren TE. Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload. Hum Factors 2016 Nov 23;33(1):17-33. [doi: 10.1177/001872089103300102]
18. Tscholl DW, Weiss M, Handschin L, Spahn DR, Nöthiger CB. User perceptions of avatar-based patient monitoring: a mixed qualitative and quantitative study. BMC Anesthesiol 2018 Dec 11;18(1):188 [FREE Full text] [doi: 10.1186/s12871-018-0650-1] [Medline: 30537934]
19. Kieseier BC, Bauer M. [Patient monitoring 4.0]. Nervenarzt 2019 Dec;90(12):1205-1206. [doi: 10.1007/s00115-019-00823-w] [Medline: 31807830]
20. Zimlichman E, Szyper-Kravitz M, Unterman A, Goldman A, Levkovich S, Shoenfeld Y. How is my patient doing? Evaluating hospitalized patients using continuous vital signs monitoring. Isr Med Assoc J 2009 Jun;11(6):382-384 [FREE Full text] [Medline: 19697594]
21. Tscholl DW, Handschin L, Neubauer P, Weiss M, Seifert B, Spahn DR, et al. Using an animated patient avatar to improve perception of vital sign information by anaesthesia professionals. Br J Anaesth 2018 Sep;121(3):662-671 [FREE Full text] [doi: 10.1016/j.bja.2018.04.024] [Medline: 30115265]
22. Tscholl DW, Rössler J, Said S, Kaserer A, Spahn DR, Nöthiger CB. Situation Awareness-Oriented Patient Monitoring with Visual Patient Technology: A Qualitative Review of the Primary Research. Sensors (Basel) 2020 Apr 09;20(7) [FREE Full text] [doi: 10.3390/s20072112] [Medline: 32283625]
23. Rössler J, Meybohm P, Spahn DR, Zacharowski K, Braun J, Nöthiger CB, et al. Improving decision making through presentation of viscoelastic tests as a 3D animated blood clot: the Visual Clot. Anaesthesia 2020 Aug;75(8):1059-1069. [doi: 10.1111/anae.14985] [Medline: 32030729]
24. Ruiz-Rabelo JF, Navarro-Rodriguez E, Di-Stasi LL, Diaz-Jimenez N, Cabrera-Bermon J, Diaz-Iglesias C, et al. Validation of the NASA-TLX Score in Ongoing Assessment of Mental Workload During a Laparoscopic Learning Curve in Bariatric Surgery. Obes Surg 2015 Dec;25(12):2451-2456. [doi: 10.1007/s11695-015-1922-1] [Medline: 26459432]
25. Tubbs-Cooley HL, Mara CA, Carle AC, Gurses AP. The NASA Task Load Index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. Intensive Crit Care Nurs 2018 Jun;46:64-69. [doi: 10.1016/j.iccn.2018.01.004] [Medline: 29449130]

26. Miyake S. [Mental Workload Assessment of Health Care Staff by NASA-TLX]. J UOEH 2020;42(1):63-75 [FREE Full text] [doi: 10.7888/juoeh.42.63] [Medline: 32213744]
27. Lowndes BR, Forsyth KL, Blocker RC, Dean PG, Truty MJ, Heller SF, et al. NASA-TLX Assessment of Surgeon Workload Variation Across Specialties. Ann Surg 2020 Apr;271(4):686-692. [doi: 10.1097/SLA.0000000000003058] [Medline: 30247331]
28. Pfarr J, Ganter MT, Spahn DR, Noethiger CB, Tscholl DW. Effects of a standardized distraction on caregivers' perceptive performance with avatar-based and conventional patient monitoring: a multicenter comparative study. J Clin Monit Comput 2019 Nov 25. [doi: 10.1007/s10877-019-00429-2] [Medline: 31768924]
29. Garot O, Rössler J, Pfarr J, Ganter MT, Spahn DR, Nöthiger CB, et al. Avatar-based versus conventional vital sign display in a central monitor for monitoring multiple patients: a multicenter computer-based laboratory study. BMC Med Inform Decis Mak 2020 Feb 10;20(1):26 [FREE Full text] [doi: 10.1186/s12911-020-1032-4] [Medline: 32041584]
30. Fink A. Survey Research Methods. In: Peterson P, Baker E, McGaw B, editors. International Encyclopedia of Education (Third Edition). Oxford: Elsevier; 2010:152-160.
31. Loft S, Sanderson P, Neal A, Mooij M. Modeling and predicting mental workload in en route air traffic control: critical review and broader implications. Hum Factors 2007 Jun;49(3):376-399. [doi: 10.1518/001872007X197017] [Medline: 17552304]
32. Mental AV, Pamela ST. Mental Workload and Situation Awareness. In: Handbook of Human Factors and Ergonomics. Hoboken, New Jersey: John Wiley & Sons, Inc; 2012.
33. Wheelock A, Suliman A, Wharton R, Babu ED, Hull L, Vincent C, et al. The Impact of Operating Room Distractions on Stress, Workload, and Teamwork. Ann Surg 2015 Jun;261(6):1079-1084. [doi: 10.1097/SLA.0000000000001051] [Medline: 26291954]
34. Binenbaum G, Musick DW, Ross HM. The development of physician confidence during surgical and medical internship. Am J Surg 2007 Jan;193(1):79-85. [doi: 10.1016/j.amjsurg.2006.07.009] [Medline: 17188093]
35. Endsley M, Jones W. Situation awareness information dominance & information warfare. In: Air Force Research Laboratory. Ohio: United States Airforce; 1997.
36. Burton J. WHO Healthy Workplace Framework and Model: Background and Supporting Literature and Practice. Geneva: World Health Organization; 2010.
37. Bowling NA, Alarcon GM, Bragg CB, Hartman MJ. A meta-analytic examination of the potential correlates and consequences of workload. Work & Stress 2015 Apr 17;29(2):95-113. [doi: 10.1080/02678373.2015.1033037]
38. de Beer LT, Pienaar J, Rothmann S. Work overload, burnout, and psychological ill-health symptoms: a three-wave mediation model of the employee health impairment process. Anxiety Stress Coping 2016 Jul;29(4):387-399. [doi: 10.1080/10615806.2015.1061123] [Medline: 26079200]
39. Carthon JM, Kutney-Lee A, Jarrín O, Sloane D, Aiken LH. Nurse staffing and postsurgical outcomes in black adults. J Am Geriatr Soc 2012 Jun;60(6):1078-1084 [FREE Full text] [doi: 10.1111/j.1532-5415.2012.03990.x] [Medline: 22690984]
40. Wallace DJ, Angus DC, Barnato AE, Kramer AA, Kahn JM. Nighttime intensivist staffing and mortality among critically ill patients. N Engl J Med 2012 May 31;366(22):2093-2101 [FREE Full text] [doi: 10.1056/NEJMsa1201918] [Medline: 22612639]
41. Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004 Oct 28;351(18):1838-1848. [doi: 10.1056/NEJMoa041406] [Medline: 15509817]
42. Grier RA, Warm JS, Dember WN, Matthews G, Galinsky TL, Parasuraman R. The vigilance decrement reflects limitations in effortful attention, not mindlessness. Hum Factors 2003;45(3):349-359. [doi: 10.1518/hfes.45.3.349.27253] [Medline: 14702988]
43. Broom MA, Capek AL, Carachi P, Akeroyd MA, Hilditch G. Critical phase distractions in anaesthesia and the sterile cockpit concept. Anaesthesia 2011 Mar;66(3):175-179 [FREE Full text] [doi: 10.1111/j.1365-2044.2011.06623.x] [Medline: 21320085]
44. Schulz CM, Burden A, Posner KL, Mincer SL, Steadman R, Wagner KJ, et al. Frequency and Type of Situational Awareness Errors Contributing to Death and Brain Damage: A Closed Claims Analysis. Anesthesiology 2017 Aug;127(2):326-337 [FREE Full text] [doi: 10.1097/ALN.0000000000001661] [Medline: 28459735]
45. Crockett CJ, Donahue BS, Vandivier DC. Distraction-Free Induction Zone: A Quality Improvement Initiative at a Large Academic Children's Hospital to Improve the Quality and Safety of Anesthetic Care for Our Patients. Anesth Analg 2019 Sep;129(3):794-803. [doi: 10.1213/ANE.0000000000003879] [Medline: 31425222]
46. Tscholl DW, Handschin L, Rössler J, Weiss M, Spahn DR, Nöthiger CB. It's not you, it's the design - common problems with patient monitoring reported by anesthesiologists: a mixed qualitative and quantitative study. BMC Anesthesiol 2019 May 28;19(1):87 [FREE Full text] [doi: 10.1186/s12871-019-0757-z] [Medline: 31138143]
47. McNeer RR, Bennett CL, Dudaryk R. Intraoperative Noise Increases Perceived Task Load and Fatigue in Anesthesiology Residents: A Simulation-Based Study. Anesth Analg 2016 Feb;122(2):512-525. [doi: 10.1213/ANE.0000000000001067] [Medline: 26797555]

48. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simul Healthc 2010 Oct;5(5):267-271. [doi: 10.1097/SIH.0b013e3181e3f329] [Medline: 21330808]
49. Hu JS, Lu J, Tan WB, Lomanto D. Training improves laparoscopic tasks performance and decreases operator workload. Surg Endosc 2016 May;30(5):1742-1746. [doi: 10.1007/s00464-015-4410-8] [Medline: 26173550]
50. Finomore VS, Shaw TH, Warm JS, Matthews G, Boles DB. Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. Hum Factors 2013 Dec;55(6):1044-1063. [doi: 10.1177/0018720813484498] [Medline: 24745198]
51. Rubio-Valdehita S, López-Núñez MI, López-Higes R, Díaz-Ramiro EM. Development of the CarMen-Q Questionnaire for mental workload assessment. Psicothema 2017 Nov;29(4):570-576. [doi: 10.7334/psicothema2017.151] [Medline: 29048320]
52. Zheng B, Jiang X, Tien G, Meneghetti A, Panton ON, Atkins MS. Workload assessment of surgeons: correlation between NASA TLX and blinks. Surg Endosc 2012 Oct;26(10):2746-2750. [doi: 10.1007/s00464-012-2268-6] [Medline: 22527300]
53. Marinescu AC, Sharples S, Ritchie AC, Sánchez López T, McDowell M, Morvan HP. Physiological Parameter Response to Variation of Mental Workload. Hum Factors 2018 Feb;60(1):31-56 [FREE Full text] [doi: 10.1177/0018720817733101] [Medline: 28965433]

Abbreviations

NASA-TLX: National Aeronautics and Space Administration-Task Load Index
ROTEM: rotational thromboelastometry

Edited by G Eysenbach; submitted 20.04.20; peer-reviewed by F Blazer, JJ Mira, AS Poncette, MA Bahrami; comments to author 11.05.20; revised version received 29.06.20; accepted 11.08.20; published 07.09.20.

Please cite as:
Said S, Gozdzik M, Roche TR, Braun J, Rössler J, Kaserer A, Spahn DR, Nöthiger CB, Tscholl DW
Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: Pooled Analysis Study Using Mixed Models
J Med Internet Res 2020;22(9):e19472
URL: http://www.jmir.org/2020/9/e19472/
doi:10.2196/19472
PMID:32780712

Original Paper

Automatic Grading of Stroke Symptoms for Rapid Assessment Using Optimized Machine Learning and 4-Limb Kinematics: Clinical Validation Study

Eunjeong Park[1], PhD; Kijeong Lee[2], MD; Taehwa Han[3], PhD; Hyo Suk Nam[4], MD, PhD

[1]Cerebro-Cardiovascular Disease Research Center, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Radiology, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Health-IT Center, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Department of Neurology, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Hyo Suk Nam, MD, PhD
Department of Neurology
Yonsei University College of Medicine
50 Yonsei-ro, Seodaemoon-gu
Seoul, 03722
Republic of Korea
Phone: 82 222280245
Email: hsnam@yuhs.ac

Abstract

Background: Subtle abnormal motor signs are indications of serious neurological diseases. Although neurological deficits require fast initiation of treatment in a restricted time, it is difficult for nonspecialists to detect and objectively assess the symptoms. In the clinical environment, diagnoses and decisions are based on clinical grading methods, including the National Institutes of Health Stroke Scale (NIHSS) score or the Medical Research Council (MRC) score, which have been used to measure motor weakness. Objective grading in various environments is necessitated for consistent agreement among patients, caregivers, paramedics, and medical staff to facilitate rapid diagnoses and dispatches to appropriate medical centers.

Objective: In this study, we aimed to develop an autonomous grading system for stroke patients. We investigated the feasibility of our new system to assess motor weakness and grade NIHSS and MRC scores of 4 limbs, similar to the clinical examinations performed by medical staff.

Methods: We implemented an automatic grading system composed of a measuring unit with wearable sensors and a grading unit with optimized machine learning. Inertial sensors were attached to measure subtle weaknesses caused by paralysis of upper and lower limbs. We collected 60 instances of data with kinematic features of motor disorders from neurological examination and demographic information of stroke patients with NIHSS 0 or 1 and MRC 7, 8, or 9 grades in a stroke unit. Training data with 240 instances were generated using a synthetic minority oversampling technique to complement the imbalanced number of data between classes and low number of training data. We trained 2 representative machine learning algorithms, an ensemble and a support vector machine (SVM), to implement auto-NIHSS and auto-MRC grading. The optimized algorithms performed a 5-fold cross-validation and were searched by Bayes optimization in 30 trials. The trained model was tested with the 60 original hold-out instances for performance evaluation in accuracy, sensitivity, specificity, and area under the receiver operating characteristics curve (AUC).

Results: The proposed system can grade NIHSS scores with an accuracy of 83.3% and an AUC of 0.912 using an optimized ensemble algorithm, and it can grade with an accuracy of 80.0% and an AUC of 0.860 using an optimized SVM algorithm. The auto-MRC grading achieved an accuracy of 76.7% and a mean AUC of 0.870 in SVM classification and an accuracy of 78.3% and a mean AUC of 0.877 in ensemble classification.

Conclusions: The automatic grading system quantifies proximal weakness in real time and assesses symptoms through automatic grading. The pilot outcomes demonstrated the feasibility of remote monitoring of motor weakness caused by stroke. The system can facilitate consistent grading with instant assessment and expedite dispatches to appropriate hospitals and treatment initiation by sharing auto-MRC and auto-NIHSS scores between prehospital and hospital responses as an objective observation.

(J Med Internet Res 2020;22(9):e20641) doi:10.2196/20641

KEYWORDS

machine learning; artificial intelligence; sensors; kinematics; stroke; telemedicine

Introduction

Motor weakness is a typical manifestation in various neurological disorders, including stroke, spinal cord injury, and traumatic brain injury. In addition, it is a major obstacle to functional recovery after the treatment of those diseases. As an example of motor weakness, unintentional drift is an indication of arm weakness; this is mainly caused by subtle damages in the motor pathway from the brain to the spinal cord [1]. If the supinator muscles in the upper limb are weaker than the pronator muscles in the presence of upper motor neuron lesion, the arm drifts downward and the palm turns toward the floor. The pathological response is for one of the arms to drift (up, down, or out). Therefore, motor weakness is a major sign in the FAST (face drooping, arm weakness, speech slurring, and time to call) protocol for stroke patients [2].

Rapid detection of such motor weakness is critical because acute treatments, including thrombolysis or thrombectomy, are performed in a constrained time window. More importantly, diagnosis can be established through bedside examination by specialists because it is a qualitative measurement. If the symptom occurs outside a hospital, a substantial time delay can cause poor outcomes for acute stroke patients [3-5]. In addition, the objective and accurate neurological assessments are not possible by mere visual examination because the examiner cannot easily trace the movement using the conventional neurological examination when there are subtle weaknesses. Therefore, systems need to automatically detect motor deficits using sensor data in real time.

However, operating such systems in a real environment requires a significant effort in integrating new systems into an emergency protocol. This is because interruptions caused by the attachment of sensors on patients' bodies and the initiation of the recording process can affect the streamlined structure of emergency protocols. However, evaluation methods are still required to identify stroke patients, as they can be instantly used in the communication among patients or caregivers, emergency call centers, and hospitals. In addition to a sensor-based measurement tool that was demonstrated useful in detecting subtle motor weakness in our previous study [6], the grading of stroke severity can be informed remotely and used in the emergency medical service (EMS) and hospital system.

In the field and in clinical environments, various grading methods exist for identifying ischemic stroke patients with motor weakness [7-10]. The National Institutes of Health Stroke Scale (NIHSS) score [11,12] and Medical Research Council (MRC) score [13,14] have been used as typical assessment indicators for stroke in the clinical environment. The rapid arterial occlusion evaluation scale, the Cincinnati stroke triage assessment tool, and the prehospital acute stroke severity scale are grading methods in the field environment. In this study, we implemented auto-NIHSS and auto-MRC systems to grade the NIHSS and modified MRC scores to assess patients in the clinical environment. We used subdivided MRC scores (10-grade MRC) instead of a 6-grade MRC to define subtle differences, as shown in Table 1.

Table 1. NIHSS and MRC grades for muscle power assessment.

Scale and grade	Description
NIHSS[a]	
0	No drift; limb holds 90° (or 45°) angle for full 10 seconds
1	Drift; limb holds 90° (or 45°) angle, but drifts down before full 10 seconds; does not hit bed or other support
2	Some effort against gravity; limb cannot reach or maintain (if cued) 90° (or 45°) angle; drifts down to bed, but has some effort against gravity
3	No effort against gravity; limb falls
4	No movement
MRC[b]	
0 (0)	No movement
1 (1)	A flicker of movement is observed or felt in the muscle
2 (1+)	Muscle moves the joint when gravity is eliminated
3 (2)	Muscle moves the joint against gravity, but not through full mechanical range of motion
4 (2+)	Muscle cannot hold the joint against resistance, but moves the joint fully against gravity
5 (3)	Muscle moves the joint fully against gravity and is capable of transient resistance, but collapses abruptly
6 (3+)	Same as grade 4 (on 6-point scale) but muscle holds the joint only against minimal resistance
7 (4)	Muscle holds the joint against a combination of gravity and moderate resistance
8 (4+)	Same as grade 4 (on 6-point scale) but muscle holds the joint against moderate to maximal resistance
9 (5)	Normal strength

[a]NIHSS: National Institutes of Health Stroke Scale.

[b]MRC: Medical Research Council.

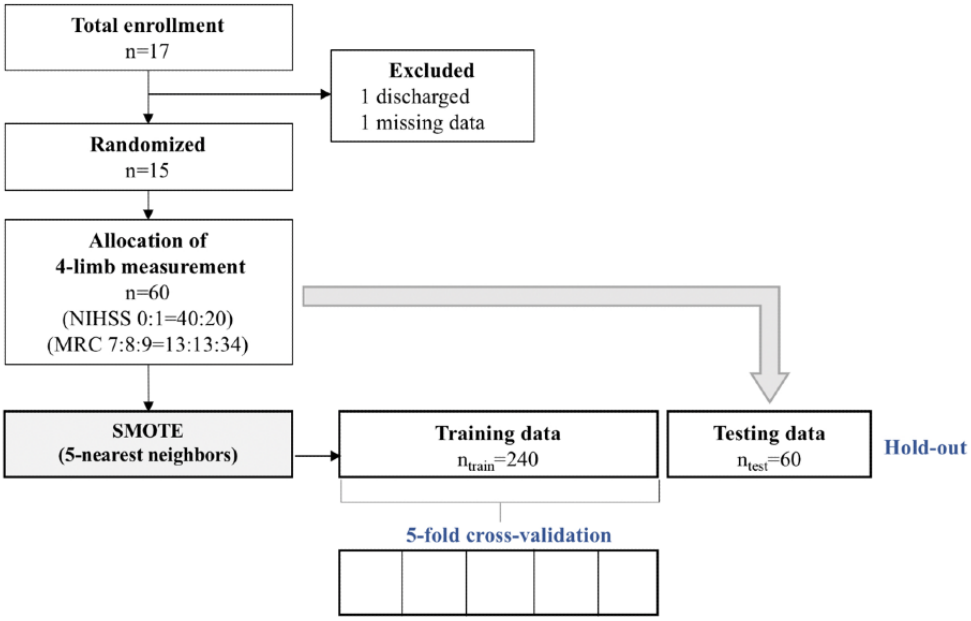
Methods

Participants and Data

A total of 17 participants were recruited; 15 participants (10 male and 5 female participants) were finally enrolled and completed 4-limb drift test trials. To estimate the scores of patients with severity, we performed the assessment shortly after admission to a stroke unit. The ages of the participants ranged from 44 to 92 years, with a mean of 68.6 years (SD 16.11). Exclusion criteria were patients (1) who had a substantial weakness that prevented arm or leg raising against gravity, (2) who were not able to sit and who had bilateral arm weakness or preexisting chronic arm weakness, and (3) who had aphasia, neglect, peripheral neuropathy, myopathy, or joint deformity. This study was approved by the Severance Hospital Institutional Review Board, and informed consent was obtained from all participants.

Figure 1 shows patient enrollment and data preparation for auto-NIHSS and auto-MRC grading. Description of data composition for training, validation and testing is detailed in the section on system design.

Figure 1. Patient enrollment and data set for automatic grading system. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale; SMOTE: synthetic minority oversampling technique.



System Design

The entire process of the system is shown in Figure 2. The system is composed of 2 parts, the measurement and the grading units. The measurement unit sets up sensors and Bluetooth connection with the primary information of patients.

We measured the upper left and upper right limb movements using sensors on both wrists of patients, who were asked to stretch and hold their arms for 20 seconds, as shown in Figure 3. For the lower left and lower right limb drift tests, patients were asked to lift and stretch their left or right leg for 20 seconds.

Figure 2. Automatic grading process. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale.

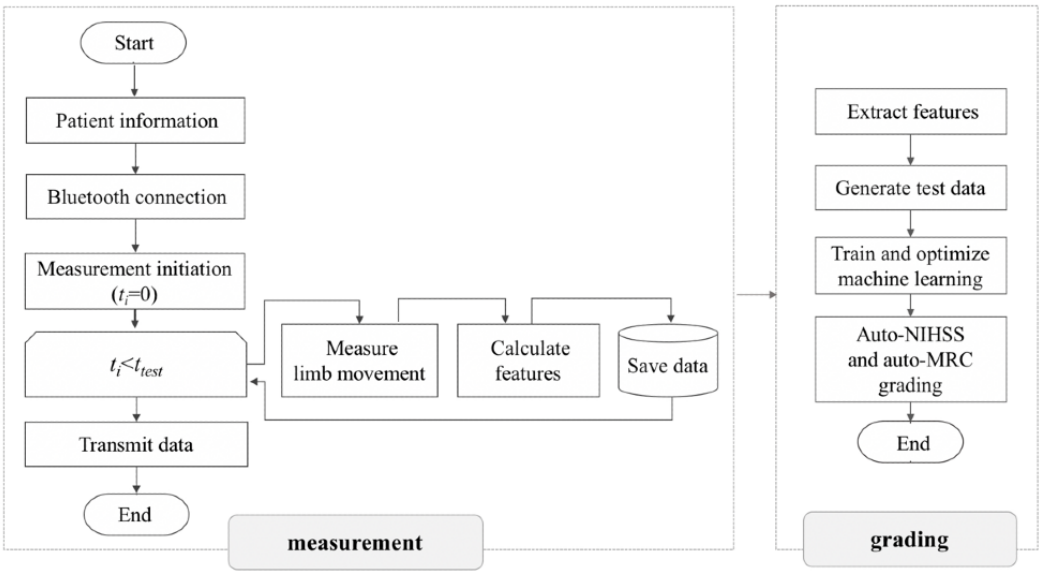


Figure 3. Schematic of upper and lower limb sensors and corresponding segment axes.

The pseudo-code of the measurement unit is shown in Multimedia Appendix 1. For each time frame *i*, the rotational transformation from the limb into the reference frame *xyz* is denoted as . The corresponding rotation matrices *R* for each angle are defined using the of the accelerometer signals for the *i*th frame. Subsequently, the degree of drift, θ_{drift}, is calculated and used in key features of machine learning classification.

After collecting the series of 4-limb movements during the test time, the grading unit analyzes the kinematic features. Subsequently, the machine learning algorithm is trained to estimate the NIHSS and MRC scores of each limb. Algorithm 2 (in Multimedia Appendix 2) shows the process of feature extraction, data generation, and model training for the optimized classification of auto-NIHSS and auto-MRC.

In the feature extraction process, features as predictors of limb paralysis were extracted using a series of measured data. In this study, the duration of the drift test (t_{test}) was set to 20 seconds; however, analysis started 10 seconds after the examination started (t_{start}) to exclude the initial dip. The average, maximum, and oscillation of drift caused by paralysis for each limb and demographic features were fed to train the machine learning algorithms.

In the data generation process, we adopted the synthetic minority oversampling technique (SMOTE) [15], leveraging the K-nearest neighbor (K-NN), to solve the imbalanced problem that is typical in machine learning studies in medicine [16-18]. The SMOTE with K-NN generated n_g samples for each grade. Therefore, $n_g c$ records were used to construct a grading model with c classes. In this study, n_g was set to 120 for auto-NIHSS (c=2) and 80 for auto-MRC (c=3) to compose the training data with 240 (t_{train}) instances. Apart from the training data, the original data set with 60 records remained for the test data, as shown in Figure 1.

In the training process, 5-fold cross-validation was applied to reduce overfitting and generalize the model [19]. In the optimization process, the fitted support vector machine (SVM), as well as ensemble models among various SVM kernels and boosting algorithms with tuned hyperparameters, were searched via Bayes optimization in 30 trials for each model [20]. The grading models were implemented and evaluated in MATLAB R2020a (MathWorks Inc) [21].

Results

Sensor Data Characteristics

The system measured the drift of 4 limbs and extracted the kinematic features, as shown in Multimedia Appendix 3. The characteristics of the patients and test data are summarized in Table 2. The grade distribution of clinical scores was not regularized between limbs, as shown in Figure 4. For example, the upper left MRC group had 10 patients graded as MRC 9, 2 patients graded as MRC 8, and 3 patients graded as MRC 7. Among 13 MRC 8 instances, 7 were evaluated as NIHSS 1, whereas 6 were evaluated NIHSS 0. We constructed auto-MRC, which discriminated instances of grades with a data ratio of 13:13:34, whereas auto-NIHSS performed binary classification with a data ratio of 40:20.

Figure 4. Grade distribution of NIHSS and MRC. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale.

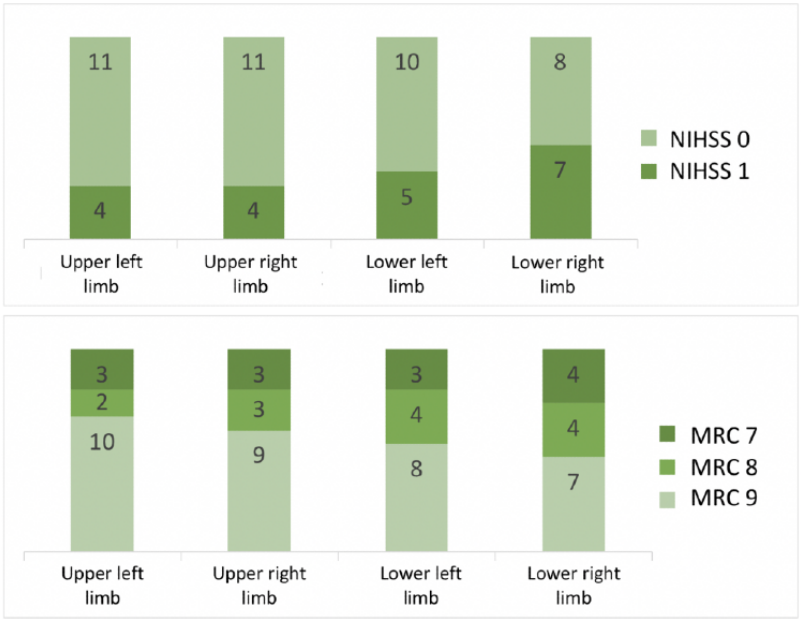


Table 2. Summary of patients and test data.

Diagnosis	Measurement												NIHSS[a] grade (MRC[b] grade)			
	ULL[c]			URL[d]			LLL[e]			LRL[f]			ULL	URL	LLL	LRL
	Mean	Max[g]	Osc[h]	Mean	Max	Osc	Mean	Max	Osc	Mean	Max	Osc				
Lt[i] internal capsule infarction	0.82	2.7	14.4	–3	–1.9	15.3	–1.19	2.1	25.1	11.81	17.4	30.8	0 (9)	0 (8)	0 (9)	1 (8)
Lt MCA[j] infarction	–9.33	–12	15.9	–6.47	–9.2	11.7	18.7	10.7	43.5	30.26	27.7	13.6	0 (8)	1 (7)	0 (8)	1 (7)
Lt MCA infarction	0.86	0	7	4.06	1.4	13.8	2.96	0	24.9	9.91	–1.6	61.5	0 (9)	0 (9)	0 (9)	1 (7)
Lt MCA infarction	3.16	4.2	14.5	2.3	3.2	19.1	0.26	1.6	12.9	4.26	8.4	21.7	0 (9)	0 (9)	0 (9)	0 (8)
Lt MCA infarction	1.92	3.6	14.6	3.14	4.2	14.5	1.84	5.7	39.5	0.75	2.9	19.6	0 (9)	0 (9)	0 (9)	0 (9)
Lt pontine infarction	–0.67	0.6	19.5	–1.37	1.3	12.9	–11.93	–10.3	16.7	–4.93	–2.1	17.3	1 (7)	0 (8)	1 (7)	1 (8)
Lt thalamic infarction	2.05	3.5	22.8	8.91	11.4	12.5	4.77	8.8	31.3	1.98	6.8	37.7	0 (9)	1 (7)	0 (9)	1 (8)
Pontine ICH[k]	–1.57	1.5	39.1	0.81	2	18.5	–3	1.2	40	3.18	5.3	16.5	1 (7)	0 (9)	1 (8)	0 (9)
Rt[l] MCA infarction	–9.96	–7.5	17.9	–1.93	–0.6	19	–2.71	0.4	18.5	–1.99	–0.3	17.2	1 (7)	0 (9)	1 (7)	0 (9)
Lt internal capsule infarction	–6	–7.9	14	–0.8	–2	11.6	1.8	0.8	18.6	11	6.5	38.5	0 (9)	0 (9)	0 (8)	0 (9)
Myelitis (no weakness)	1.3	2.9	18.6	–0.56	0.1	11.7	–1.23	1.2	24.1	–1.14	0.7	24	0 (9)	0 (9)	0 (9)	0 (9)
Rt MCA infarction	–4.97	–6.4	19.2	0.7	0	13.1	13.9	7	49.3	6.31	2.3	34.3	0 (9)	0 (9)	1 (7)	0 (9)
Myasthenia gravis	–0.64	1.3	19.2	1.1	2.7	14.4	–1.97	0	18.5	–0.64	2.7	22.6	0 (9)	0 (9)	0 (9)	0 (9)
Lt pontine infarction	15.5	5.4	41.1	23.5	12	54	6.3	2.2	26.1	5.3	0.6	46.2	0 (9)	1 (7)	0 (9)	1 (7)
Pontine hemorrhage	–0.83	1.1	19	–2.72	1.3	26.6	1.69	3.3	13.6	–7.52	–0.8	54.5	1 (8)	1 (8)	1 (8)	1 (7)

[a]NIHSS: National Institutes of Health Stroke Scale.
[b]MRC: Medical Research Council.
[c]ULL: upper left limb.
[d]URL: upper right limb.
[e]LLL: lower left limb.
[f]LRL: lower right limb.
[g]Max: maximum.
[h]Osc: oscillation.
[i]Lt: left.
[j]MCA: middle cerebral artery.
[k]ICH: intracerebral hemorrhage.
[l]Rt: right.

Evaluation Outcomes

We evaluated the performance of the system in terms of the accuracy, sensitivity, specificity, precision, F1 score, and area under the receiver operating characteristics curve (AUC) with a confusion matrix.

The statistical plots in Figure 5 show the patterns of the average, maximum, and oscillation of the 4-limb features of each NIHSS grade. Auto-NIHSS discriminated those features, as shown in the confusion matrices in Figure 6. The result shows that the proposed autonomous grading achieved an accuracy of at least 80% and that the overall accuracy was 81.7%, as shown in the summary of performance in Table 3. The AUC of auto-NIHSS reached 0.912, as depicted in the receiver operating characteristics curves in Figure 6. The sensitivity of the NIHSS grading reached 0.825 with the SVM and 0.875 with the ensemble. The specificity was 0.750 for both models.

Auto-MRC discriminates instances into 3 MRC grades, and the statistical plots of movement features are depicted in Figure 7. The mean AUC was 0.870 for the SVM and 0.877 for the ensemble, as shown in Figure 8. Table 4 shows the summarized performance of auto-MRC; the average accuracy, sensitivity, and specificity for the MRC grading were 0.775, 0.717, and 0.876, respectively.

Figure 5. Statistical plots of 4-limb features of NIHSS grades. NIHSS: National Institutes of Health Stroke Scale.

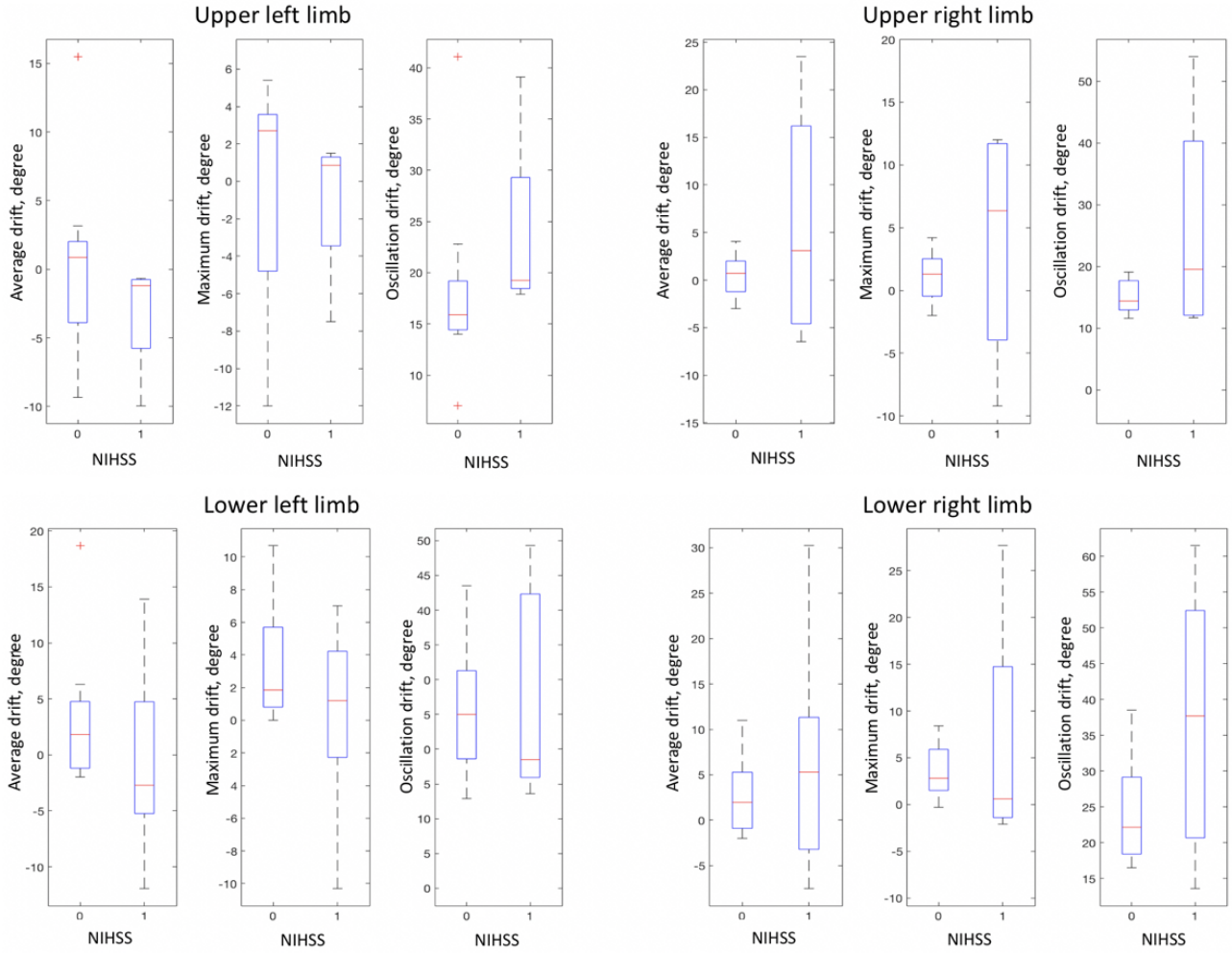


Figure 6. Confusion matrix and receiver operating characteristic of auto-NIHSS grading using (A) support vector machine and (B) ensemble learning. AUC: area under the receiver operating characteristics curve; NIHSS: National Institutes of Health Stroke Scale.

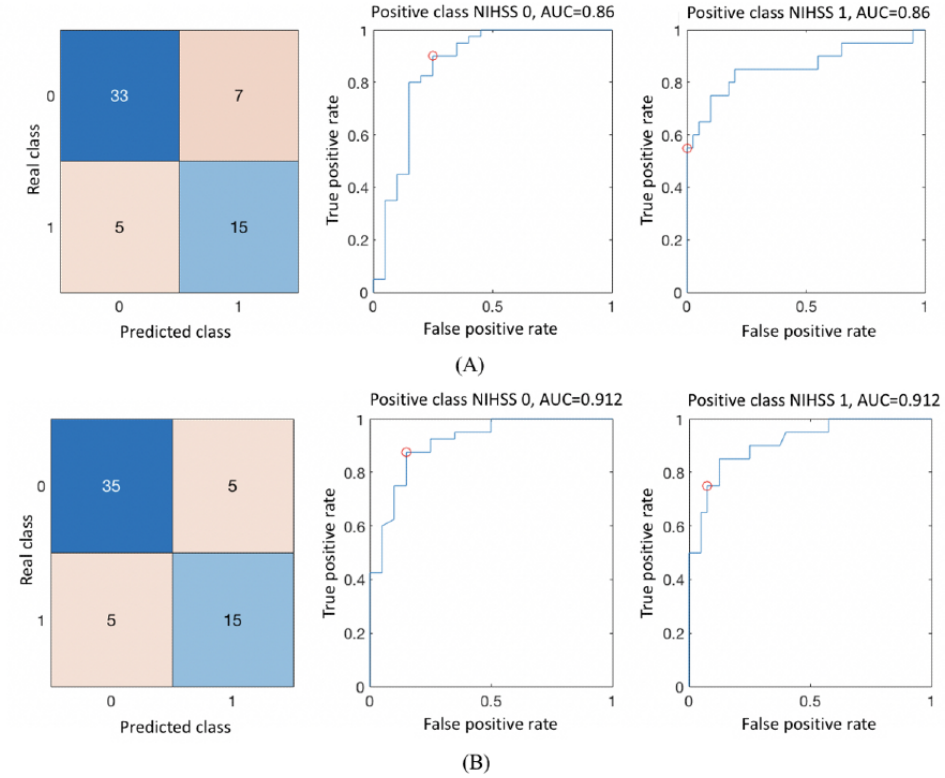


Figure 7. Statistical plots of 4-limb features of MRC grades. MRC: Medical Research Council.

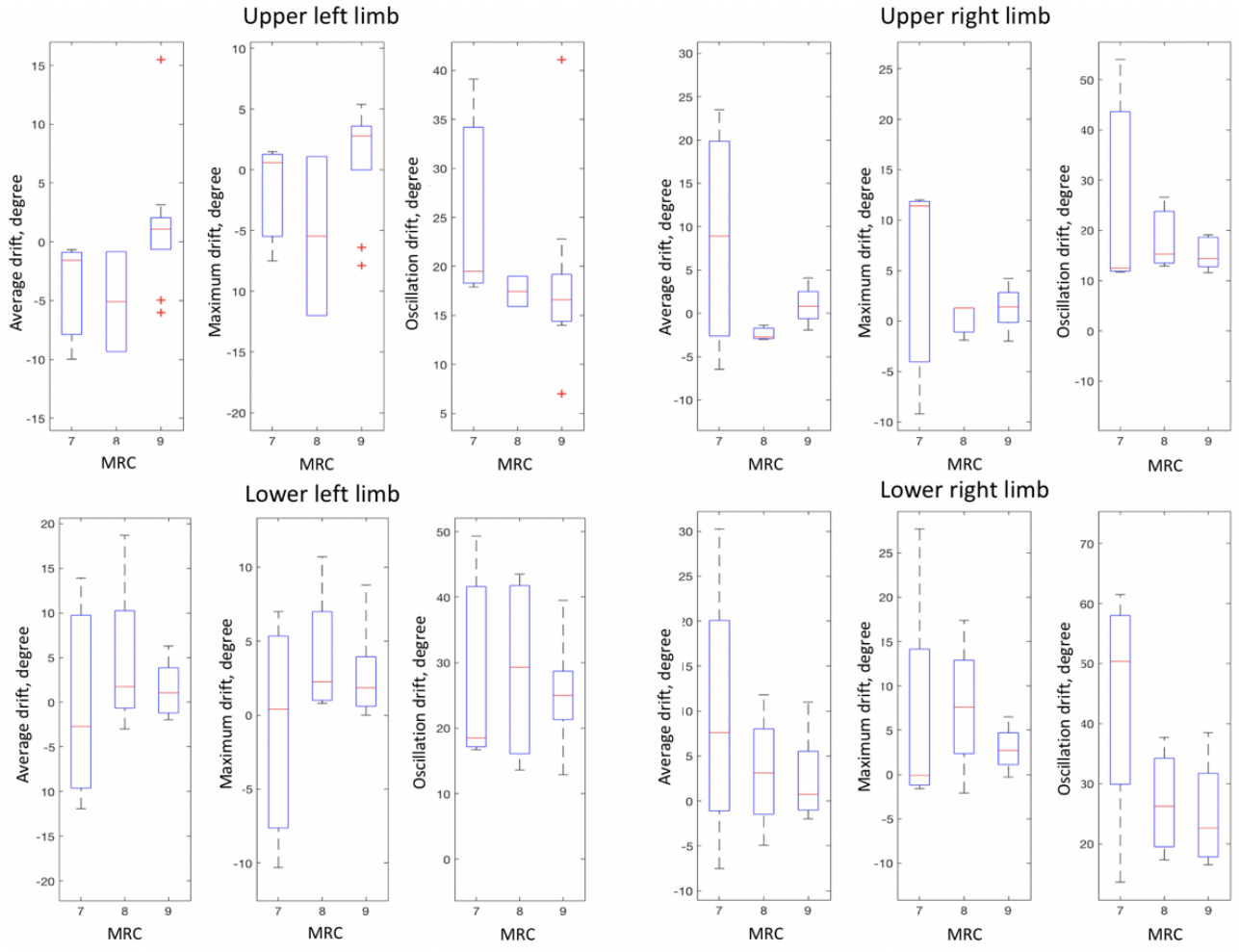


Figure 8. Confusion matrix and receiver operating characteristic of auto-MRC grading using (A) support vector machine and (B) ensemble learning. AUC: area under the receiver operating characteristics curve; MRC: Medical Research Council.

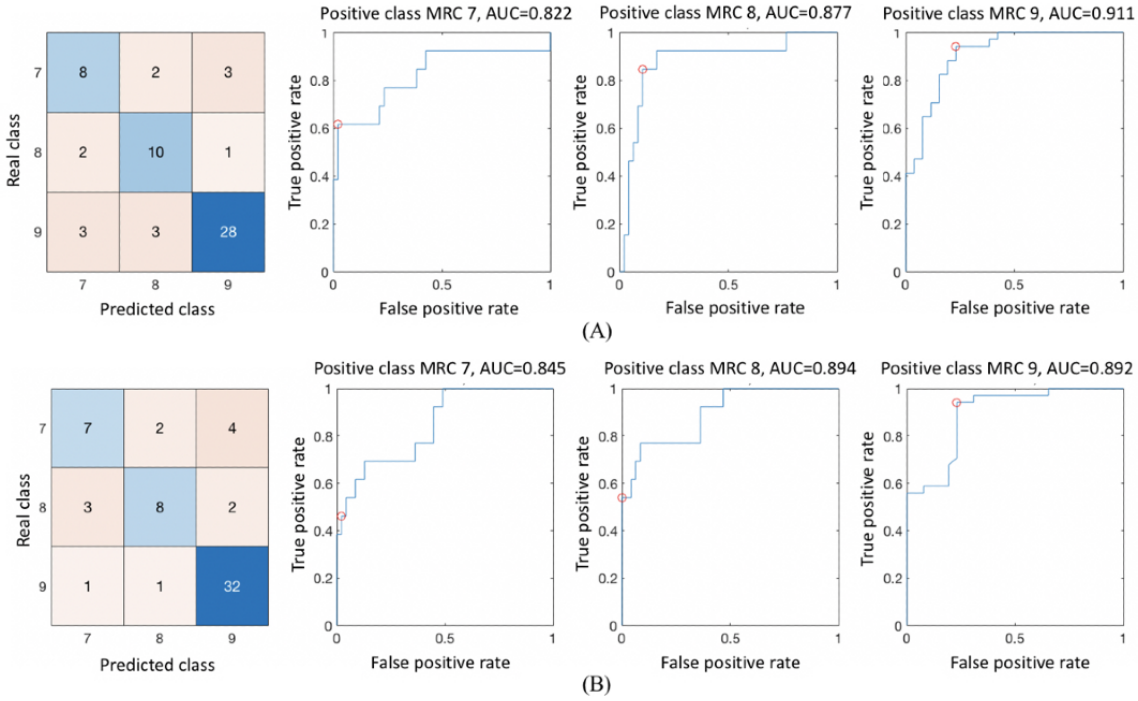


Table 3. Performance of auto-NIHSS grading.

Auto-NIHSS[a] grading	Accuracy	Sensitivity	Specificity	Precision	F1 score
SVM[b]	0.800	0.825	0.750	0.868	0.846
Ensemble	0.833	0.875	0.750	0.875	0.875

[a]NIHSS: National Institutes of Health Stroke Scale.

[b]SVM: support vector machine.

Table 4. Performance of auto-MRC grading.

Auto-MRC[a] grading	Accuracy	Sensitivity	Specificity	Precision	F1 score
SVM[b]	0.767	0.736	0.878	0.719	0.726
Ensemble	0.783	0.698	0.873	0.735	0.713

[a]MRC: Medical Research Council.

[b]SVM: support vector machine.

Discussion

Importance of Objective and Fast Assessment of Stroke Severity

The notion “time is brain” is valid in treating stroke patients. Intravenous tissue plasminogen activator (IV tPA) within 4.5 hours of stroke onset is the only therapy for acute ischemic stroke [22]. Subsequently, endovascular thrombectomy (EVT) has been a standard of care for patients with acute ischemic stroke caused by large artery occlusion within 6 to 24 hours of onset, based on successful large randomized clinical trials [23]. Reperfusion therapy, including IV tPA and EVT, for acute ischemic stroke is time sensitive (ie, an earlier treatment yields a better outcome). As the onset-to-intervention time is composed of prehospital and in-hospital phases, patients who arrive early have more chances of appropriate treatment [24-27]. Delays in hospital admission and the preparation before treatment affect the prognosis of patients [28]. In Goyal et al [24], the authors reported that the most significant issue was getting the correct patient to the correct hospital quickly. In Sukumaran et al [27], strategies for stroke patient workflow optimization were suggested by analyzing and solving prehospital and preprocedural bottlenecks. The interhospital transfer is directly associated with delays in onset to reperfusion time, which results in the poor outcome of stroke patients; therefore, the timely triage of patients is a significant bottleneck [27].

The importance of accurate and objective assessments of stroke severity in telemedicine and telestroke strategies has been discussed in numerous studies [29]. In particular, the timing constraint in performing reperfusion therapy, which has been shown to significantly reduce mortality, invokes the

development of efficient systems and protocols in prehospital care or emergency medical systems. Researchers have addressed the fact that the rapid and accurate evaluation of stroke severity can aid in identifying patients for treatments and accelerate an urgent streamlined process. In the study by Andsberg et al [30], a prehospital ambulance stroke test was performed to score the severity of stroke through commands, answers, and observations. The remote assessment of stroke using smartphones was proposed and compared with bedside examination in calculating the NIHSS score [31]. However, most assessments in those systems used conservative observation or campaigns that were subjective and unreliable between testers. Modern communication, sensor technology, and machine learning can solve this problem through accurate measurements and the fast determination of assessment in a prehospital or remote environment [29,32,33]. A previous study evaluated arm function in activities using kinematic exposure variation analysis and inertial sensors [34]. A mobile-based walk test was developed to report patients' walking ability [35], and upper limb impairments in stroke patients were measured using inertial sensors in the home environment [33]. Such sensor-based testing enables objective evaluation regardless of the testers or place.

Utility of Consistent Grading Method as an Agreement Between Prehospital and Hospital Environment

The necessity of a controlled test is revealed in the results of previous studies for monitoring daily living. Motor recovery was monitored using accelerometers, and the NIHSS motor index was estimated in the study by Gubbi et al [36]. However, the movement in daily living limited the accuracy of estimation to 56% for the low index. Activity monitoring in most sensor-based studies involved trials that were not approved by clinical protocols. Those systems limited extensibility as a standard of remote monitoring systems, although they were efficient in tracking the progress or the treatment outcome.

In addition to rapid and accurate measurements, we aimed to increase the utility of the assessment system in the prehospital and hospital environments. At every phase of the prehospital process, consistent methods to conduct assessments can reduce errors and delays in communication among the participants of a community's emergency group. Therefore, automatic scoring can facilitate agreement in assessments among patients, caregivers, paramedics, and medical staff. With regard to bottleneck analysis in acute stroke treatment, the rapid identification of neurological deficits and assessment of motor grading will aid EMS personnel in transporting patients to a comprehensive stroke center because hospitals may be limited in terms of stroke unit availability and resources. In Berglund et al [26], the importance of stroke identification without meeting the patient or without neurological examination was asserted; the time to treatment can be decreased with the high-priority dispatch of ambulances through early identification of stroke from emergency calls. In the hyper acute stroke alarm study [25], researchers observed that higher prehospital priority levels of stroke improved thrombolysis frequency and time to stroke unit. The stroke identification by EMS dispatchers during emergency calls varied between 31% and 57%, as identifying stroke can be a challenge without examination [26].

Therefore, we developed an automatic grading system, leveraging multiclassification of machine learning using typically performed tests and grading in clinics. Our proposed solution uses controlled observations of drift tests in clinics and can estimate the assessment by neurologists. Therefore, the scores by the automatic grading system can be instantly used for communication in an objective manner.

Data and Techniques for Clinical Scoring by Machine Learning

A considerable number of studies have used artificial intelligence, including machine learning, to estimate clinical scores and assess patients or provide warnings regarding adverse events [37-40]. In those studies, a series of various techniques were used according to the scale of scores, the capacity of collected data, and the skewness of data. Following the significant development of enhanced algorithms, data with significant meaning have gained importance. However, as addressed in Li et al [41], real-world data have a long-tail pattern with a significant imbalance in quality and quantity. Many algorithms have used public big data to develop new algorithms and build models; however, real-world applications have completely different data quality and quantity and cannot directly apply those models. This situation is particularly severe in medicine, as discussed in Hulsen et al [42]. The availability of qualified data differs by disease, severity of disease in patients, and difficulty of collection [43]. Big data from electronic medical records that are already facilitated in hospital information systems can be used in comparatively easy tasks for medical artificial intelligence. The recent success of medical artificial intelligence requires significant effort and cost in collecting and labeling data [44,45]. In addition, machine learning for sporadic events in emergencies or patients with rare diseases is affected by data deficiency. This is because interventions for collecting data can affect the prognosis of treatment due to the possible delay in the rapid streamlining of treatment processes. Previous feasibility studies have stated that the difficulty in real-time capturing of acute neurological disorders was the main limitation in the research [33,46].

The learning models with imbalanced data were affected by low precision or recall in the validation and test phases, although they achieved high accuracy for a large number of data in the majority groups [47]. Recently, techniques to solve this data skewness, including data augmentation, transfer learning, and deep imbalanced learning, were emphasized [48-51]. Studies on deep learning that extract filtered features derived from raw data have attempted to solve the problem by knowledge transfer from pretrained models [52,53] or with data augmentation [54,55]. Machine learning with records can cope with the imbalance problem through sampling, cost-sensitive learning, boosting algorithms, and skew-related performance metrics [47,56]. We used the SMOTE to balance between classes in the training phase and applied techniques, including RUSBoost, in optimized ensemble machine learning. To compare different models according to their precision on each class, the F measure is typically used as a performance metric [57]; additionally, we validated the performance of the proposed solution using the AUC and F1 scores. Consequently, the performances of auto-NIHSS and auto-MRC indicated the acceptable AUC,

sensitivity, specificity, and F1 score as real-world applications with data skewness.

Conclusion

Accurate monitoring and grading of motor weakness are critical for the appropriate assessment of stroke severity, particularly for reliable and consistent evaluations. We developed an automatic grading system to assess proximal motor weakness using the kinematic features of unintended drift of 4 limbs. We trained optimized machine learning models and obtained promising results in scoring NIHSS and MRC. The objective scoring of neurological deficits can be used to identify stroke patients, dispatch patients to the appropriate medical center, and expedite treatment preparation.

Acknowledgments

This research was supported by a grant funded by the Ministry of Science and ICT (NRF-2020R1A2C1013152) and by a grant of the Korea Health Technology Research and Development Project funded by the Ministry of Health and Welfare (HI19C0481, HC19C0028), Republic of Korea.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Algorithm of measurement unit for automatic grading.
[PNG File , 699 KB - jmir_v22i9e20641_app1.png]

Multimedia Appendix 2
Algorithm of the grading unit for extracting features and training machine learning algorithms. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale; SMOTE: synthetic minority oversampling technique.
[PNG File , 1029 KB - jmir_v22i9e20641_app2.png]

Multimedia Appendix 3
Sample measurement of unintended drift of limbs.
[PNG File , 1259 KB - jmir_v22i9e20641_app3.png]

References

1. Darcy P, Moughty AM. Pronator Drift. N Engl J Med 2013 Oct 17;369(16):e20. [doi: 10.1056/nejmicm1213343]
2. Nor AM, McAllister C, Louw S, Dyker A, Davis M, Jenkinson D, et al. Agreement Between Ambulance Paramedic- and Physician-Recorded Neurological Signs With Face Arm Speech Test (FAST) in Acute Stroke Patients. Stroke 2004 Jun;35(6):1355-1359. [doi: 10.1161/01.str.0000128529.63156.c5]
3. Kim J, Fonarow GC, Smith EE, Reeves MJ, Navalkele DD, Grotta JC, et al. Treatment With Tissue Plasminogen Activator in the Golden Hour and the Shape of the 4.5-Hour Time-Benefit Curve in the National United States Get With The Guidelines-Stroke Population. Circulation 2017 Jan 10;135(2):128-139. [doi: 10.1161/circulationaha.116.023336]
4. Kodankandath TV, Wright P, Power PM, De Geronimo M, Libman RB, Kwiatkowski T, et al. Improving Transfer Times for Acute Ischemic Stroke Patients to a Comprehensive Stroke Center. J Stroke Cerebrovasc Dis 2017 Jan;26(1):192-195. [doi: 10.1016/j.jstrokecerebrovasdis.2016.09.008] [Medline: 27743926]
5. Threlkeld ZD, Kozak B, McCoy D, Cole S, Martin C, Singh V. Collaborative Interventions Reduce Time-to-Thrombolysis for Acute Ischemic Stroke in a Public Safety Net Hospital. J Stroke Cerebrovasc Dis 2017 Jul;26(7):1500-1505. [doi: 10.1016/j.jstrokecerebrovasdis.2017.03.004] [Medline: 28396187]
6. Park E, Chang H, Nam HS. Use of Machine Learning Classifiers and Sensor Data to Detect Neurological Deficit in Stroke Patients. J Med Internet Res 2017 Apr 18;19(4):e120 [FREE Full text] [doi: 10.2196/jmir.7092] [Medline: 28420599]
7. Singer OC, Dvorak F, du Mesnil de Rochemont R, Lanfermann H, Sitzer M, Neumann-Haefelin T. A Simple 3-Item Stroke Scale. Stroke 2005 Apr;36(4):773-776. [doi: 10.1161/01.str.0000157591.61322.df]
8. Pérez de la Ossa N, Carrera D, Gorchs M, Querol M, Millán M, Gomis M, et al. Design and Validation of a Prehospital Stroke Scale to Predict Large Arterial Occlusion. Stroke 2014 Jan;45(1):87-91. [doi: 10.1161/strokeaha.113.003071]
9. Katz BS, McMullan JT, Sucharew H, Adeoye O, Broderick JP. Design and Validation of a Prehospital Scale to Predict Stroke Severity. Stroke 2015 Jun;46(6):1508-1512. [doi: 10.1161/strokeaha.115.008804]
10. Hastrup S, Damgaard D, Johnsen SP, Andersen G. Prehospital Acute Stroke Severity Scale to Predict Large Artery Occlusion. Stroke 2016 Jul;47(7):1772-1776. [doi: 10.1161/strokeaha.115.012482]
11. Heldner MR, Jung S, Zubler C, Mordasini P, Weck A, Mono M, et al. Outcome of patients with occlusions of the internal carotid artery or the main stem of the middle cerebral artery with NIHSS score of less than 5: comparison between

thrombolysed and non-thrombolysed patients. J Neurol Neurosurg Psychiatry 2015 Jul;86(7):755-760. [doi: 10.1136/jnnp-2014-308401] [Medline: 25266203]
12. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. Stroke 2000 May;31(4):858-862. [doi: 10.1161/01.str.31.4.858] [Medline: 10753988]
13. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax 1999 Jul;54(7):581-586 [FREE Full text] [doi: 10.1136/thx.54.7.581] [Medline: 10377201]
14. Paternostro-Sluga T, Grim-Stieger M, Posch M, Schuhfried O, Vacariu G, Mittermaier C, et al. Reliability and validity of the Medical Research Council (MRC) scale and a modified scale for testing muscle strength in patients with radial palsy. J Rehabil Med 2008 Aug;40(8):665-671 [FREE Full text] [doi: 10.2340/16501977-0235] [Medline: 19020701]
15. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
16. Verbiest N, Ramentol E, Cornelis C, Herrera F. Preprocessing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection. Applied Soft Computing 2014 Sep;22:511-517. [doi: 10.1016/j.asoc.2014.05.023]
17. Yijing L, Haixiang G, Xiao L, Yanan L, Jinling L. Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. Knowledge-Based Systems 2016 Feb;94:88-104. [doi: 10.1016/j.knosys.2015.11.013]
18. Douzas G, Bacao F, Last F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Information Sciences 2018 Oct;465:1-20. [doi: 10.1016/j.ins.2018.06.056]
19. Burman P. A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. Biometrika 1989 Sep;76(3):503. [doi: 10.2307/2336116]
20. Snoek J, Larochelle H, Adams R. Practical Bayesian optimization of machine learning algorithms. 2012 Presented at: Annual Advances in Neural Information Processing Systems 2012; Dec 3-6, 2012; Lake Tahoe, NV URL: https://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf
21. MATLAB R2020a. URL: https://mathworks.com [accessed 2020-08-25]
22. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. Thrombolysis with Alteplase 3 to 4.5 Hours after Acute Ischemic Stroke. N Engl J Med 2008 Sep 25;359(13):1317-1329. [doi: 10.1056/nejmoa0804656]
23. Goyal M, Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. The Lancet 2016 Apr;387(10029):1723-1731. [doi: 10.1016/s0140-6736(16)00163-x]
24. Goyal M, Jadhav AP, Wilson AT, Nogueira RG, Menon BK. Shifting bottlenecks in acute stroke treatment. J Neurointerv Surg 2016 Dec;8(11):1099-1100. [doi: 10.1136/neurintsurg-2015-012151] [Medline: 26674088]
25. Berglund A, Svensson L, Sjöstrand C, von Arbin M, von Euler M, Wahlgren N, et al. Higher Prehospital Priority Level of Stroke Improves Thrombolysis Frequency and Time to Stroke Unit. Stroke 2012 Oct;43(10):2666-2670. [doi: 10.1161/strokeaha.112.652644]
26. Berglund A, von Euler M, Schenck-Gustafsson K, Castrén M, Bohm K. Identification of stroke during the emergency call: a descriptive study of callers' presentation of stroke. BMJ Open 2015 May 28;5(4):e007661 [FREE Full text] [doi: 10.1136/bmjopen-2015-007661] [Medline: 25922106]
27. Sukumaran M, Cantrell D, Ansari S, Huryley M, Shaibani A, Potts M. Stroke patient workflow optimization. Endovasc Today 2019 Feb;18(2):46-50 [FREE Full text]
28. Itrat A, Taqui A, Cerejo R, Briggs F, Cho S, Organek N, Cleveland Pre-Hospital Acute Stroke Treatment Group. Telemedicine in Prehospital Stroke Evaluation and Thrombolysis: Taking Stroke Treatment to the Doorstep. JAMA Neurol 2016 Mar;73(2):162-168. [doi: 10.1001/jamaneurol.2015.3849] [Medline: 26641366]
29. Hess DC, Audebert HJ. The history and future of telestroke. Nat Rev Neurol 2013 Jul;9(6):340-350. [doi: 10.1038/nrneurol.2013.86] [Medline: 23649102]
30. Andsberg G, Esbjörnsson M, Olofsson A, Lindgren A, Norrving B, von Euler M. PreHospital Ambulance Stroke Test - pilot study of a novel stroke test. Scand J Trauma Resusc Emerg Med 2017 May 11;25(1):37 [FREE Full text] [doi: 10.1186/s13049-017-0377-x] [Medline: 28399897]
31. Anderson ER, Smith B, Ido M, Frankel M. Remote assessment of stroke using the iPhone 4. J Stroke Cerebrovasc Dis 2013 May;22(4):340-344. [doi: 10.1016/j.jstrokecerebrovasdis.2011.09.013] [Medline: 22018507]
32. Alber M, Buganza Tepole A, Cannon WR, De S, Dura-Bernal S, Garikipati K, et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit Med 2019;2:115 [FREE Full text] [doi: 10.1038/s41746-019-0193-y] [Medline: 31799423]
33. Held JPO, Klaassen B, Eenhoorn A, van Beijnum BJF, Buurke JH, Veltink PH, et al. Inertial Sensor Measurements of Upper-Limb Kinematics in Stroke Patients in Clinic and Home Environment. Front Bioeng Biotechnol 2018;6:27 [FREE Full text] [doi: 10.3389/fbioe.2018.00027] [Medline: 29707537]
34. Ertzgaard P, Öhberg F, Gerdle B, Grip H. A new way of assessing arm function in activity using kinematic Exposure Variation Analysis and portable inertial sensors--A validity study. Man Ther 2016 Mar;21:241-249. [doi: 10.1016/j.math.2015.09.004] [Medline: 26456185]

35. Salvi D, Poffley E, Orchard E, Tarassenko L. The Mobile-Based 6-Minute Walk Test: Usability Study and Algorithm Development and Validation. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13756 [FREE Full text] [doi: 10.2196/13756] [Medline: 31899457]
36. Gubbi J, Rao AS, Fang K, Yan B, Palaniswami M. Motor recovery monitoring using acceleration measurements in post acute stroke patients. BioMed Eng OnLine 2013;12(1):33. [doi: 10.1186/1475-925x-12-33]
37. Lei H, Huang Z, Zhang J, Yang Z, Tan E, Zhou F, et al. Joint detection and clinical score prediction in Parkinson's disease via multi-modal sparse learning. Expert Systems with Applications 2017 Sep;80:284-296. [doi: 10.1016/j.eswa.2017.03.038]
38. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019 Aug;25(8):1301-1309 [FREE Full text] [doi: 10.1038/s41591-019-0508-1] [Medline: 31308507]
39. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Critical Care Medicine 2018;46(4):547-553. [doi: 10.1097/ccm.0000000000002936]
40. Kim J, Chae M, Chang H, Kim Y, Park E. Predicting Cardiac Arrest and Respiratory Failure Using Feasible Artificial Intelligence with Simple Trajectories of Patient Data. J Clin Med 2019 Aug 29;8(9):1336 [FREE Full text] [doi: 10.3390/jcm8091336] [Medline: 31470543]
41. Li Q, Li Y, Gao J, Su L, Zhao B, Demirbas M, et al. A confidence-aware approach for truth discovery on long-tail data. Proc VLDB Endow 2014 Dec;8(4):425-436. [doi: 10.14778/2735496.2735505]
42. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From Big Data to Precision Medicine. Front Med (Lausanne) 2019;6:34 [FREE Full text] [doi: 10.3389/fmed.2019.00034] [Medline: 30881956]
43. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
44. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
45. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056]
46. Qassem T. Emerging technologies for dementia patient monitoring. In: Xhafa F, Moore P, Tadros G, editors. Advanced Technological Solutions for E-Health and Dementia Patient Monitoring. Dallas, TX: IGI Global; 2015:62-104.
47. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A Hybrid Approach to Alleviating Class Imbalance. IEEE Trans Syst, Man, Cybern A 2010 Jan;40(1):185-197. [doi: 10.1109/tsmca.2009.2029559]
48. Ammar H, Eaton E, Ruvolo P, Taylor M. Unsupervised cross-domain transfer in policy gradient reinforcement learning via manifold alignment. 2015 Presented at: Twenty-Ninth AAAI Conference on Artificial Intelligence; Jan 25-30, 2015; Austin, TX.
49. Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/tkde.2009.191]
50. Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. 2018 Presented at: 2018 International interdisciplinary PhD workshop (IIPhDW); May 9-12, 2018; Swinoujscie, Poland. [doi: 10.1109/iiphdw.2018.8388338]
51. Huang C, Li Y, Chen CL, Tang X. Deep Imbalanced Learning for Face Recognition and Attribute Prediction. IEEE Trans Pattern Anal Mach Intell 2019:1-1. [doi: 10.1109/tpami.2019.2914680]
52. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. Magn Reson Med 2019 Jan;81(1):116-128 [FREE Full text] [doi: 10.1002/mrm.27355] [Medline: 29774597]
53. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imag 2016 Jul 01;3(3):034501. [doi: 10.1117/1.jmi.3.3.034501]
54. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. 2017 Presented at: AMIA 2017 Annual Symposium Proceedings; Nov 4-8, 2017; Washington, DC.
55. Shin H, Tenenholtz NA, Rogers JK, Schwarz CG, Senjem ML, Gunter JL, et al. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. 2018 Presented at: International workshop on simulation and synthesis in medical imaging; Sep 6, 2018; Granada, Spain.
56. Jeni L, Cohn J, De LTF. Facing imbalanced data--recommendations for the use of performance metrics. 2013 Presented at: Humaine association conference on affective computing and intelligent interaction; Sep 2-5, 2013; Geneva, Switzerland. [doi: 10.1109/acii.2013.47]
57. López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences 2013 Nov;250:113-141. [doi: 10.1016/j.ins.2013.07.007]

Abbreviations

AUC: area under the receiver operating characteristic curve
EMS: emergency medical service
EVT: endovascular thrombectomy
IV tPA: intravenous tissue plasminogen activator
K-NN: K-nearest neighbor
MRC: Medical Research Council
NIHSS: National Institutes of Health Stroke Scale
SMOTE: synthetic minority oversampling technique
SVM: support vector machine

Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by Y Arai, KL Ong; comments to author 18.06.20; revised version received 13.08.20; accepted 13.08.20; published 16.09.20.

Please cite as:
Park E, Lee K, Han T, Nam HS
Automatic Grading of Stroke Symptoms for Rapid Assessment Using Optimized Machine Learning and 4-Limb Kinematics: Clinical Validation Study
J Med Internet Res 2020;22(9):e20641
URL: http://www.jmir.org/2020/9/e20641/
doi:10.2196/20641
PMID:32936079

Original Paper

A Rapid Electronic Cognitive Assessment Measure for Multiple Sclerosis: Validation of Cognitive Reaction, an Electronic Version of the Symbol Digit Modalities Test

Rod M Middleton[1], BSc, MBA; Owen R Pearson[2], MBChB; Gillian Ingram[2], MBChB; Elaine M Craig[1], BA, MPhil; William J Rodgers[1], PhD; Hannah Downing-Wood[3], BSc; Joseph Hill[3], BSc; Katherine Tuite-Dalton[1], BSc; Christopher Roberts[1], BSc; Lynne Watson[2], BSc; David V Ford[1], MBA; Richard Nicholas[3], MBBS, FRCP; UK MS Register Research Group

[1]Population Data Science, Swansea University Medical School, Swansea University, Swansea, United Kingdom

[2]Department of Neurology, Morriston Hospital, Swansea Bay National Health Service Trust, Swansea, United Kingdom

[3]Department of Neurology, Charing Cross Hospital, Imperial College London, London, United Kingdom

Corresponding Author:
Rod M Middleton, BSc, MBA
Population Data Science
Swansea University Medical School
Swansea University
Data Science Building
Singleton Park
Swansea, SA2 8PP
United Kingdom
Phone: 44 1792606760
Email: r.m.middleton@swansea.ac.uk

Abstract

Background: Incorporating cognitive testing into routine clinical practice is a challenge in multiple sclerosis (MS), given the wide spectrum of both cognitive and physical impairments people can have and the time that testing requires. Shortened paper and verbal assessments predominate but still are not used routinely. Computer-based tests are becoming more widespread; however, changes in how a paper test is implemented can impact what exactly is being assessed in an individual. The Symbol Digit Modalities Test (SDMT) is one validated test that forms part of the cognitive batteries used in MS and has some computer-based versions. We developed a tablet-based SDMT variant that has the potential to be ultimately deployed to patients' own devices.

Objective: This paper aims to develop, validate, and deploy a computer-based SDMT variant, the Cognition Reaction (CoRe) test, that can reliably replicate the characteristics of the paper-based SDMT.

Methods: We carried out analysis using Pearson and intraclass correlations, as well as a Bland-Altman comparison, to examine consistency between the SDMT and CoRe tests and for test-retest reliability. The SDMT and CoRe tests were evaluated for sensitivity to disability levels and age. A novel metric in CoRe was found: question answering velocity could be calculated. This was evaluated in relation to disability levels and age for people with MS and compared with a group of healthy control volunteers.

Results: SDMT and CoRe test scores were highly correlated and consistent with 1-month retest values. Lower scores were seen in patients with higher age and some effect was seen with increasing disability. There was no learning effect evident. Question answering velocity demonstrated a small increase in speed over the 90-second duration of the test in people with MS and healthy controls.

Conclusions: This study validates a computer-based alternative to the SDMT that can be used in clinics and beyond. It enables accurate recording of elements of cognition relevant in MS but offers additional metrics that may offer further value to clinicians and people with MS.

(J Med Internet Res 2020;22(9):e18234) doi:10.2196/18234

KEYWORDS

cognition; multiple sclerosis; eHealth; electronic assessment; patient reported outcomes; neurology

Introduction

Background

Multiple sclerosis (MS) is an inflammatory demyelinating and degenerative disease of the central nervous system and the most common nontraumatic cause of disability in young adults worldwide [1]. The dominant phenotype is characterized by relapses (attacks) and remissions, known as relapsing-remitting MS (RRMS). In the majority of those affected with RRMS, the condition evolves, within 10 to 15 years, into secondary progressive MS (SPMS). About 15% of people with MS develop primary progressive MS (PPMS), characterized by progressive neurological dysfunction from onset [2].

Motor impairment forms the most overt impact of MS but cognitive impairment affects up to 40% of people with MS, rising to 80% in those with the progressive forms of the disease [3]. It has substantial impact on disability and can, when present in isolation, limit employment prospects [4]. However, in the early stages of MS, formal cognitive testing can show minimal changes in a wide variety of domains [5]. Later, as the disease advances, the picture becomes more coherent, with impairments in speed of information processing, attention, episodic memory, and executive function dominating. These impact independence and mood and can lead to social isolation [6].

Cognitive testing itself can be demanding on patients, causing difficulties for those with attentional disorders, fatigue, and physical limitations [7]. The time and attention required in a busy clinic environment makes test delivery in a routine context a challenge for both patient and assessors. To this end, in MS, a number of simplified tests of cognition have been developed for clinical use. These include the Brief International Cognitive Assessment for MS [8], the Brief Repeatable Battery of Neuropsychological Tests [9], and the Minimal Assessment of Cognitive Function in Multiple Sclerosis [10]. In most cases, these tests are still largely paper- or apparatus-based exercises completed in front of an assessor and take the form of a battery of tests that incorporate multiple testing methodologies.

One common element of the MS testing batteries is the Symbol Digit Modalities Test (SDMT) [11]. It assesses organic cerebral dysfunction and has a proven history as an effective outcome measure in a number of MS trials [10,11] and in other conditions [12]. The SDMT consists of matching symbols against digits within 90 seconds, the result being the total number of correct answers. Participants are given a practice number of attempts and then perform the timed assessment. The implementation of the test typically takes 5 minutes, including instruction and demonstration. The responses can be written or spoken out loud and recorded by the assessor [13].

A number of electronic variants of the SDMT have been developed [14,15], but as yet, they are not used routinely to assess cognitive impairment [16]. Their implementation varies from the original paper test, but the impact of these slight variations is as yet unclear, as impairment in individuals with MS can vary widely with different elements, such as fatigue, which can slow reactions, and physical issues such as ataxia or weakness, which can introduce further variability if a screen or keyboard needs to be manipulated. This is a further challenge if a test is to be administered without an assessor present. However, the computer-based approaches have the potential to offer additional information over the paper-based or oral approaches, as additional metrics can be quantified and these may enhance the information available from the test.

The United Kingdom Multiple Sclerosis Register (UKMSR) was established in 2011 as a means of capturing real-world evidence of living with MS in the United Kingdom. There are comprehensive data on 11,387 people with MS registered on the UKMSR via the internet and more than 13,000 consented clinically via a network of National Health Service (NHS) centers [17]. An online portal facilitates collection of longitudinal patient-reported outcomes (PROs) and real-world evidence of living with MS, but none of the instruments currently capture cognitive function. Given the need to understand in more depth the performance characteristics of electronic testing and the key role of cognitive impairment in MS, we developed an electronic variant of the SDMT that could be implemented rapidly and routinely at clinical centers to address this need. Ultimately, as an electronic register, if this type of testing is validated, then it could be also carried out in the patient's home, which would also help patients who are unable to physically attend clinics.

Objectives

This paper aims to develop, validate, and deploy a computer-based SDMT variant, the Cognition Reaction (CoRe) test, that can reliably replicate the characteristics of the paper-based SDMT and assess its utility for deployment as a meaningful measure to assess cognition in an MS population.

Methods

Population

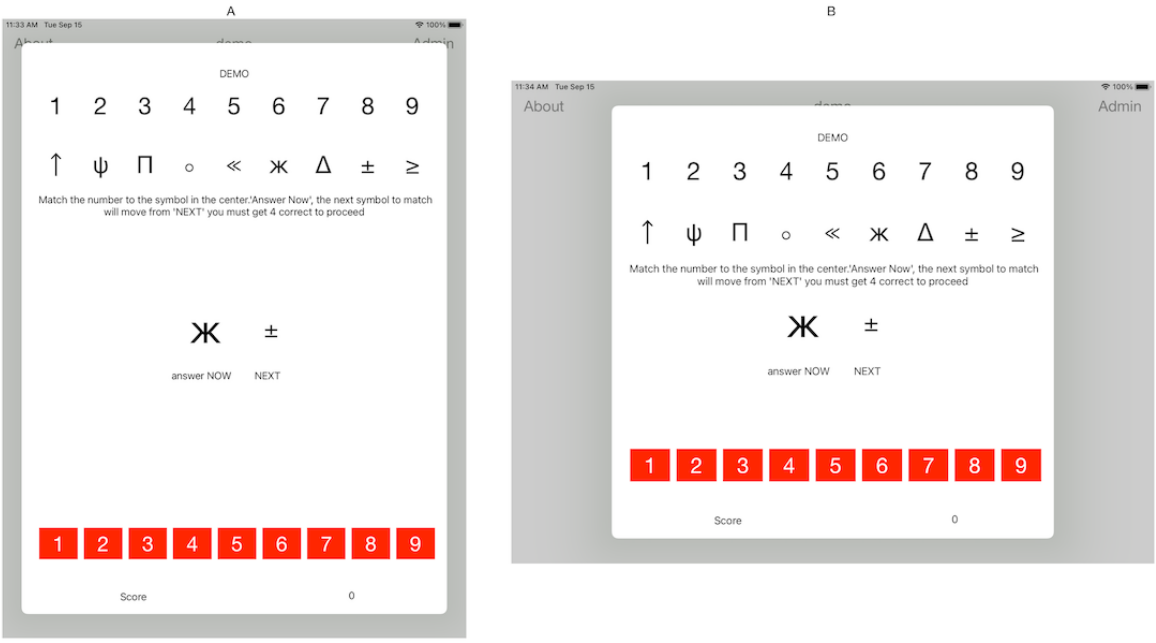
All participants gave informed consent, and the study has ethical approval from South West Central Bristol Ethics Committee (16/SW/0194). Participants were recruited from Morriston Hospital, Swansea Bay University Health Board and Charing Cross Hospital, Imperial College Healthcare NHS Trust. The people with MS that took part in the study were recruited at either progressive MS teaching days or as part of their routine visits to their respective hospitals. Demographic data and an Expanded Disability Status Score (EDSS) [18] were recorded at the time of testing. Healthy volunteers were recruited from Swansea University Medical School and Imperial College London to provide a control group of test scores with anonymized demographic data. Healthy volunteers were recruited from among the staff at the two clinical sites and included a mix of staff and PhD students from Swansea University. None of the healthy controls had MS and no one approached refused. All participants had completed at least full

formal secondary education. There were no declared visual problems in the population.

CoRe Test App

The Cognition Reaction (CoRe) test was inspired by the SDMT; however, there are some key differences. The CoRe test presents 9 different symbols displayed at the top of the screen, with corresponding numbers, 1 through 9, underneath. The symbols are randomized every time the app is launched, and the center of the screen displays 2 symbols, the one to be identified now and the next one. At the bottom of the screen, there are a number of buttons labelled 1 through 9 that participants tap to match the central symbol on the screen. Data recorded include the number of symbols accurately tapped, as for the SDMT, but in addition, CoRe automatically registers the time between responses and the number of incorrect responses. Further details of the app are presented in Multimedia Appendix 1 [19,20]. The app is entirely self-contained, with no requirement for internet access. The CoRe test app can be seen in Figure 1.

Figure 1. Cognitive Reaction test app shown running in portrait and landscape modes.



App Testing

For the MS population, participants first completed the paper SDMT using the traditional written method, requiring the paper test, a pen, and a stopwatch. Following this, participants were handed an iPad and given an introduction by a researcher from the UKMSR team, merely demonstrating the 2 orientations that the device could be placed in. The orientation that participants chose was not recorded as part of this assessment. They were then invited to follow the written directions on the app. They were first presented with a demonstration mode and encouraged to run through at least once. A score of 4, which was displayed on the screen, was required to progress to the main test. This could be repeated if desired. Once ready, participants hit “start” and were given 90 seconds to complete the test. A countdown timer was displayed on the screen of the iPad. Visual acuity was not formally assessed, and no participants claimed that they could not see the icons on the tablet screen. Test environments were controlled for noise and disturbance. Some participants were retested 1 month later in the same environment to determine the consistency of the results.

Statistical Analysis

Analysis was carried out using the Pandas library for Python (version 3.773) [21] and the R statistical language (version 3.6.0; R Foundation for Statistical Computing) [22]. Graphs and images were generated using Seaborn [23] and ggplot2 (version 3.0.0) [24]. Correlation was used to compare the validity of the paper and electronic versions of the tests and the test-retest reliability of the CoRe test. Pearson *r* was calculated for test scores from the CoRe test and the SDMT, with mean difference evaluated using a 2-tailed paired samples *t* test and differences in variances compared using a Pitman-Morgan test for paired samples. Intraclass correlation was also performed on the first and second CoRe and SDMT test results. A Bland-Altman analysis was used as an additional measure of equivalency. The sensitivity of the CoRe and SDMT scores to disability levels and age were measured using analysis of variance (ANOVA) statistics, with post hoc Tukey tests used to determine any significant differences between groups.

To utilize the additional data generated by the CoRe test, the question answering velocity (QAV) was quantified as a measure of cognitive function. This was defined as the total number of correct answers given at a time divided by total current time elapsed in the test (correct answers/seconds). Multivariate linear regression was performed to determine if any relationship existed between the QAV and the time period of the questionnaire. The CoRe test lasts a total of 90 seconds, and responses were divided into thirds to study the rates of change over the first, second, and third sections of responses for each patient. For analysis, EDSS scores were divided into 3 categories: low (EDSS of 0-2.5), medium (EDSS of 3-5.5), and

high (EDSS of 6-10), as was age, with categories of 18-34 years, 35-54 years, and >55 years.

Results

Demographics

A total of 102 people with MS were recruited to the study (Table 1), of whom 30 returned within 1 month for a repeat test. All patients were over 18 years of age and had no significant comorbidities that would exclude them from being able to complete the paper or CoRe tests. No participants were excluded from the study, and none reported a relapse of MS at any point in the testing. Mean age of the people with MS cohort tested was younger than the overall MS Register population, with a slightly lower proportion of patients with PPMS and SPMS (Table 1). A total of 45 anonymous healthy controls were tested during the development of the app; apart from not completing an initial paper SDMT, conditions were similar to the MS cohort. Both healthy controls and people with MS had completed at least 12 years of education.

Table 1. Demographics of cohort and healthy controls undertaking the CoRe test. The UKMSR population is shown for comparison.

Characteristic	Total UKMSR[a]	CoRe[b] cohort (MS[c])	CoRe cohort (healthy controls)	Cohort difference[d], chi-square test (*df*)	Cohort difference[d], *t* test (*df*)	*P* value
Total participants, n	11,387	102	45	N/A[e]	N/A	N/A
Gender, n (%)				0.3 (1)	N/A	.57
Female	8387 (73.7)	70 (68.6)	28 (62.2)			
Male	3000 (26.3)	32 (31.4)	17 (37.8)			
MS type, n (%)				N/A	N/A	N/A
RRMS[f]	5988 (52.6)	86 (83.2)	N/A			
PPMS[g]	1492 (13.1)	5 (5.6)	N/A			
SPMS[h]	2945 (25.9)	9 (9.3)	N/A			
Other	962 (8.4)	2 (1.9)	N/A			
Age (years), mean (SD)	53.6 (11.8)	44.0 (11.0)	38.1 (11.9)	N/A	2.891 (145)	.004
Age at diagnosis (years), mean (SD)	39.2 (10.3)	34.6 (10.6)	N/A	N/A	N/A	N/A
EDSS[i], median (range)	6 (0-9.5)	3.5 (1-8)	N/A	N/A	N/A	N/A

[a]UKMSR: United Kingdom Multiple Sclerosis Register.

[b]CoRe: Cognitive Reaction.

[c]MS: multiple sclerosis.

[d]Difference between people with multiple sclerosis and healthy controls.

[e]N/A: not applicable.

[f]RRMS: relapsing-remitting multiple sclerosis.

[g]PPMS: primary progressive multiple sclerosis.

[h]SPMS: secondary progressive multiple sclerosis.

[i]EDSS: Expanded Disability Status Score.

CoRe Test in People With MS and Control Group: Comparison of Total Correct Responses

The first set of CoRe test scores for people with MS were compared with those of the healthy control group. Mean test results for people with MS were 39.0 (SD 13.3), while mean scores for the healthy control group were 56.1 (SD 15.9). An unpaired *t* test found that people with MS had significantly lower scores (t_{145}=–6.769; *P*<.001), with no significant difference in variance between the groups ($F_{101,44}$=0.701; *P*=.15).

CoRe Test and SDMT in People With MS: Comparison of Total Correct Responses

People with MS completed both the CoRe test and SDMT together on 2 occasions, 1 month apart. The first test response distributions for the CoRe test and SDMT were normally distributed (Shapiro-Wilk tests with *P*=.48 and *P*=.61, respectively) and were strongly correlated (Pearson r_{100}=0.800; *P*<.001). First test participants scored a mean of 4.40 responses lower for the CoRe test compared with the SDMT, as seen in Table 2 (paired samples t_{101}=5.390; *P*<.001), but there was no significant difference in the variance (Pitman-Morgan test: t_{100}=–0.879; *P*=.38), with good agreement between tests (Figure 2). When the CoRe test and SDMT were repeated for a second time, the mean CoRe test score was not significantly lower than

the SDMT (1.4 responses difference; t_{29}=0.954; *P*=.35). Again, there was a strong correlation between the second CoRe test and second SDMT (Pearson r_{28}=0.842; *P*<.001). Table 2 shows the baseline and retest responses for those who completed it.

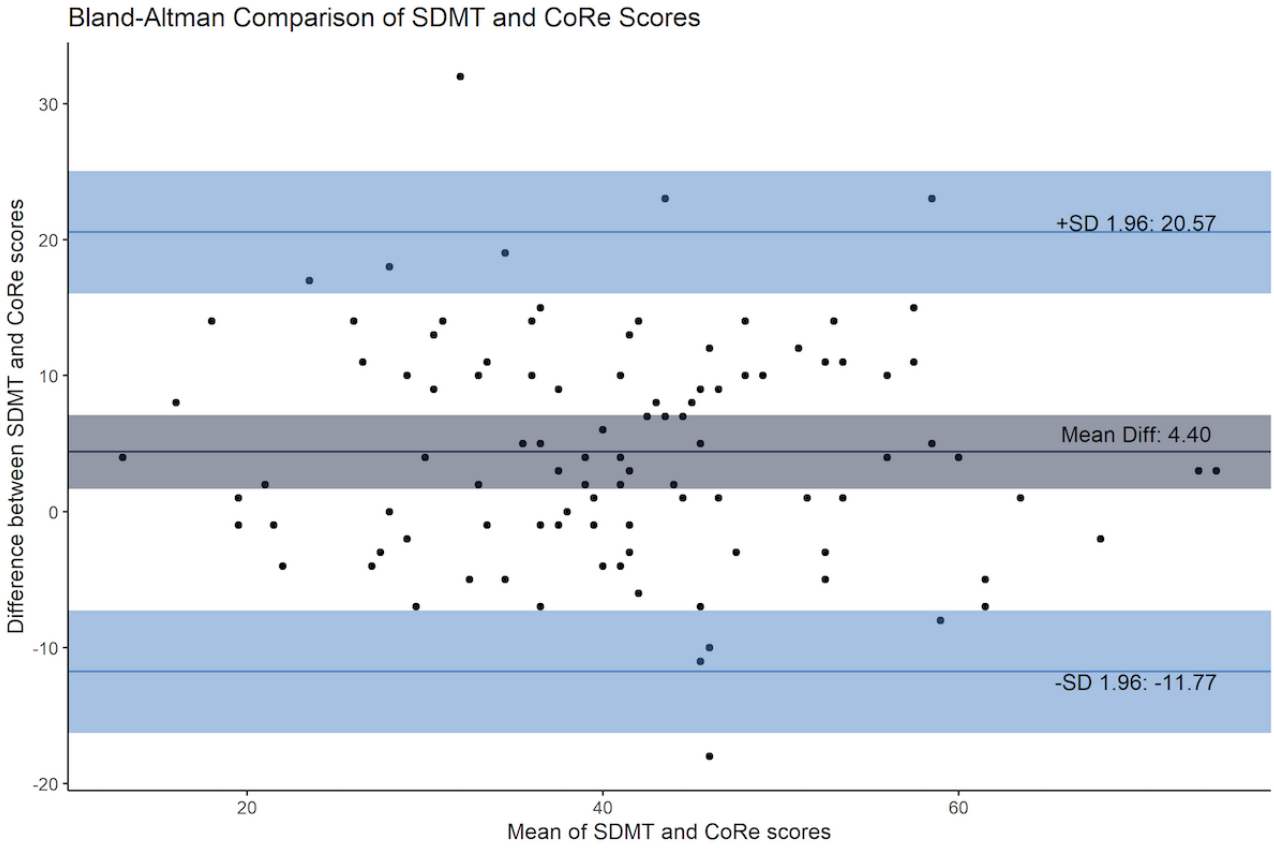
Table 2. Baseline and retest SDMT and CoRe test total responses at baseline and retest 1 month later.

Test	Participants, n	Score, mean (SD), range
Baseline		
SDMT[a]	102	43.4 (12.6), 15-76
CoRe[b] test	102	39.0 (13.3), 11-73
Retest		
SDMT	30	41.9 (14.6), 14-76
CoRe test	30	40.5 (13.9), 20-70

[a]SDMT: Symbol Digit Modalities Test.

[b]CoRe: Cognitive Reaction.

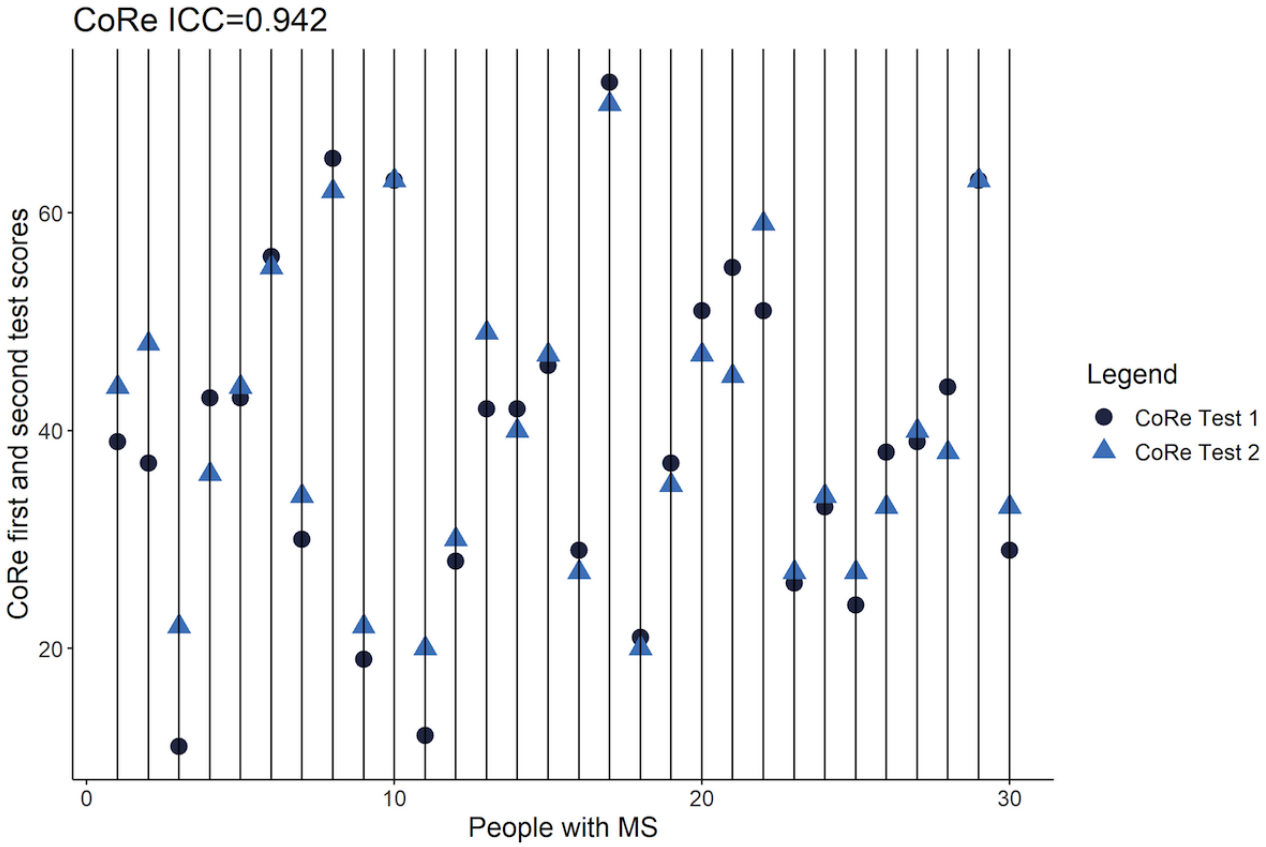
Figure 2. Bland-Altman comparison of first CoRe test with paper SDMT scores. CoRe: Cognitive Reaction; SDMT: Symbol Digit Modalities Test.



CoRe Test and SDMT Test-Retest Reliability

First and second CoRe test and SDMT scores were evaluated for test-retest reliability and scores at a 1-month interval. The CoRe tests were normally distributed and demonstrated consistency (Pearson correlation coefficient *r*=0.947; t_{28}=15.60; *P*<.001). Differences in means were normal (Shapiro-Wilk test *P*=.81) and not significantly different (t_{29}=–0.944; *P*=.35), with equal variances (Pitman-Morgan t_{28}=1.784; *P*=.09). The intraclass correlation coefficient between the first and second CoRe tests was found to be 0.942 (95% CI 0.882-0.0972; $F_{29,30}$=33.2; *P*<.001) (Figure 3).

Figure 3. Intraclass correlation coefficients between the first and retested CoRe tests. CoRe: Cognitive Reaction; ICC: intraclass correlation coefficient; MS: multiple sclerosis.



Test-retest correlations were observed in the same people completing the SDMT at a 1-month interval. Scores were normally distributed and consistent (Pearson correlation r=0.936; t_{28}=14.052; P<.001) and differences in means were normal (Shapiro-Wilk test P=.44) and not significantly different (t_{29}=–0.919; P=.37), with equal variances (Pitman-Morgan t_{28}=–0.743; P=.46). The intraclass correlation coefficient between the first and second SDMT tests was found to be 0.935 (95% CI 0.869-0.968; $F_{29,30}$=29.6; P<.001).

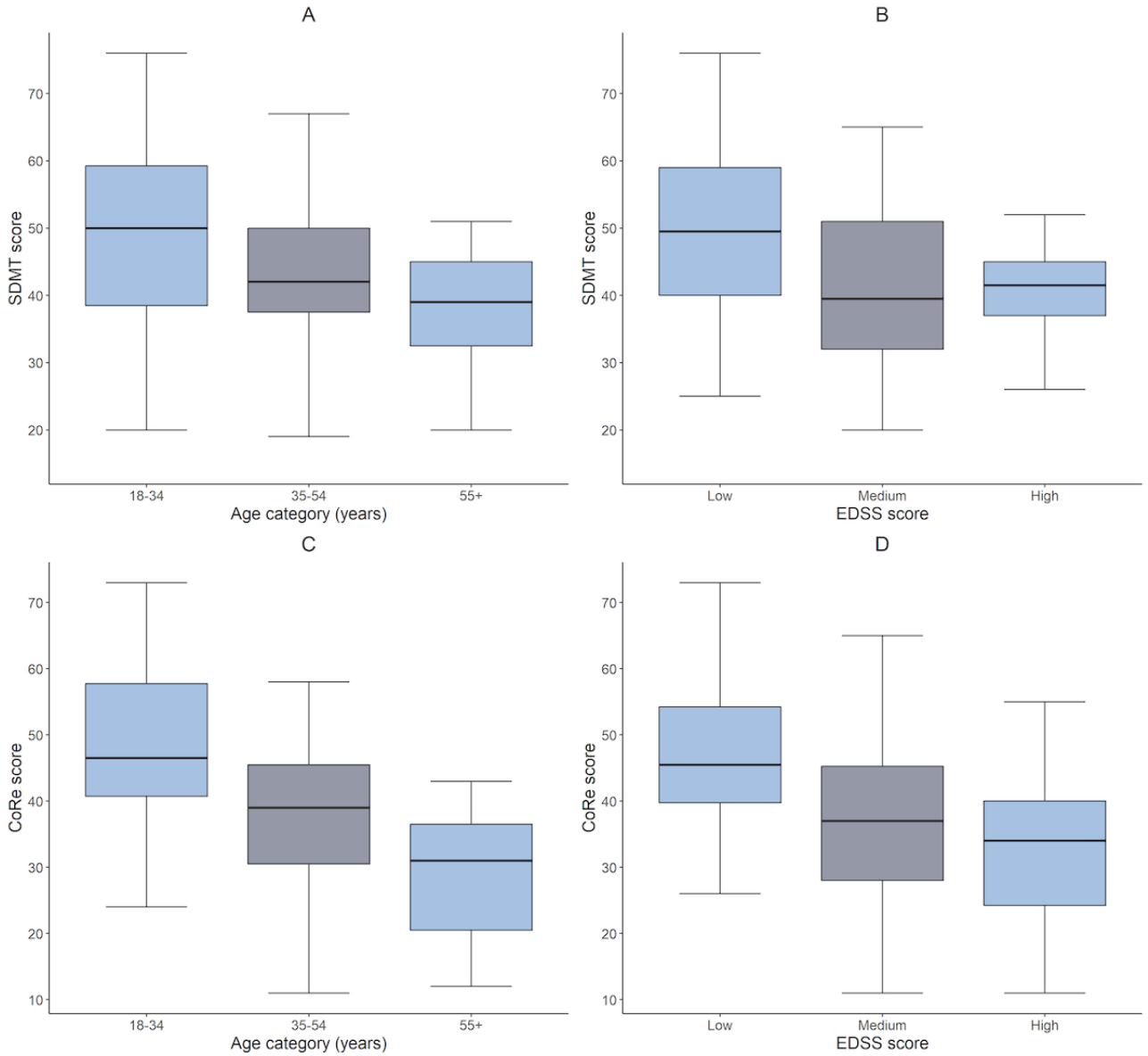
CoRe Test Total Correct Response Score Is Impacted by Age and Disability in MS, Whereas SDMT Is Only Affected by Disability

We examined the impact of age, gender, and EDSS on the total correct responses (Figure 4). An ANOVA for SDMT scores with respect to age and EDSS found no significant impact of age (aged 18-34 years: mean 48.1, SD 15.5; aged 35-54 years: mean 43.2, SD 11.8; and 55+ years: mean 38.3, SD 9.0; F_2=1.036; P=.36), but significance for EDSS (low EDSS: mean 49.8, SD 12.9; medium EDSS: mean 41.4, SD 12.3; high EDSS: mean 38.6, SD 9.8; F_2=8.574; P<.001); post hoc Tukey tests showed higher scores in those in the lowest EDSS category compared with those in the highest EDSS category (P<.001) and compared with the medium EDSS category (P=.01). No significant difference was found between the low and medium EDSS categories.

In contrast, an ANOVA for CoRe test scores showed a significant difference in the total responses with age (aged 18-34 years: mean 48.6, SD 13.5; aged 35-54 years: mean 38.3, SD 11.6; and >55 years: mean 28.9, SD 9.8; F_2=8.633; P<.001) and EDSS (low EDSS: mean 47.4, SD 11.6; medium EDSS: mean 36.8, SD 12.7; high EDSS: mean 32.1, SD 10.7; F_2=18.151; P<.001). Post hoc Tukey tests showed those in the age group of 18 to 34 years had significantly higher scores than those in the 34 to 54 years (P=.01) and 55+ years group (P=.001), with no difference between the medium and high age groups. The lowest EDSS category was associated with higher CoRe test scores than both other groups (P<.001), with no difference between the medium and high EDSS groups.

Gender was not found to be significant for either SDMT or CoRe test scores.

Figure 4. Mean SDMT and CoRe scores with age categories and EDSS scores. CoRe: Cognitive Reaction; EDSS: Expanded Disability Status Score; SDMT: Symbol Digit Modalities Test.



Speed of Reaction (Question Answering Velocity) Derived From the CoRe Test Increases Throughout the Test and Correlates With Age, Gender, and Disability

Due to the way data are acquired for the CoRe test, we were able to measure the speed of reaction to each question and calculate the QAV as correct answers over time elapsed (seconds) continuously throughout the assessment. There was a significant range of QAV over the time frame of the test in people with MS, as illustrated in Figure 5, which shows the two individuals with the lowest and highest scores in the CoRe test. Breaking down the total correct answers into 3 sections also allowed us to quantify the change in QAV over the course of the CoRe test. Multiple linear regression models with the variables age, gender, and EDSS in people with MS found that QAV increased in each third of the test in people with MS and healthy controls (Table 3).

Figure 5. A polynomial regression of QAV for those people with MS with the lowest and highest scores in the cohort. CoRe: Cognitive Reaction; EDSS: Expanded Disability Status Score; MS: multiple sclerosis; QAV: question answering velocity.

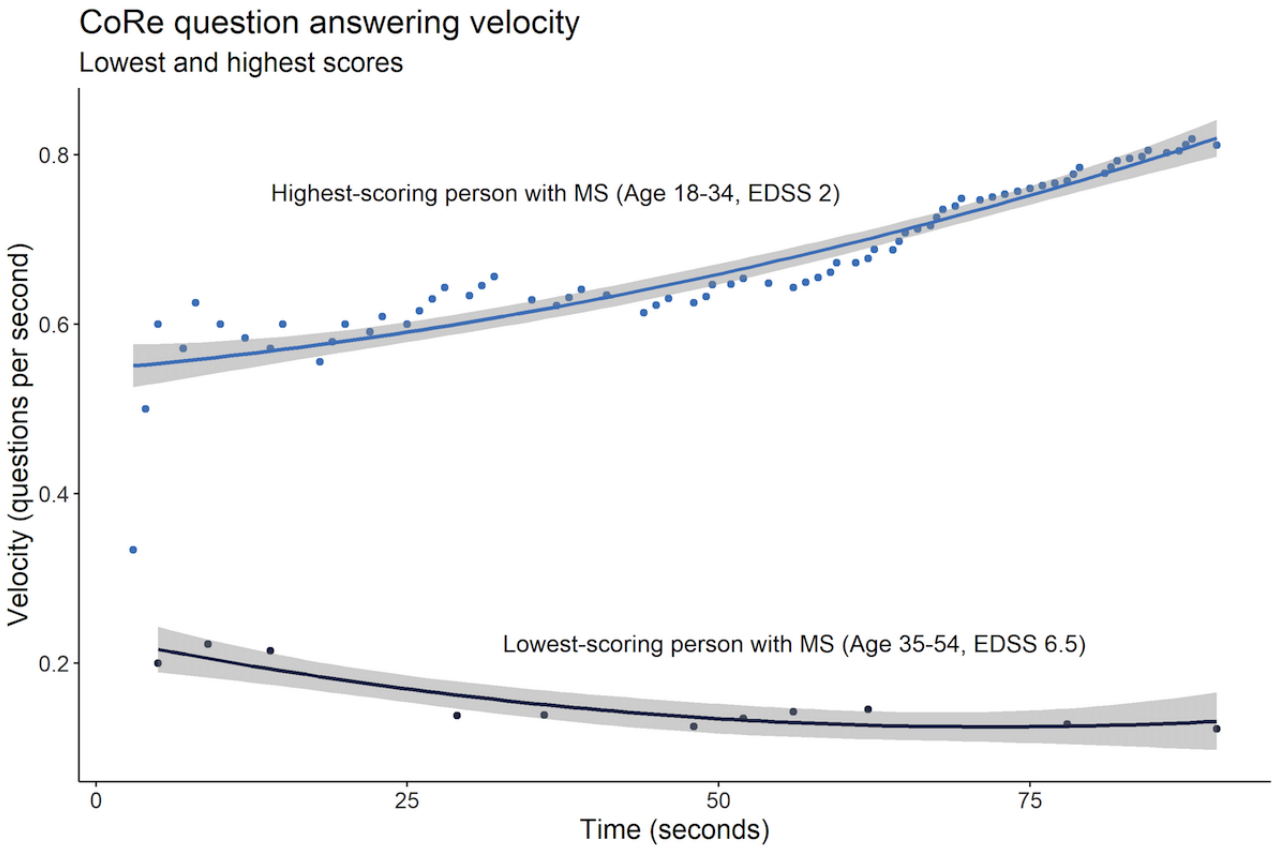


Table 3. Multivariate models in people with MS (R^2=0.396; $F_{5,3973}$=520.4; *P*<.001) and healthy controls (R^2=0.323; $F_{4,2521}$=300.1; *P*<.001) for QAV over the time frame of the Cognitive Reaction test, with additional covariates age and gender. EDSS scores are given for people with MS only.

Variable	QAV[a] of people with MS[b]		QAV of healthy controls	
	β coefficient (95% CI)	*P* value	β coefficient (95% CI)	*P* value
Second section compared to first	.045 (0.037 to 0.053)	<.001	.070 (0.056 to 0.085)	<.001
Third section compared to first	.071 (0.063 to 0.080)	<.001	.110 (0.094 to 0.123)	<.001
Age	–.005 (–0.005 to –0.006)	<.001	–.008 (–0.007 to –0.008)	<.001
Female gender	.049 (0.041 to 0.056)	<.001	–.043 (–0.055 to –0.031)	<.001
EDSS[c]	–.017 (–0.015 to –0.019)	<.001	N/A[d]	N/A

[a]QAV: question answering velocity.

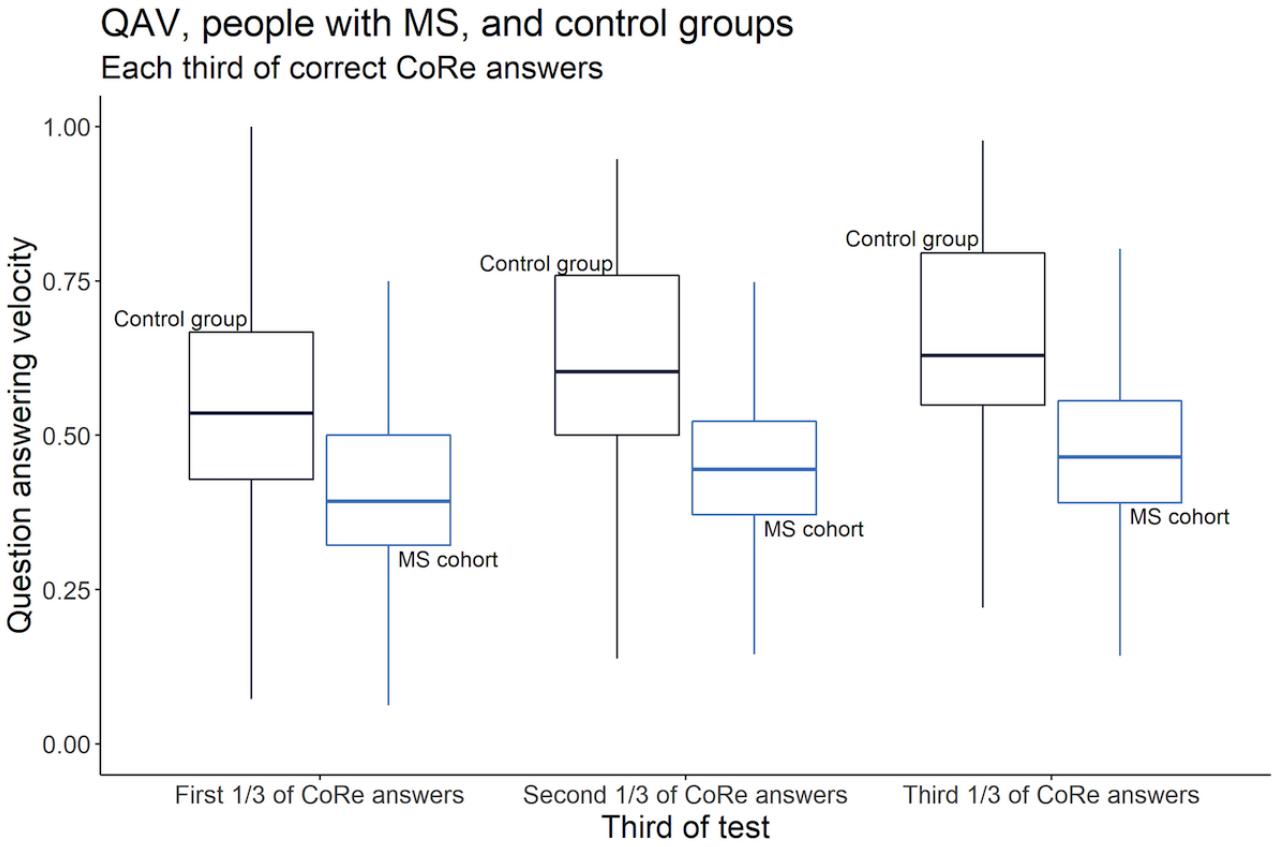
[b]MS: multiple sclerosis.

[c]EDSS: Expanded Disability Status Score.

[d]N/A: not applicable.

Both groups answered more quickly as the test progressed (the control group at an even faster rate than people with MS), with the second and third sections of their correct answers being completed in less time than the first. The gradient is similar in both populations (Figure 6). In both populations, increased age was associated with slowing of QAV by 0.007 to 0.008 questions per second for each year increase in age. For control participants, female gender was associated with a slowing of QAV by 0.034 questions per second, whereas in people with MS, female gender was associated with an increase in QAV of 0.049 questions per second. However, disability slowed QAV by 0.017 questions per second for every increase in EDSS by 1 point (Table 3).

Figure 6. Comparison of increase in speed between each test third for healthy and MS populations. CoRe: Cognitive Reaction; MS: multiple sclerosis; QAV: question answering velocity.



We next directly compared the variables associated with CoRe test QAV and the CoRe test total response score. A regression using the variables age, gender, and EDSS score found that the CoRe test QAV was significantly impacted by all 3 factors, whereas the CoRe test score (total correct answers) found significant impacts only from EDSS and age (Table 4).

Table 4. Impact of age, gender, and EDSS on total response score (R^2=0.383; $F_{3,98}$=20.3; *P*<.001) in people with MS cohort.

Variable	CoRe[a] test score	
	β coefficient (95% CI)	*P* value
EDSS[b]	–2.103 (–3.390 to –0.808)	.002
Age	–.489 (–0.713 to –0.265)	<.001
Female gender	4.413 (–0.155 to 8.981)	.06

[a]CoRe: Cognitive Reaction.

[b]EDSS: Expanded Disability Status Score.

Discussion

Summary of Findings

This study aims to validate an electronic variant of the SDMT, comparing the CoRe test with the established paper-based SDMT within an MS cohort in 2 independent UK centers, examining its overall reliability and suitability. In addition, we quantified an additional metric that can be extracted from the electronic implementation. The total response scores for the CoRe test were on average lower than the SDMT but showed good correlation with the paper test, though there are clear differences in responses across age groups. Having the understanding that the CoRe performs similarly across these deviations allows it to be compared with the paper-based test, though it is not a like-for-like match. However, the consistency of the test and its utility remain clear. The CoRe test showed consistent responses over time and demonstrated similar test-retest properties to the SDMT, as with other electronic implementations [14]. These findings suggest that the CoRe test is an appropriate alternative to measure of cognitive ability as assessed by the SDMT.

We confirmed that a reduction in correct responses for both the SDMT and CoRe test correlates with increasing disability, but in addition, a reduction in correct responses correlates with increasing age in people with MS. Using the advantages of an electronic implementation, we were able to measure the QAV and found that both people with MS and healthy controls

increase their QAV throughout the test and also that in both groups, an increased QAV correlates with younger age and male gender. This implies these correlations are not associated with MS-specific cognitive decline. However, increased QAV is also associated with lower disability, only present in those with MS. In our testing, increasing age showed a reduction in correct responses over the test. This finding corresponds with other SDMT testing in populations [25], and there is evidence for older participants performing poorly over the duration of the test, with studies showing decreased reaction times (about 0.5 ms/year) [26] in simple reaction-style tests in older people. There is also the effect of older people's familiarity with tablet computers [27] that could have some impact on this. This will be investigated in future testing.

There are a number of computer-based variants of the SDMT, one of the first being the computerized Symbol Digit Modalities Test (c-SDMT) [14], which showed excellent sensitivity in 119 people with MS versus 38 healthy controls, with people with MS performing significantly worse than the healthy controls. Use of the c-SDMT has not become widespread, most likely due to the technology platform that it was developed on and the stringent test description (Windows PC, 19-inch screen with participant at 15 inches from the screen), making deployment challenging. A more recent implementation of a computer-based SDMT is the processing speed test (PST) [15], which was also tested against a healthy control population and forms one element of the Floodlight assessment tool [28]. The PST showed similar results as we have demonstrated and has shown high reliability when reproduced within Floodlight on patients' own devices. Small differences in implementation of the same paper-based test can impact what is being tested and need to be understood. The CoRe implementation requires the screen to be touched, which adds a visuospatial element to the assessment, and this will have an impact in some people with MS. It also presents 2 symbols in random order as opposed to a standard sheet of symbols; this change means that there is less likely to be a learning effect on retesting. A key issue with computer-based implementation is the impact of rapid hardware and software development, which results in a need to develop applications that can adapt to a changing environment. Another issue is the variety of devices, such as desktops, laptops, and smartphones, that are currently in use, especially if the test is to be performed without an assessor present. CoRe has been developed to run at multiple screen sizes and on different devices, with an interface—two symbols seen at a time—that is suited to a small screen. This will have to be tested separately.

Prior studies, and our results, show that data produced by electronic tests are consistent, repeatable, and have utility to clinicians, informing on a vital aspect of patient care [29]. The scores on both paper SDMT and the CoRe test fall with increasing disability. The CoRe test is more sensitive than the SDMT to age, with the SDMT being only affected in populations older than 55 years [27]. The electronic CoRe test allows greater analysis of this effect, demonstrating slower mean response times in higher ages and disability groups. There is some evidence that there may be a gender difference in cognitive tests [30], with males and females performing differently at various ages in different test types. Notably, this is seen with visual reaction times, and this would be consistent with the implementation of the test presented here. The fact that this extra variable of reaction time (QAV) can be measured as part of the CoRe test could have clinical or research utility in the future. Having additional quantifiable clinical measurement information via a simple-to-implement and rapid test could hopefully have some relevance to everyday clinical practice, research, and medication trials. Benedict et al [13] state that the current definition of "NEDA" (no evidence of disease activity) is predicated on largely physical outcome measures, but cognition is so fundamental to socialization, employment, and quality of life beyond pure health care that a prolonged measurement of cognitive aspects could add a compelling dimension to our understanding of disease activity.

Limitations

We identified some limitations with this study. First, there were few people with MS with progressive disease and advanced disability, and we did not have complete directly measured cognitive assessments. In addition, the population that took 1-month follow-up tests was limited, and we have only tested this on a single type of device here. The 1-month period chosen for retest represents the hospital visit pattern for some patients on a particular disease-modifying therapy. Differing retest periods should be tested in the future. Although testing was performed in the presence of a researcher, they had little or no input on the actual test itself—though this has been shown to not have effect on these types of tests [31]. We also did not consider the orientation of the device as having any effect. This could also be incorporated into future testing on other devices.

Given that the CoRe test is consistent and repeatable, we intend to test the app on other devices, including laptops and a variety of smartphones. This will facilitate completion of the test away from the clinic and will enable us to integrate the CoRe test into the range of PROs captured by the UKMSR. Additionally, this will allow us to carry out testing among participants with higher disability and more progressive disease at different intervals to ensure that the test maintains its reliability and repeatability. We recognize that the CoRe is not an exact replacement for SDMT. It is an entirely new test [32], but it is comparable and measurable compared with the SDMT.

Conclusion

The CoRe implementation of the SDMT test is reliable and correlates with the paper-based SDMT, while also offering the additional metric of patient reaction time (QAV). This will allow clinicians and researchers to capture important additional metrics in people with MS, and potentially in other diseases, quickly and reliably on existing tablet hardware.

Acknowledgments

We would like to thank all the participants of the UK MS Register but especially those people who directly took part in this research. Additionally, we thank the healthy participants that took part in the development of CoRe, and all the hard-working researchers and clinicians that take such a keen interest in the development of novel measures.

Group Collaborators:

Dr Judy Archer, Mid Yorkshire Hospitals NHS Trust

Dr Carlo Canepa, James Paget University Hospitals NHS Trust

Dr Viquar Chamoun, Buckingham Healthcare NHS Trust

Professor Jeremy Chataway, University College London Hospitals NHS Trust

Dr Abhijit Chaudhubl, Barking Havering and Redbridge Hospitals NHS Trust

Professor Alasdair Coles, Cambridge University Hospitals NHS Trust

Dr Matt Craner, Oxford University Hospitals NHS Trust

Professor H Emsley, Lancashire Teaching Hospitals NHS Trust

Dr Nikos Evangelou, Nottingham University Hospitals NHS Trust

Dr Leonora Finisku, Brighton and Sussex Hospitals NHS Trust

Dr Helen Ford, Leeds Teaching Hospitals NHS Trust

Ms Annmieke Fox, Poole Hospital NHS Trust

Ms Julie Foxton, Royal Berkshire NHS Trust

Dr Andrew Gale, Luton and Dunstable Hospital NHS Trust

Dr Ian Galea, University Hospital Southampton NHS Trust

Dr Andrew Graham, Ipswich Hospital NHS Trust

Dr Joe Guadango, Newcastle Upon Tyne Hospitals NHS Trust

Dr Dreedharan Harikrishnan, East Kent Hospitals University NHS Trust

Dr Tim Harrower, Royal Devon and Exeter NHS Trust

Dr Jeremy Hobart, Plymouth Hospitals NHS Trust

Dr Chris Kipps, Hampshire Hospitals NHS Trust

Ms Jo Kitley, Portsmouth Hospitals NHS Trust

Dr Monica Marta, Southend University Hospital NHS Trust

Dr Gavin McDonnell, Belfast Health and Social Care Trust

Dr Brendan McLean, Royal Cornwall Hospitals NHS Trust

Ms Charlotte Owen, Shrewsbury and Telford Hospital NHS Trust

Dr Laura Petzold, Maidstone and Tunbridge Wells NHS Trust

Dr David Rog, Salford Royal Hospital NHS Trust

Dr Klaus Schmierer, Barts Health NHS Trust

Professor Basil Sharrack, Sheffield Teaching Hospitals NHS Trust

Dr Zbignew Slowinski, Airedale NHS Trust

Dr Agne Straukiene, Torbay and South Devon NHS Trust

Dr Sigurlaug Sveinsbornsdottir, Basildon and Thurrock University Hospitals NHS Trust

Dr Stewart Webb, NHS Greater Glasgow and Clyde

Dr Heather Wilson, Royal Free London NHS Trust

Conflicts of Interest

None declared.

Multimedia Appendix 1
CoRe test Application Details.
[DOCX File , 14 KB - jmir_v22i9e18234_app1.docx]

References

1. Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. The Lancet 2017 Apr;389(10076):1336-1346. [doi: 10.1016/s0140-6736(16)30959-x]
2. Leray E, Yaouanq J, Le Page E, Coustans M, Laplaud D, Oger J, et al. Evidence for a two-stage disability progression in multiple sclerosis. Brain 2010 Jul;133(Pt 7):1900-1913 [FREE Full text] [doi: 10.1093/brain/awq076] [Medline: 20423930]
3. Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. Neurology 1991 May;41(5):685-691. [doi: 10.1212/wnl.41.5.685] [Medline: 2027484]
4. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. Lancet Neurol 2008 Dec;7(12):1139-1151. [doi: 10.1016/S1474-4422(08)70259-X] [Medline: 19007738]
5. de Gobbi Porto FH, Spíndola L, de Oliveira MO, Figuerêdo do Vale PH, Orsini M, Nitrini R, et al. A score based on screening tests to differentiate mild cognitive impairment from subjective memory complaints. Neurol Int 2013;5(3):e16 [FREE Full text] [doi: 10.4081/ni.2013.e16] [Medline: 24147213]
6. Kobelt G, Langdon D, Jönsson L. The effect of self-assessed fatigue and subjective cognitive impairment on work capacity: The case of multiple sclerosis. Mult Scler 2019 Apr;25(5):740-749. [doi: 10.1177/1352458518769837] [Medline: 29663869]
7. Sumowski JF, Benedict R, Enzinger C, Filippi M, Geurts JJ, Hamalainen P, et al. Cognition in multiple sclerosis. Neurology 2018 Jan 17;90(6):278-288. [doi: 10.1212/wnl.0000000000004977]
8. Langdon D, Amato M, Boringa J, Brochet B, Foley F, Fredrikson S, et al. Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Mult Scler 2012 Jun;18(6):891-898. [doi: 10.1177/1352458511431076] [Medline: 22190573]
9. Bever C, Grattan L, Panitch H, Johnson K. The Brief Repeatable Battery of Neuropsychological Tests for Multiple Sclerosis: a preliminary serial study. Mult Scler 1995 Nov;1(3):165-169. [doi: 10.1177/135245859500100306] [Medline: 9345448]
10. Benedict RHB, Cookfair D, Gavett R, Gunther M, Munschauer F, Garg N, et al. Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J Int Neuropsychol Soc 2006 Jul;12(4):549-558. [doi: 10.1017/s1355617706060723] [Medline: 16981607]
11. Sheridan LK, Fitzgerald HE, Adams KM, Nigg JT, Martel MM, Puttler LI, et al. Normative Symbol Digit Modalities Test performance in a community-based sample. Arch Clin Neuropsychol 2006 Jan;21(1):23-28. [doi: 10.1016/j.acn.2005.07.003] [Medline: 16139470]
12. Sachdev P, Anstey K, Parslow R, Wen W, Maller J, Kumar R, et al. Pulmonary function, cognitive impairment and brain atrophy in a middle-aged community sample. Dement Geriatr Cogn Disord 2006;21(5-6):300-308. [doi: 10.1159/000091438] [Medline: 16484809]
13. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, Multiple Sclerosis Outcome Assessments Consortium. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr;23(5):721-733. [doi: 10.1177/1352458517690821] [Medline: 28206827]
14. Akbar N, Honarmand K, Kou N, Feinstein A. Validity of a computerized version of the symbol digit modalities test in multiple sclerosis. J Neurol 2011 Mar;258(3):373-379. [doi: 10.1007/s00415-010-5760-8] [Medline: 20924594]
15. Rao SM, Losinski G, Mourany L, Schindler D, Mamone B, Reece C, et al. Processing speed test: Validation of a self-administered, iPad-based tool for screening cognitive dysfunction in a clinic setting. Mult Scler 2017 Dec;23(14):1929-1937. [doi: 10.1177/1352458516688955] [Medline: 28080262]
16. Klein OA, das Nair R, Ablewhite J, Drummond A. Assessment and management of cognitive problems in people with multiple sclerosis: A National Survey of Clinical Practice. Int J Clin Pract 2018 Dec 03;73(3):e13300. [doi: 10.1111/ijcp.13300] [Medline: 30507025]
17. Middleton R, Rodgers W, Chataway J, Schmierer K, Rog D, Galea I, et al. Validating the portal population of the United Kingdom Multiple Sclerosis Register. Mult Scler Relat Disord 2018 Aug;24:3-10. [doi: 10.1016/j.msard.2018.05.015] [Medline: 29860199]
18. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
19. Kremenek T. Swift 4.0 Released!. Swift Blog. 2017 Sep. URL: https://swift.org/blog/swift-4-0-released/ [accessed 2019-03-05]
20. Xcode 10. Apple Inc. Cupertino, CA: Apple Inc; 2017. URL: https://developer.apple.com/documentation/xcode-release-notes/xcode-10-release-notes [accessed 2019-03-05]

21. McKinney W. Data Structures for Statistical Computing in Python. 2010 Presented at: Proceedings of the 9th Python in Science Conference; June 28-July 3, 2010; Austin, TX p. 51-56. [doi: 10.25080/majora-92bf1922-00a]
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2013. URL: http://www.R-project.org/ [accessed 2019-03-05]
23. Waskom M, et al. seaborn: v0.9.0. mwaskom/seaborn.: Zenodo; 2018 Jul. URL: https://seaborn.pydata.org/ [accessed 2019-03-05]
24. Wickham H. ggplot2: Elegant Graphics for Data Analysis. ggplot2. New York, NY: Springer-Verlag; 2016. URL: https://ggplot2-book.org/ [accessed 2019-03-05]
25. Kiely KM, Butterworth P, Watson N, Wooden M. The Symbol Digit Modalities Test: Normative data from a large nationally representative sample of Australians. Arch Clin Neuropsychol 2014 Dec;29(8):767-775. [doi: 10.1093/arclin/acu055] [Medline: 25352087]
26. Woods DL, Wyma JM, Yund EW, Herron TJ, Reed B. Factors influencing the latency of simple reaction time. Front Hum Neurosci 2015;9:131 [FREE Full text] [doi: 10.3389/fnhum.2015.00131] [Medline: 25859198]
27. Vaportzis E, Clausen MG, Gow AJ. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
28. Midaglia L, Mulero P, Montalban X, Graves J, Hauser SL, Julian L, et al. Adherence and Satisfaction of Smartphone- and Smartwatch-Based Remote Active Testing and Passive Monitoring in People With Multiple Sclerosis: Nonrandomized Interventional Feasibility Study. J Med Internet Res 2019 Aug 30;21(8):e14863 [FREE Full text] [doi: 10.2196/14863] [Medline: 31471961]
29. Benedict RH, Wahlig E, Bakshi R, Fishman I, Munschauer F, Zivadinov R, et al. Predicting quality of life in multiple sclerosis: accounting for physical disability, fatigue, cognition, mood disorder, personality, and behavior change. J Neurol Sci 2005 Apr 15;231(1-2):29-34. [doi: 10.1016/j.jns.2004.12.009] [Medline: 15792817]
30. Jorm AF, Anstey KJ, Christensen H, Rodgers B. Gender differences in cognitive abilities: The mediating role of health state and health habits. Intelligence 2004 Jan;32(1):7-23. [doi: 10.1016/j.intell.2003.08.001]
31. Wojcik CM, Rao SM, Schembri AJ, Drake AS, Maruff P, Schindler D, et al. Necessity of technicians for computerized neuropsychological assessment devices in multiple sclerosis. Mult Scler 2020 Jan;26(1):109-113. [doi: 10.1177/1352458518813287] [Medline: 30465463]
32. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol 2012 May;27(3):362-373 [FREE Full text] [doi: 10.1093/arclin/acs027] [Medline: 22382386]

Abbreviations

ANOVA: analysis of variance
CoRe: Cognitive Reaction
c-SDMT: computerized Symbol Digit Modalities Test
EDSS: Expanded Disability Status Score
MS: multiple sclerosis
NHS: National Health Service
PPMS: primary progressive multiple sclerosis
PRO: patient-reported outcome
PST: processing speed test
QAV: question answering velocity
RRMS: relapsing-remitting multiple sclerosis
SDMT: Symbol Digit Modalities Test
SPMS: secondary progressive multiple sclerosis
UKMSR: United Kingdom Multiple Sclerosis Register

Edited by G Eysenbach; submitted 13.02.20; peer-reviewed by T Muto, D Golan, P Wipfler, N Nazari; comments to author 12.06.20; revised version received 06.07.20; accepted 02.08.20; published 23.09.20.

Please cite as:
Middleton RM, Pearson OR, Ingram G, Craig EM, Rodgers WJ, Downing-Wood H, Hill J, Tuite-Dalton K, Roberts C, Watson L, Ford DV, Nicholas R, UK MS Register Research Group
A Rapid Electronic Cognitive Assessment Measure for Multiple Sclerosis: Validation of Cognitive Reaction, an Electronic Version of the Symbol Digit Modalities Test
J Med Internet Res 2020;22(9):e18234
URL: http://www.jmir.org/2020/9/e18234/
doi:10.2196/18234
PMID:32965240

Original Paper

Mobile Insight in Risk, Resilience, and Online Referral (MIRROR): Psychometric Evaluation of an Online Self-Help Test

Merel Marjolein van Herpen[1,2], MA; Manon A Boeschoten[3*], PhD; Hans te Brake[1*], PhD; Niels van der Aa[3], PhD; Miranda Olff[2,4], PhD

[1]ARQ Centre of Expertise for the Impact of Disasters and Crises, Diemen, Netherlands

[2]Department of Psychiatry, Amsterdam Neuroscience & Public Health, Amsterdam University Medical Center, Amsterdam, Netherlands

[3]ARQ Centre '45, Oegstgeest/Diemen, Netherlands

[4]ARQ National Psychotrauma Centre, Diemen, Netherlands

[*]these authors contributed equally

Corresponding Author:
Merel Marjolein van Herpen, MA
ARQ Centre of Expertise for the Impact of Disasters and Crises
Nienoord 5
Diemen,
Netherlands
Phone: 31 610082023
Email: m.van.herpen@impact.arq.org

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e29776

Abstract

Background: Most people who experience a potentially traumatic event (PTE) recover on their own. A small group of individuals develops psychological complaints, but this is often not detected in time or guidance to care is suboptimal. To identify these individuals and encourage them to seek help, a web-based self-help test called Mobile Insight in Risk, Resilience, and Online Referral (MIRROR) was developed. MIRROR takes an innovative approach since it integrates both negative and positive outcomes of PTEs and time since the event and provides direct feedback to the user.

Objective: The goal of this study was to assess MIRROR's use, examine its psychometric properties (factor structure, internal consistency, and convergent and divergent validity), and evaluate how well it classifies respondents into different outcome categories compared with reference measures.

Methods: MIRROR was embedded in the website of Victim Support Netherlands so visitors could use it. We compared MIRROR's outcomes to reference measures of PTSD symptoms (PTSD Checklist for DSM-5), depression, anxiety, stress (Depression Anxiety Stress Scale–21), psychological resilience (Resilience Evaluation Scale), and positive mental health (Mental Health Continuum Short Form).

Results: In 6 months, 1112 respondents completed MIRROR, of whom 663 also completed the reference measures. Results showed good internal consistency (interitem correlations range .24 to .55, corrected item-total correlations range .30 to .54, and Cronbach alpha coefficient range .62 to .68), and convergent and divergent validity (Pearson correlations range –.259 to .665). Exploratory and confirmatory factor analyses (EFA+CFA) yielded a 2-factor model with good model fit (CFA model fit indices: χ^2_{19}=107.8, P<.001, CFI=.965, TLI=.948, RMSEA=.065), conceptual meaning, and parsimony. MIRROR correctly classified respondents into different outcome categories compared with the reference measures.

Conclusions: MIRROR is a valid and reliable self-help test to identify negative (PTSD complaints) and positive outcomes (psychosocial functioning and resilience) of PTEs. MIRROR is an easily accessible online tool that can help people who have experienced a PTE to timely identify psychological complaints and find appropriate support, a tool that might be highly needed in times like the coronavirus pandemic.

(J Med Internet Res 2020;22(9):e19716) doi:10.2196/19716

KEYWORDS

potentially traumatic events; mobile mental health; self-help; online; resilience; posttraumatic stress disorder

Introduction

Most people will experience at least one potentially traumatic event (PTE) in their lives [1-5]. The impact of PTEs is not the same for every individual. Research shows that most individuals are able to maintain a healthy level of functioning or resilience after experiencing a PTE and psychological complaints usually diminish over time without professional support [1,6-10]. However, a small but significant group of individuals develops psychological complaints such as posttraumatic stress disorder (PTSD) that require care [2].

Experiencing psychological complaints a few days to weeks after a PTE is often considered normal [11-13]. The National Institute for Health and Care Excellence (NICE) advises to consider active monitoring—also known as watchful waiting—following a PTE (ie, regular monitoring of people with some PTSD symptoms within 1 month of the event) [14]. The European Network for Traumatic Stress (TENTS) guideline for post-disaster psychosocial care advises against formal screening of everyone affected by a PTE but stresses the importance of identifying individuals in need of support. Once PTSD has been diagnosed, early treatment is advised [14-18]. It could be concluded, then, that support for people who have experienced a PTE is necessary, preferably early, and easily accessible.

Unfortunately, the small but significant group that develops persisting psychological complaints is often not detected in time or guidance to care is suboptimal [19,20]. Guidance to care can be hindered due to people not recognizing their symptoms or having self-stigma, which prevents them from seeking help [21-24]. In addition, health care facilities may lack the resources to be able to reach people who have experienced a PTE and identify the ones who need support [23,25]. Also, general practitioners may not recognize PTSD symptoms [26] or other psychological complaints [27].

In order to prevent the development and persistence of trauma-related complaints, timely and accurate identification is needed [23,28]. Short and easy-to-use screening instruments could enable individuals at risk of developing psychological complaints to self-identify and monitor possible symptoms after PTEs. Moreover, providing online or mobile self-help tests can aid in timely identification of symptoms in people who have experienced a PTE, providing more information regarding normal psychological responses and encouraging help seeking [29,30].

Multiple studies show that when one chooses to assist people who have experienced a PTE, it is important to support self-reliance and resilience [1,11,14]. Normalizing and validating emotional responses can promote the capacity to deal with these emotions [11]. Also, the extent to which individuals identify themselves as being resilient is considered to positively influence post-trauma outcomes [31,32]. Several self-report screening instruments are available to predict PTSD, such as the Trauma Screening Questionnaire, Impact of Event Scale–Revised or PTSD Checklist for the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (PCL-5) [33,34]. However, most instruments only screen for complaints and do not inquire about protective factors such as psychological resilience and psychosocial functioning [33,34]. In addition, most screening instruments do not consider the time period that has passed since the event. Such information is necessary to determine whether reported complaints can be appraised as normal given the stressful event just happened or whether referral to care is needed [14]. By not including time in classifying responses, screening can overlook or misappraise the different response trajectories that have been found after PTEs [9].

To incorporate above guideline advice and address the aforementioned concerns in the early support of people who have experienced a PTE, Mobile Insight in Risk, Resilience, and Online Referral (MIRROR) was developed. MIRROR is a web-based self-help test with the potential to reach large groups of people who are seeking reassurance on how they are coping. MIRROR takes an innovative approach since it integrates both negative and positive outcomes of PTEs and time since the event. This was realized by creating a new questionnaire based on existing measures on resilience, functioning, and PTSD, and by developing a new algorithm that takes into account multiple factors. In compliance with NICE, TENTS, and DSM-5 guidelines [14,15,35], MIRROR's algorithm includes the following as main weight factors: severity of complaints, time passed since event, and level of psychosocial functioning. MIRROR provides users with personal advice based on respondent answers with relevant follow-up support options such as a reminder for self-monitoring and contact information for consultation. Giving personal feedback to users is recommended to augment the use of mobile self-tests after PTEs [36]. Also, arranging active monitoring with follow-up within 1 month is advised [14]. Of relevance, no difference has been found between responses on a PTSD self-report administered via a mobile device versus paper administration [37]. MIRROR aims to contribute to the early identification of those likely to develop psychological complaints and encourage them to seek help. At the same time, MIRROR aims to support self-reliance by facilitating self-monitoring and self-recovery through follow-up support options.

While it is recognized that mobile apps have the potential to improve timely identification of complaints and delivery of mental health support after PTEs, there is very little research on their validity, reliability, and effectiveness [29,30,38,39]. Therefore, the aims of this study were to assess MIRROR's use, examine MIRROR's psychometric properties (factor structure, internal consistency, and convergent and divergent validity) and evaluate how well MIRROR classifies respondents into different outcome categories compared with reference measures.

Methods

Mobile Insight in Risk, Resilience, and Online Referral (MIRROR)

A multidisciplinary team of professionals in the fields of psychotrauma (clinicians, researchers, and policy officers) and victim and crisis support developed MIRROR. The items and algorithm were based on existing protocols—DSM-5 and the *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (ICD-10) [35,40]—best practices and recommendations of the Dutch National Multidisciplinary Guideline on Psychosocial Support in Disasters and Crises [41], and international guidelines for PTSD and postdisaster psychosocial care [14,15].

MIRROR consists of 2 parts. Part 1 includes items regarding event-related characteristics: type of event, measured with all events of the Dutch version of the Life Events Checklist for the DSM-5 (LEC-5) [42], time passed since the event (measured in weeks), and relation to the event (happened to me, learned about it, witnessed it, part of my job). Part 2 consists of 8 items divided in 3 sections. The first concerns PTSD core symptoms (4 items in total; 1 about intrusion, 2 about avoidance, and 1 about arousal). The items are developed based on the clusters in the DSM-IV, DSM-5, ICD-10, and ICD-11. Higher scores reflect more PTSD symptoms. The second concerns the item "how would you rate your present functioning (at work/home)," based on the widely used Global Assessment of Functioning (GAF) score for which higher scores reflect a higher level of functioning. The third concerns resilience (3 items in total; about social support, self-reliance, and problem solving), based on the resilience concept as introduced by Van der Meer et al [43]. Higher scores reflect more resilience. PTSD and resilience items are answered on a 5-point response scale, ranging from 1 (never) to 5 (all the time). Functioning is rated on a scale from 1 to 10.

MIRROR's algorithm aims to identify PTSD symptoms, psychosocial functioning, and resilience; normalize complaints (ie, reassuring users that it is normal to experience distress shortly after a PTE); and stimulate seeking support in users with persisting complaints. See Multimedia Appendix 1 for an overview of the possible outcomes of the algorithm. In the algorithm, MIRROR's PTSD scale and functioning item are classified in 3 levels: low, moderate, and high. Resilience is categorized as either low or high. The categorizations are based on the aforementioned existing protocols and best practices. MIRROR's algorithm differentiates 3 phases of time passed since the event: (1) less than 1 week ago, (2) between 1 and 4 weeks, and (3) more than 4 weeks or reoccurring. These were based on the assumption that complaints after PTEs may occur but will generally diminish over time, as most people recover on their own [6]. Therefore, the occurrence of PTSD core complaints with moderate to low functioning shortly after an adverse event can be seen as normal [11-13], but if complaints and moderate to low functioning are present after 1 month, guidance to care is needed [14-18].

MIRROR summarizes the outcome of its algorithm to respondents as either green, orange, or red. Together with this color outcome, respondents receive personal advice. The color outcome is based on the level of complaints, functioning, and time passed since the event. MIRROR's resilience scale is not included in the color outcome because based on current research it is unclear precisely how resilience interacts with the development of PTSD complaints and functioning after PTEs. Nonetheless, resilience is integrated in the personal advice to stimulate the use of social support. If respondents score low on resilience they are encouraged to seek support from those close to them and individuals who have experienced similar events.

A green outcome indicates few complaints and/or sufficient functioning, and the accompanying advice states no further action is needed. An orange outcome indicates complaints and moderate functioning in combination with a PTE that happened only recently (ie, less than 1 month). The accompanying advice is directed at normalizing complaints combined with promoting watchful waiting and encouraging setting a reminder to use MIRROR again in 2 weeks to assess if complaints have diminished. The red outcome indicates significant complaints (ie, low functioning or complaints with moderate to low functioning for a longer period or due to a reoccurring event) which have persisted for more than 1 month. Therefore, the advice aims to encourage the user to seek consultation with a general practitioner or to contact Victim Support Netherlands. MIRROR provides follow-up support options with its advice, such as the opportunity to get in touch with people who have had similar experiences, reading information about dealing with stress reactions, or setting a reminder to use MIRROR again in 2 weeks.

Participants and Procedure

MIRROR was available in the Dutch language and open for each visitor on the website of Victim Support Netherlands (Slachtofferhulp Nederland). The specifically targeted sample consisted of website visitors who were automatically led to MIRROR when searching for information regarding stress reactions following a PTE. MIRROR is a responsive website; respondents did not have to download it. MIRROR can be used on mobile and nonmobile devices. To evaluate the psychometric properties of MIRROR, we added a research survey with reference measures (see details in Measures) after the MIRROR questions. Data collection took place during a period of 6 months. We tested the usability and technical functionality of MIRROR and the research survey before making it available. Each item was presented on a new webpage.

Before starting MIRROR, respondents were invited to participate in the research survey. Participants were informed regarding the purpose of the study, duration time of the survey, and data storage. Participation was voluntary and completely anonymous. Respondents received no incentive for completing MIRROR or the research survey. They were asked for informed consent to use their data for research purposes, in accordance with the European General Data Protection Regulation. The Medical Ethical Committee of Amsterdam University Medical Center exempted this study from formal review (W18_364 #18.435).

Data collection took place between February and August 2019. Only original answers were saved in the database. That is, if respondents went back to change their answers once they already

received their advice, changes were not saved. We followed data cleaning recommendations by Birnbaum [44] and Wood et al [45]. Data were discarded when respondents did not complete all survey items. In case of identical answers on all items of the different reference measures, other systematic answering patterns, or obvious unusual missing answers on certain measures, we reviewed individual results thoroughly and discarded the data in case of doubt.

Measures

Posttraumatic Stress Disorder Symptoms

To measure PTSD symptoms, we used the Dutch version of the PCL-5 [46,47]. The PCL-5 consists of 20 items and measures symptoms of intrusion (cluster B, 5 items), avoidance (cluster C, 2 items), negative alterations in cognitions and mood (cluster D, 7 items), and alterations in arousal and reactivity (cluster E, 6 items) in the past month. All items are answered on a 5-point scale, ranging from 0 (not at all) to 4 (extremely). The PCL-5 showed good psychometric properties in different languages [48-50]. The total score was calculated by adding all item scores. Scale scores per cluster were calculated by adding the scores of the corresponding items. Higher scores reflect more severe symptoms. Cronbach alphas in our sample ranged between .77 and .86 for the B, C, D, and E clusters. The DSM-5 rule to determine a provisional PTSD diagnosis was followed. This entails treating each item with a minimum score of 2 as a symptom endorsed and requiring at least one B symptom, one C symptom, two D symptoms, and two E symptoms [46].

Depression, Anxiety, and Stress

To assess other common psychological complaints after PTEs, we used the Dutch short version of the Depression Anxiety Stress Scale (DASS-21) measuring depression (7 items), anxiety (7 items), and stress (7 items) [51,52]. The DASS-21 is a valid and reliable measure [53,54]. Item scores were summed to calculate scale scores and the total score. Higher scores reflect more severe symptoms. In our sample, Cronbach alphas were .92, .86, and .86 for depression, anxiety, and stress scales, respectively. A 4-point response scale measures the extent to which each state has been experienced over the past week ranging from 0 (did not apply to me at all) to 3 (applied to me very much, or most of the time). To determine cutoff values, DASS-21 scale scores were multiplied by two, in accordance with the scale's manual [52]. The manual provides cutoff scores for a Dutch clinical sample. These discriminate the following categories: normal (depression <9, anxiety <7, stress <14), mild (depression 10-13, anxiety 8-9, stress 15-18), moderate (depression 14-20, anxiety 10-14, stress 19-25), severe (depression 21-27, anxiety 15-19, stress 26-33) and extremely severe (depression >28, anxiety >20, stress >34).

Psychological Resilience

We used the Resilience Evaluation Scale (RES) to assess psychological resilience [43]. The 9 items are rated on a 5-point scale ranging from 0 (strongly disagree) to 4 (strongly agree). We calculated the total score by adding all items. Higher scores reflect more psychological resilience. The RES is a valid and reliable measure [43]. In this sample, Cronbach alpha of the total scale was .88.

Positive Mental Health

We assessed positive mental health with the Dutch version of the Mental Health Continuum Short Form (MHC-SF) [55,56]. The MHC-SF measures emotional well-being (3 items), social well-being (5 items), and psychological well-being (6 items). Items were rated on a 6-point scale ranging from 0 (never) to 5 (every day). The MHC-SF is a valid and reliable instrument [56,57]. We calculated the total score by summing all item scores. Higher scores reflect more positive mental health. In this sample, Cronbach alpha of the total scale was .93.

Google Analytics

Google Analytics data were collected between March and August 2019 to examine MIRROR's use. Due to technical problems, data from February 2019 were missing. The data provide information on the number of unique visits per page, type of device used, and number of visitors who have started MIRROR (defined as a unique page visit on MIRROR's start page) and who have finished MIRROR (defined as unique page visit on MIRROR's outcome and advice page). Google Analytics cannot determine to what extent the follow-up options were used, but it can detect how many respondents have visited the follow-up support option pages.

Statistical Analyses

Sample and Use

Since participation in the research survey was optional, this resulted in two 2 samples. The MIRROR-only sample consists of respondents who only completed MIRROR. The validation sample includes respondents who completed MIRROR and the accompanying survey with reference measures before receiving their advice. The total sample combines these two samples, consisting of all respondents. To examine if the validation sample was representative of the MIRROR user, we used independent-samples *t* tests in SPSS Statistics version 23 (IBM Corporation) to compare the MIRROR-only sample with the validation sample based on their MIRROR scores and event-related characteristics.

We used the total sample to evaluate MIRROR's use and examine MIRROR's factor structure and internal consistency because for these analyses only data from MIRROR were needed. We used the validation sample to examine MIRROR's convergent and divergent validity and evaluate how well MIRROR classifies respondents into different outcome categories because for these analyses data from MIRROR as well as reference measures from the accompanied survey were needed.

Factor Structure

We used Mplus version 8 (Muthen & Muthen) to conduct exploratory factor analysis (EFA) using geomin rotation and confirmatory analysis (CFA). EFA assumes that any item may be associated with any factor. CFA specifies expected relationships between items and their underlying latent factors. Because items of MIRROR's PTSD and resilience section were categorical, they were treated as ordinal and therefore the means and variance adjusted weighted least square (WLSMV) estimator was used. An underlying normal distribution was assumed for

each ordinal item, where the 5 response categories were divided by 4 thresholds estimated from the data. MIRROR's functioning item has 10 response categories and was treated as continuous. Because MIRROR's factor structure was not tested before, several models with different numbers of latent factors were examined using EFA. To assess the model with the optimal number of latent factors needed to adequately account for the correlations among item scores, we used Kaiser criterion (ie, eigenvalues of the latent factors >1) and model fit statistics. The model with the best balance between model fit, parsimony, and conceptual interpretability was selected as the most optimal model. Subsequently, CFA was used to test the optimal model based on EFA. The difference in goodness-of-fit between nested models was evaluated with the difftest option in Mplus for appropriate chi-square difference testing with the WLSMV estimator [58]. The chi-square difference test is highly sensitive to sample size such that even trivial differences between two nested models may be significant [59]. Therefore, we also assessed the difference in comparative fit index (CFI). A difference in CFI <0.01 indicates a better fit of the nested model compared with the more complex model [59]. For EFA and CFA, the model fit indices CFI, Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA) were used to evaluate model fit. Model fit can be considered good when CFI and TLI are close to .95, and RMSEA <.06 [60]. If RMSEA <.08, model fit can be considered adequate [60].

Internal Consistency

We evaluated internal consistency of MIRROR's PTSD and resilience section with interitem correlations, corrected item-total correlations, and Cronbach alpha in SPSS Statistics version 23. Internal consistency of MIRROR's functioning section could not be evaluated since it is represented by only one item. When most interitem correlations are in the recommended range of .15 to .50 (moderate magnitude) and Cronbach alpha for the scale is >.80, internal consistency can be considered good [61]. Cronbach alpha is a function of scale length and therefore is likely to be lower for MIRROR's scales since they consist of 3 or 4 items [61]. Corrected item-total correlations were computed to assess whether item scores regarding PTSD and resilience are associated with overall PTSD and resilience scores.

Convergent and Divergent Validity

To evaluate MIRROR's convergent and divergent validity, we calculated Pearson correlations between the MIRROR scales and reference measures. Convergent and divergent validity can be considered good when the correlations between a scale and equivalent measure (eg, MIRROR's PTSD scale and the PTSD scale of the PCL-5) are significant and high while correlations between this scale and other related measures (eg, MIRROR's PTSD scale and depression scale of the DASS-21) are lower and moderate or modest in magnitude.

Classification Quality

To evaluate how well MIRROR classifies respondents into a red, orange, or green outcome, we tested whether respondents in these three outcome categories differed on related reference measures by using cross-tabs and analysis of variance (ANOVA). If the assumption of equal variances was violated, we used the Welch *F*-test and Games-Howell post hoc test. MIRROR's PTSD scale score was calculated by summing the 4 PTSD items. Higher scores reflect more severe symptoms. MIRROR's resilience scale score was calculated by a summing the 3 items. Higher scores reflect more resilience. Provisional PTSD diagnosis based on PCL-5 were used to classify respondents. To examine the distribution on depression, anxiety and stress symptoms, respondents were classified by comparing their scores to a Dutch clinical reference group. Respondents with normal and mild complaints compared with the reference group were classified into one group representing subclinical complaints. Respondents with average, severe, and very severe complaints compared with the reference group were classified into another group, representing clinical complaints. Since no reference groups were available with regard to the RES and MHC-SF, the sample was divided into tertiles (ie, 3 groups of equal size divided by the 33rd and 66th percentile) based on the total scores of the RES and MHC-SF. With regard to the RES, the first tertile (scores ≤17) was assumed to represent relatively low psychological resilience, the second tertile (scores from 18 to 24) relatively moderate psychological resilience, and the third tertile (scores ≥25) relatively high psychological resilience. With regard to the MHC-SF, the first (scores ≤23), second (scores from 24 to 47), and third tertile (scores ≥48) were assumed to represent relatively low, moderate, and high positive mental health, respectively.

Results

Sample and Use

MIRROR was completed 1314 times in the study period of 6 months. In total, 51.90% (682/1314) of respondents started the research survey. We deleted 51 respondents who indicated they used MIRROR on behalf of a family member, partner, friend, or colleague who experienced a PTE. We deleted 37 repeated measurements, completed by respondents who set a reminder. We excluded 95 respondents because they did not complete all research survey items. After thorough investigation of the answering patterns, we deleted 19 respondents because of unusual answering patterns. A total of 84.63% (1112/1314) of respondents were retained in the total sample, of whom 59.62% (validation sample, 663/1112) also completed all questionnaires of the accompanying research survey.

Table 1 presents the MIRROR scores, outcomes, and event-related characteristics for the MIRROR-only and validation sample. We found no significant difference between the samples on MIRROR's PTSD scale: t_{1110}=–.401, *P*=.69; resilience scale: t_{1110}=.752, *P*=.45; or level of functioning t_{1110}=1.547, *P*=.12. We found a significant association between sample and MIRROR outcome: $\chi^2_{2,n=1112}$=18.99, *P*<.001; the validation sample consisted of more respondents with the red MIRROR outcome than the MIRROR-only sample. The event-related characteristics for both samples were similar, see Table 1. Overall, the validation sample can be considered representative of all MIRROR users in this study period. In the validation sample, 74.2% (492/663) of respondents were female. Almost half (300/663, 45.3%) of respondents were aged between

21 and 40 years. Tables 2 and 3 present the frequency distributions for MIRROR's response categories.

Table 1. Mobile Insight in Risk, Resilience and Online Referral (MIRROR) scores, outcomes, and event-related characteristics for the validation sample and MIRROR-only sample.

MIRROR[a]	Validation[b] (n=663)	MIRROR[b] only (n=449)
MIRROR scores, mean (SD)		
MIRROR PTSD[c] scale	14.88 (3.39)	14.80 (3.28)
MIRROR functioning	4.92 (1.96)	5.11 (1.94)
MIRROR resilience scale	10.08 (2.36)	10.91 (2.37)
MIRROR[a] outcome[b], n (%)		
Red	409 (61.7)	224 (49.9)
Orange	230 (34.7)	214 (47.7)
Green	24 (3.6)	11 (2.4)
Type of event (LEC-5[d]), n (%)		
Another very stressful event or experience	216 (32.6)	150 (33.4)
Transportation accident	115 (17.4)	107 (23.8)
Physical assault	109 (16.5)	50 (11.1)
Sudden accidental death	38 (5.7)	20 (4.5)
Serious accident at work, home, or during recreation	33 (5.0)	28 (6.2)
Sexual assault	33 (5.0)	18 (4.0)
Assault with a weapon	30 (4.5)	25 (5.6)
Other unwanted or uncomfortable sexual experience	30 (4.5)	14 (3.1)
Sudden violent death	24 (3.6)	16(3.6)
Severe human suffering	14 (2.1)	5 (1.1)
Life-threatening illness or injury	10 (1.5)	5 (1.1)
Fire or explosion	9 (1.4)	4 (0.9)
Combat or exposure to a war zone	1 (0.2)	0 (0)
Captivity	0 (0)	4 (0.9)
Serious injury, harm, or death caused by you to someone else	0 (0)	3 (0.7)
Natural disaster	0 (0)	0 (0)
Relation to the event, n (%)		
Event happened to me	480 (72.5)	311 (69.3)
I witnessed the event	129 (19.5)	94 (20.9)
I learned about the event	42 (6.3)	35 (7.8)
Other[e]	11 (1.7)	9 (2.0)
Work-related, n (%)		
No	586 (88.4)	379 (84.4)
Yes	77 (11.6)	70 (15.6)
Time since the event, n (%)		
Less than 1 week	241 (36.3)	218 (48.6)
Over 4 weeks	214 (32.3)	113 (25.2)
Between 1 and 4 weeks	144 (21.7)	90 (20.0)
It happens repeatedly	64 (9.7)	28 (6.2)

[a]MIRROR: Mobile Insight in Risk, Resilience, and Online Referral.

[b]Significant association between sample and MIRROR outcome, $P<.001$.

[c]PTSD: posttraumatic stress disorder.

[d]LEC-5: Life Events Checklist for DSM-5.

[e]If respondents could not select one of the event relations (happened to me, witnessed it, learned about it, work-related), they are asked to specify their relation to the event.

Table 2. Frequency distribution in percentages of Mobile Insight in Risk, Resilience and Online Referral (MIRROR) item response categories, items 1-4 and 6-8 (n=1112).

Scale and item number	Never	Rarely	Sometimes	Often	All the time
PTSD[a]					
1	2.7	5.7	16.6	38.5	36.5
2	5.1	8.5	19.3	27.4	39.6
3	9.3	13.8	26.9	22.9	27.1
4	8.5	11.4	26.7	26.8	26.6
Resilience					
6	5.3	8.5	21.7	35.3	29.3
7	7.3	15.6	35.2	30.5	11.5
8	5.2	15.1	45.6	28.1	5.9

[a]PTSD: posttraumatic stress disorder.

Table 3. Frequency distribution in percentages of Mobile Insight in Risk, Resilience and Online Referral (MIRROR) item response categories, item 5 (n=1112).

Scale and item number	1	2	3	4	5	6	7	8	9	10
Functioning										
5	4.9	6.9	10.4	15.8	19.8	20.9	12.1	6.3	1.6	1.3

A detailed overview of the scores of the validation sample on the reference measures can be found in Multimedia Appendix 2. Overall, these show a high level of complaints in our sample and rather low levels of psychological resilience and positive mental health (also see Table 7 and Figure 1 for reference measures of each MIRROR outcome category).

Google Analytics data provided insight into MIRROR's use. The number of visitors who started MIRROR was 2555, of whom 2247 (87.95%) finished it. The original database contained 1314 entries. This discrepancy can be explained by users having the opportunity to refuse to have their data saved before starting. Of all users, 47.59% (1216/2555) chose this option. Furthermore, of the follow-up support options, the "seek contact with Victim Support Netherlands" page had most views (411 unique views), followed by "more information" (293 unique views), "send your advice to yourself or someone else" (235 unique views), "seek contact with people who have had similar experiences" (209 unique views), and "set a reminder" (161 unique views). A total of 28.7% (113/394) of respondents who received the orange outcome and were advised to complete MIRROR again in 2 weeks immediately set a reminder to complete MIRROR again in 2 weeks. A total of 22.1% (25/113) did so at the time of data analyses. The most often used device was the smartphone (1566/2555, 61.29%), followed by desktop (794/2555, 31.08%), and tablet (195/2555, 7.63%).

Factor Structure

Table 4 presents the factor loadings for the 2-factor and 3-factor solution model of MIRROR as estimated by EFA. EFA yielded a 3-factor solution with good model fit based on all fit indices. The Kaiser criterion was met for the first 2 factors, eigenvalues of the third through eighth factor were <1. The 3-factor solution separated MIRROR's PTSD items into 2 factors: 1 with the intrusion item and 1 with the avoidance and arousal/reactivity items. However, item 2 ("have you become jumpy and/or vigilant since the event?") cross-loaded significantly on 2 factors within the model, with only a small difference between the 2 factor loadings (λ=0.030). This indicates that item 2 did not sufficiently distinguish between both factors. The 3-factor solution clustered the functioning item with the resilience items into a third factor.

Table 4. Geomin rotated factor loadings for the 2-factor and 3-factor solution model of Mobile Insight in Risk, Resilience and Online Referral (MIRROR) as estimated by exploratory factor analysis (n=1112).

MIRROR[a] items	2-factor solution[b]		3-factor solution[c]		
	F1	F2	F1	F2	F3
1. Are you troubled by images of or thoughts about the event?[d]	0.525*	–0.004	0.813*	0.015	0.018
2. Have you become jumpy and/or vigilant since the event?[e]	0.585*	–0.009	0.308*	0.338*	–0.012
3. Do you try to avoid things that are related to the event?[f]	0.789*	0.071	–0.000	1.078*	0.245*
4. Do you try to avoid thinking about the event?[g]	0.648*	–0.016	0.208*	0.459*	–0.019
5. How would you rate your present functioning (at work/home)?[h]	–0.153*	0.354*	–0.213*	0.004	0.360*
6. Do you experience support from those close to you?[i]	0.081*	0.388*	0.160*	–0.064	0.374*
7. Are you confident in yourself?[j]	0.006	0.827*	0.010	–0.021	0.827*
8. Are you able to deal with any problems you encounter?[k]	–0.015	0.730*	–0.074	0.018	0.718*

[a]Mobile Insight in Risk, Resilience and Online Referral.

[b]Model fit indices for the 2-factor solution: χ^2_{13}=88.7, *P*<.001, CFI=.969, TLI=.933, RMSEA=.072.

[c]Model fit indices for the 3-factor solution: χ^2_7=12.6, *P*=.084, CFI=.998, TLI=.991, RMSEA=.027.

[d]Eigenvalue 2.777,

[e]Eigenvalue 1.466.

[f]Eigenvalue .927.

[g]Eigenvalue .715.

[h]Eigenvalue .668.

[i]Eigenvalue .640.

[j]Eigenvalue .437.

[k]Eigenvalue .369.

*P<.05.

EFA yielded a 2-factor solution with adequate model fit. The RMSEA and TLI indicated adequate model fit and CFI indicated good model fit (Table 4). The Kaiser criterion was met for the first 2 factors; eigenvalues of the third through eighth factor were <1. The first factor of the 2-factor solution consisted of the PTSD items and the second factor consisted of the functioning and resilience items. No cross-loadings were observed in this model.

Next, we conducted CFA to further compare the 2- and 3-factor model that resulted from EFA. Table 5 presents the model fit indices based on CFA of both aforementioned models. The model fit indices were similar for both models; the CFI and TLI indicated good model fit, the RMSEA acceptable model fit. As indicated by the significant χ^2 difference test, the 2-factor model has worse model fit compared with the 3-factor model ($\chi^2_{2,n=1112}$=13.63, *P*=.001). However, the difference in CFI is <0.01, indicating the 2-factor model does not have worse model fit. We selected the 2-factor model as the best-fitting model to our data, given the χ^2 difference test is sensitive to sample size, the CFI difference is <.001, and it is more parsimonious and better interpretable at a conceptual level compared with the 3-factor model. The 2-factor model represents a clear distinction between negatively formulated outcomes (PTSD complaints) and positively formulated outcomes (psychosocial functioning and resilience) of PTEs. The positively formulated outcomes combine psychosocial functioning, social support, self-reliance and problem solving. We therefore propose to rename this factor psychosocial resources.

Table 5. Confirmatory factor analysis model fit indices (n=1112).

Model	χ^2	*P* value	df[a]	CFI[b]	TLI[c]	RMSEA[d]
Two-factor solution	107.78	<.001	19	0.965	0.948	0.065
Three-factor solution	95.868	<.001	17	0.969	0.949	0.064

[a]df: degree of freedom.

[b]CFI: comparative fit index.

[c]TLI: Tucker-Lewis index.

[d]RMSEA: root mean square error of approximation.

Internal Consistency

Interitem correlations of MIRROR's PTSD complaints scale ranged between .28 and .48 with a mean of .34. All of the interitem correlations of the PTSD scale were in the recommended range of moderate magnitude of .15 to .50, indicating that this scale has high internal consistency in combination with a differentiated item set. Corrected item-total correlations for this scale ranged between .39 and .54 with a mean of .46, indicating that high scores on the PTSD items are associated with high scores on the overall PTSD scale of MIRROR. Cronbach alpha coefficient for MIRROR's PTSD scale was .68.

Interitem correlations of MIRROR's resilience scale ranged between .24 and .55, with a mean of .36. In addition, 1 out of 3 interitem correlations was higher than the recommended range of moderate magnitude of .15 to .50 (between "are you confident in yourself" and "are you able to deal with any problems you encounter"), indicating that this scale has high internal consistency in combination with a differentiated item set. Corrected item-total correlations ranged between .30 and .52 with a mean of .44, indicating that high scores on the resilience items are associated with high scores on the overall resilience scale of MIRROR. Cronbach alpha coefficient for MIRROR's resilience scale was .62.

Convergent and Divergent Validity

Pearson correlations between MIRROR and reference measures are presented in Table 6. MIRROR's PTSD scale showed strongest correlations with PTSD as measured with the PCL-5, followed by a lower but still substantial correlation with psychological complaints as assessed with the DASS-21. The weakest correlations were observed between PTSD symptom severity as assessed with MIRROR and psychological resilience and positive mental health. MIRROR's resilience scale showed strongest correlation with psychological resilience (RES), followed by a slightly lower correlation with positive mental health, psychological complaints (DASS-21), and PTSD (PCL-5). MIRROR's functioning item showed strongest correlations with psychological complaints (DASS-21) followed by PTSD (PCL-5) with lower correlations with positive mental health (MHC-SF) and psychological resilience (RES). In conclusion, the correlational structure indicates good convergent and divergent validity of MIRROR's PTSD subscale. The correlational structure with regard to MIRROR's resilience scale and functioning item indicates adequate convergent and divergent validity.

Table 6. Correlations between Mobile Insight in Risk, Resilience and Online Referral (MIRROR) subscales and reference measures (n=663).

MIRROR	PTSD[a]	*P* value	Resilience	*P* value	Functioning	*P* value
PCL-5[b]	.665	<.001	–.507	<.001	–.442	<.001
DASS-21[c]	.486	<.001	–.539	<.001	–.449	<.001
RES[d]	–.265	<.001	.612	<.001	.279	<.001
MHC-SF[e]	–.259	<.001	.603	<.001	.319	<.001

[a]PTSD: posttraumatic stress disorder.

[b]PCL-5: PTSD Checklist for DSM-5.

[c]DASS-21: Depression Anxiety Stress scale.

[d]RES: Resilience Evaluation Scale.

[e]MHC-SF: Mental Health Continuum Short Form.

Classification Quality

We expected respondents with the red MIRROR outcome to report more PTSD symptoms and depression, anxiety, and stress complaints; lower psychological resilience; and positive mental health compared with respondents with the green and orange MIRROR outcome. Table 7 presents the means and standard deviations on the reference measures for each MIRROR outcome category. Figure 1 shows the classification percentages on reference measures for each MIRROR outcome category. Both Table 7 and Figure 1 show that respondents with the red MIRROR outcome category report higher complaints and lower psychological resilience and positive mental health compared with the orange and green MIRROR outcome category.

Table 7. Means and standard deviations of reference measures for each Mobile Insight in Risk, Resilience and Online Referral (MIRROR) outcome category (n=663).

MIRROR[a] outcome category (n)	Green (n=24), mean (SD)	Orange (n=200), mean (SD)	Red (n=439), mean (SD)
PTSD[b] (PCL-5[c])	18.04 (12.49)	36.09 (15.77)	46.13 (14.04)
Depression (DASS-21[d])	4.08 (8.10)	11.73 (11.54)	19.66 (11.54)
Anxiety (DASS-21)	5.25 (6.72)	14.03 (10.27)	18.04 (10.30)
Stress (DASS-21)	10.42 (7.32)	17.60 (9.20)	22.49 (9.37)
Psychological resilience (RES[e])	25.58 (5.11)	22.04 (6.02)	18.82 (7.15)
Positive mental health (MHC-SF[f])	50.0 (12.05)	43.11 (14.89)	31.42 (14.28)

[a]MIRROR: Mobile Insight in Risk, Resilience and Online Referral.

[b]PTSD: posttraumatic stress disorder.

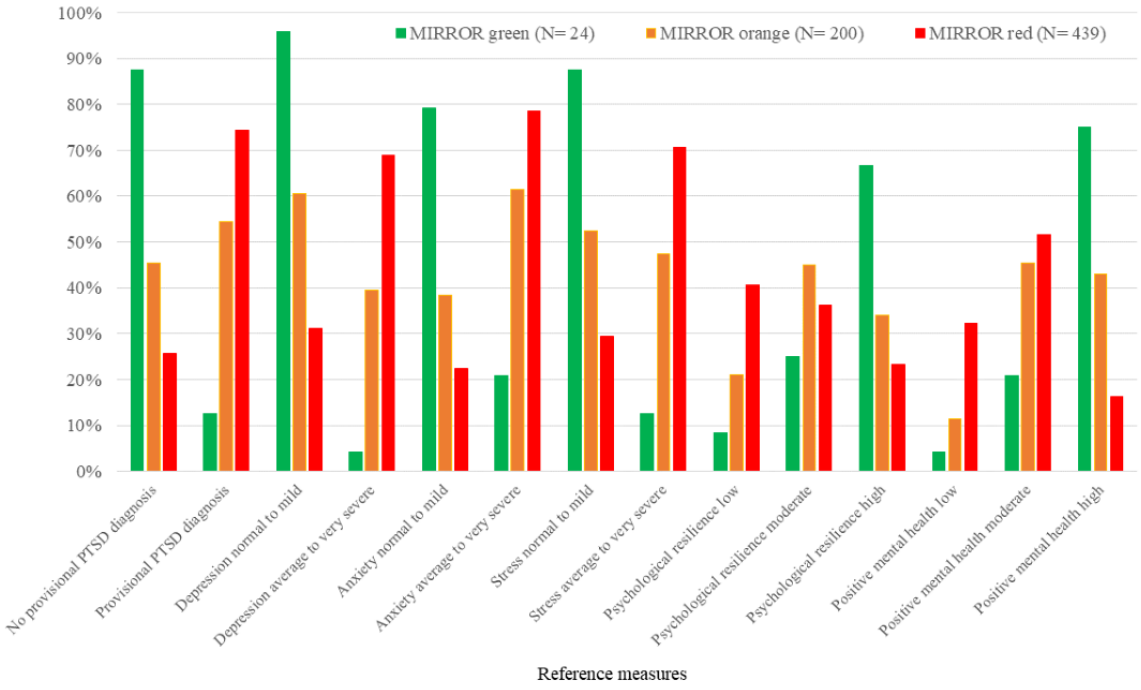
[c]PCL-5: PTSD Checklist for DSM-5.

[d]DASS-21: Depression Anxiety Stress Scale.

[e]RES: Resilience Evaluation Scale.

[f]MHC-SF: Mental Health Continuum Short Form.

Figure 1. Classification percentages on reference measures of each Mobile Insight in Risk, Resilience, and Online Referral (MIRROR) outcome category.



We conducted several 1-way between-groups ANOVAs to investigate the difference in mean scores on the reference measures between MIRROR outcome categories. As can be seen, negative outcomes were highest for the red MIRROR outcome category and positive outcomes highest for the green outcome category. The ANOVA results are shown in Table 8. We found significant differences in PTSD symptoms; depression anxiety, and stress; psychological resilience; and positive mental health between groups. Post hoc tests revealed that PTSD symptoms and depression, anxiety, and stress complaints were significantly different between all groups (P<.001). Psychological resilience was significantly higher for the green and orange MIRROR outcome category compared with the red category (P<.001). It was also significantly higher for the green category compared with the orange category (P=.01). Positive mental health was significantly higher for the green and orange category compared with the red category (P<.001). There was no significant difference between the green and orange category (P=.07).

Table 8. One-way between-groups analyses of variance with Mobile Insight in Risk, Resilience and Online Referral (MIRROR) outcome categories and reference measures.

Analysis of variance	*F*-test	Cohen *d*	df[a] between groups	df within groups	*P* value
PTSD[b] symptoms[c]	73.32	.168	2	62.90	<.001
Depression[c]	65.21	.136	2	65.81	<.001
Anxiety[c]	42.48	.072	2	67.37	<.001
Stress	34.15	.094	2	660.0	<.001
Psychological resilience[c]	30.13	.068	2	65.44	<.001
Positive mental health	57.79	.069	2	660.00	<.001

[a]df: degree of freedom.

[b]PTSD: posttraumatic stress disorder.

[c]The assumption of equal variances was violated. Therefore, the Welch *F*-test and Games-Howell post hoc test were used.

Discussion

Principal Findings and Comparison With Prior Work

The purpose of this study was to evaluate the use and psychometric and classification properties of MIRROR. MIRROR is an innovative web-based self-help test to identify individuals who develop psychological complaints after a PTE, encourage them to seek help, and support self-reliance. Our results indicated that MIRROR is a valid and reliable self-help test to identify negative outcomes (PTSD core symptoms) and positive outcomes (psychosocial functioning and resilience). MIRROR is able to correctly classify respondents according to their PTSD complaints and scores on reference measures. During the study period, 87.95% (2247/2555) of respondents who started MIRROR completed it.

We found that MIRROR's presupposed model of 3 factors (PTSD symptoms, psychosocial functioning, and resilience) did not fit our data best. Instead, a 2-factor solution showed good model fit, conceptual meaning, and maximum parsimony. This model separates MIRROR's PTSD items from the functioning and resilience items (social support, self-reliance, and problem solving). In retrospect, the grouping of the functioning and resilience items is not entirely surprising. If we assume stress to be the result of an imbalance between perceived external and internal demands and perceived personal and social resources [62], it is likely that this distinction between demands and resources is reflected in the way people cope with PTEs. We propose to call the factor psychosocial resources. In accordance with this distinction, the 2-factor model clearly separates negative (PTSD complaints) and positive (psychosocial resources) outcomes of PTEs. This is in line with the general notion that PTSD and psychosocial resources are separate constructs [63-65].

The convergent and divergent validity of MIRROR is supported by the correlations that were found between MIRROR and the reference measures. The results indicate good convergent and divergent validity for MIRROR's PTSD items. As expected, MIRROR's PTSD showed strongest correlations with PTSD (assessed with the PCL-5), followed by a lower but substantial correlation with psychological complaints (measured with the DASS-21). MIRROR's PTSD items showed low correlations with positive reference measures (assessed with the RES and MHC-SF). The results indicate adequate convergent and divergent validity for MIRROR's resilience items but less distinct than MIRROR's PTSD. MIRROR's resilience items showed strongest correlations with psychological resilience, followed by slightly lower but substantial correlations with the other reference measures. The results in this study correspond with the finding of Van der Meer et al [43] who found the RES total scale to be positively associated with established measures for resilience, self-esteem, self-efficacy, and global functioning and negatively associated with PTSD symptoms. Furthermore, the different patterns of correlations for MIRROR's PTSD and resilience scales agrees with the notion that PTSD and resilience are two separate constructs [63-65]. MIRROR's functioning item showed the strongest correlation with psychological complaints and PTSD and lower correlations with the positive reference measures. This indicates adequate convergent and divergent validity. The factor analyses revealed that functioning belongs to the resilience items of MIRROR. However, the correlation between MIRROR's functioning item and psychological complaints and PTSD is in line with studies that show that psychosocial functioning can be impaired by psychological complaints [64,66,67].

We found that both MIRROR's PTSD and resilience scales show good internal consistency. The Cronbach alpha coefficients for these scales are relatively low (.68 and .62, respectively), but this is not unusual given the (intentionally) short scales of MIRROR and given that Cronbach alpha is a function of scale length [61]. Because MIRROR contains only few items, we calculated interitem and item-total correlations. The results indicate that both scales have high internal consistency and that high scores on the items are associated with high scores on the overall scales.

MIRROR was able to correctly classify respondents into green (no further action needed), orange (encourage self-monitoring), or red (encourage seeking consultation) outcome categories and advice compared with the other measures. Results showed that respondents with a red outcome reported having more severe PTSD symptoms; more severe depression, anxiety, and stress complaints; and lower psychological resilience and positive mental health compared with respondents with a green or orange

outcome. The occurrence of PTSD and other stress-related complaints like depression following traumatic exposure is in line with former results [68]. It is important to recognize that MIRROR is specifically evaluating the risk of developing PTSD instead of other mental health outcomes of PTEs such as depression, anxiety, and substance abuse. If a respondent experiences low functioning, they will receive advice to seek consultation with their general practitioner despite the level of their PTSD complaints. This is based on the assumption that low functioning but no PTSD complaints may indicate that other problems could be at hand such as depression, anxiety, or substance abuse. Importantly, MIRROR appears to adequately identify users with more severe complaints and validly advises them to seek help. Our results seem to underline the relevance of including the factor "time since the event" in MIRROR's algorithm. According to the PCL-5, 54.5% (109/200) of the respondents with the orange outcome had a provisional PTSD diagnosis. However, their complaints could still diminish, considering the event happened only recently for these respondents and research has shown that in most individuals complaints usually diminish over time [1,2,11]. Therefore, in accordance with international guidelines [14], respondents with the orange outcome are advised to monitor how their complaints develop (by setting a reminder to use MIRROR again in 2 weeks).

The evaluation of MIRROR's use with Google Analytics showed that the number of users of MIRROR was substantial (n=2555), and the completion rate was high (2247/2555, 87.95%). These results are in line with former studies on apps assessing and monitoring mental health after PTEs indicating high use [29,30,36] and high completion rate [49]. In general, the follow-up options were visited less frequently (161 to 411 unique visits) than the outcome and advice page (2247 unique visits). A reason for this could be that receiving MIRROR's outcome and advice is sufficient initial support for people who have experienced a PTE, providing insight into how they are coping. A total of 28.7% (113/194) of respondents who were advised to complete MIRROR again in 2 weeks immediately set a reminder, suggesting MIRROR is able to support self-monitoring. Unfortunately, this study's design and considerations of ethical nature did not enable us to assess use in more depth.

Future Research and Limitations

Although guidelines on screening for PTSD complaints and postdisaster psychosocial care are widely available [7,15,69-71], the challenge remains how to reach and identify people at risk of developing psychological complaints after a PTE on a large scale. Future research could focus on investigating the implementation of MIRROR on a larger scale—for example, after terrorist attacks or natural disasters. Literature is inconclusive about the benefits versus disadvantages of formal screening of an entire population after a disaster or crisis [14,15,69,72]. Because of limited evidence of effectivity and sensitivity of screening, organizational efforts related to screening, and the often scarce resources available [25,73], it is generally not recommended to perform formal screening of complaints among all involved people following incidents. At the same time, we know that early recognition and timely referral to help are essential for preventing and treating traumatic stress symptoms. This is supported by evidence of the effectiveness of early psychological interventions for individuals prescreened with traumatic stress symptoms shortly following trauma and no benefits in those not prescreened for these symptoms [16]. Mobile apps such as MIRROR can make a contribution to solving the screening dilemma by supporting low key, accessible, and easy-to-use self-assessment and -monitoring. In this view, MIRROR could be implemented as a first step in the support for people who have experienced a PTE, before having to consult professional care [29,36]. MIRROR might lower the barrier to seek help given its open accessibility and anonymity. Future research could focus on acquiring longitudinal data of MIRROR to assess the development of complaints, functioning, and resilience over time and establish MIRROR's ability to correctly classify users accordingly. Also, qualitative research might clarify what actions users take as a result of MIRROR's personal advice.

Our study has some limitations. In our validation sample, 74.2% (492/663) of respondents were female, and 45.3% (300/663) of respondents were aged between 21 and 40 years. This could lead to selection bias and limited generalizability of the results, which is common with open internet surveys [74]. However, our sample is a specifically targeted sample because it consisted of visitors of the website of Victim Support Netherlands. Considering website visitors were automatically led to MIRROR when searching for information regarding stress reactions following a PTE, a high prevalence of psychological complaints after traumatic exposure in our sample could be expected. Moreover, research has shown that women have a higher risk of developing PTSD compared with men [75], they are more likely to seek medical or health-related information online [76], and young people use the internet as their main source of information, and this is also true for mental health concerns [77,78]. This demonstrates that the targeted sample was reached. The main strength of this study is by comparing MIRROR to more broadly used reference measures, it contributes to the highly needed evidence base of mobile apps with the potential to improve timely identification of psychological complaints [29,30,79].

Conclusions

This study shows that MIRROR is a psychometrically sound, anonymous, and easily accessible self-help test for people who have experienced a PTE. It is able to identify both negative (PTSD symptoms) and positive (psychosocial resources) outcomes of PTEs and classify respondents in accordance with reference measures. This study will hopefully contribute to enhancing adequate and timely identification of people who suffer from psychological complaints after PTEs.

Acknowledgments

We would like to thank Victim Support Netherlands (Slachtofferhulp Nederland) for their help in facilitating this study. This study was partially funded by the Interreg North-West Europe Programme which invested in the e-mental health innovation and transnational implementation platform North West Europe (eMEN) project, an EU-wide platform for e-mental health innovation and implementation formed by private and public partners in northwest Europe [80]. This study was also funded by ARQ National Psychotrauma Centre and Victim Support Netherlands.

Conflicts of Interest

This study has been conducted by the independent research center ARQ Centre of Expertise for the Impact of Disasters and Crises and ARQ Centre ’45. The funders (ARQ National Psychotrauma Centre, Interreg North-West Europe, and Victim Support Netherlands) had no influence on the outcomes of this study.

Multimedia Appendix 1
Overview of outcomes in the Mobile Insight in Risk, Resilience, and Online Referral (MIRROR) instrument.
[DOCX File , 24 KB - jmir_v22i9e19716_app1.docx]

Multimedia Appendix 2
Sample characteristics on reference measures and demography (n=663).
[DOCX File , 28 KB - jmir_v22i9e19716_app2.docx]

References

1. Bonanno GA, Westphal M, Mancini AD. Resilience to loss and potential trauma. Annu Rev Clin Psychol 2011;7:511-535. [doi: 10.1146/annurev-clinpsy-032210-104526] [Medline: 21091190]
2. de Vries G, Olff M. The lifetime prevalence of traumatic events and posttraumatic stress disorder in the Netherlands. J Trauma Stress 2009 Aug;22(4):259-267. [doi: 10.1002/jts.20429] [Medline: 19645050]
3. Benjet C, Bromet E, Karam EG, Kessler RC, McLaughlin KA, Ruscio AM, et al. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. Psychol Med 2016 Jan;46(2):327-343 [FREE Full text] [doi: 10.1017/S0033291715001981] [Medline: 26511595]
4. Kessler RC, Aguilar-Gaxiola S, Alonso J, Benjet C, Bromet EJ, Cardoso G, et al. Trauma and PTSD in the WHO World Mental Health Surveys. Eur J Psychotraumatol 2017;8(sup5):1353383 [FREE Full text] [doi: 10.1080/20008198.2017.1353383] [Medline: 29075426]
5. Knipscheer J, Sleijpen M, Frank L, de Graaf R, Kleber R, Ten Have M, et al. Prevalence of potentially traumatic events, other life events and subsequent reactions indicative for posttraumatic stress disorder in the Netherlands: a general population study based on the trauma screening questionnaire. Int J Environ Res Public Health 2020 Mar 06;17(5):1725 [FREE Full text] [doi: 10.3390/ijerph17051725] [Medline: 32155752]
6. Bonanno G. Resilience in the face of potential trauma. Curr Dir Psychol Sci 2005;14(3):135-138 [FREE Full text] [doi: 10.1111/j.0963-7214.2005.00347.x]
7. Silver RC, Holman EA, McIntosh DN, Poulin M, Gil-Rivas V. Nationwide longitudinal study of psychological responses to September 11. JAMA 2002 Sep 11;288(10):1235-1244. [doi: 10.1001/jama.288.10.1235] [Medline: 12215130]
8. Forbes D, Creamer M, Phelps A, Bryant R, McFarlane A, Devilly GJ, et al. Australian guidelines for the treatment of adults with acute stress disorder and post-traumatic stress disorder. Aust N Z J Psychiatry 2007 Aug;41(8):637-648. [doi: 10.1080/00048670701449161] [Medline: 17620160]
9. Galatzer-Levy IR, Huang SH, Bonanno GA. Trajectories of resilience and dysfunction following potential trauma: a review and statistical evaluation. Clin Psychol Rev 2018 Jul;63:41-55. [doi: 10.1016/j.cpr.2018.05.008] [Medline: 29902711]
10. Olff M, Amstadter A, Armour C, Birkeland MS, Bui E, Cloitre M, et al. A decennial review of psychotraumatology: what did we learn and where are we going? Eur J Psychotraumatol 2019;10(1):1672948 [FREE Full text] [doi: 10.1080/20008198.2019.1672948] [Medline: 31897268]
11. Hobfoll SE, Watson P, Bell CC, Bryant RA, Brymer MJ, Friedman MJ, et al. Five essential elements of immediate and mid-term mass trauma intervention: empirical evidence. Psychiatry 2007;70(4):283-315. [doi: 10.1521/psyc.2007.70.4.283] [Medline: 18181708]
12. Bryant RA, Harvey AG, Guthrie RM, Moulds ML. Acute psychophysiological arousal and posttraumatic stress disorder: a two-year prospective study. J Trauma Stress 2003 Oct;16(5):439-443. [doi: 10.1023/A:1025750209553] [Medline: 14584627]
13. Bryant RA. Acute stress reactions: can biological responses predict posttraumatic stress disorder? CNS Spectr 2003 Sep;8(9):668-674. [doi: 10.1017/s1092852900008853] [Medline: 15079140]
14. Post-traumatic stress disorder NICE guideline NG116. National Institute of Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng116/resources/posttraumatic-stress-disorder-pdf-66141601777861 [accessed 2020-09-02]

15. Bisson JI, Tavakoly B, Witteveen AB, Ajdukovic D, Jehel L, Johansen VJ, et al. TENTS guidelines: development of post-disaster psychosocial care guidelines through a Delphi process. Br J Psychiatry 2010 Jan;196(1):69-74. [doi: 10.1192/bjp.bp.109.066266] [Medline: 20044665]
16. Roberts NP, Kitchiner NJ, Kenardy J, Lewis CE, Bisson JI. Early psychological intervention following recent trauma: a systematic review and meta-analysis. Eur J Psychotraumatol 2019;10(1):1695486 [FREE Full text] [doi: 10.1080/20008198.2019.1695486] [Medline: 31853332]
17. Boelen PA, Olff M, Smid GE. Traumatic loss: mental health consequences and implications for treatment and prevention. Eur J Psychotraumatol 2019 Apr 15;10(1):1591331. [doi: 10.1080/20008198.2019.1591331]
18. Oosterbaan V, Covers MLV, Bicanic IAE, Huntjens RJC, de Jongh A. Do early interventions prevent PTSD? A systematic review and meta-analysis of the safety and efficacy of early interventions after sexual assault. Eur J Psychotraumatol 2019;10(1):1682932 [FREE Full text] [doi: 10.1080/20008198.2019.1682932] [Medline: 31762949]
19. Stene LE, Wentzel-Larsen T, Dyb G. Healthcare needs, experiences and satisfaction after terrorism: a longitudinal study of survivors from the Utøya attack. Front Psychol 2016;7:1809 [FREE Full text] [doi: 10.3389/fpsyg.2016.01809] [Medline: 27933008]
20. Jeavons S. Long-term needs of motor vehicle accident victims: are they being met? Aust Health Rev 2001;24(1):128-135. [doi: 10.1071/ah010128] [Medline: 11357727]
21. Cheng H, Wang C, McDermott RC, Kridel M, Rislin JL. Self-stigma, mental health literacy, and attitudes toward seeking psychological help. J Couns Devel 2018 Jan 04;96(1):64-74. [doi: 10.1002/jcad.12178]
22. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
23. Shalev AY, Ankri YLE, Peleg T, Israeli-Shalev Y, Freedman S. Barriers to receiving early care for PTSD: results from the Jerusalem trauma outreach and prevention study. Psychiatr Serv 2011 Jul;62(7):765-773. [doi: 10.1176/ps.62.7.pss6207_0765] [Medline: 21724790]
24. Stuber J, Galea S, Boscarino JA, Schlesinger M. Was there unmet mental health need after the September 11, 2001 terrorist attacks? Soc Psychiatry Psychiatr Epidemiol 2006 Mar;41(3):230-240. [doi: 10.1007/s00127-005-0022-2] [Medline: 16424968]
25. Brewin CR, Fuchkan N, Huntley Z, Robertson M, Thompson M, Scragg P, et al. Outreach and screening following the 2005 London bombings: usage and outcomes. Psychol Med 2010 Dec;40(12):2049-2057 [FREE Full text] [doi: 10.1017/S0033291710000206] [Medline: 20178677]
26. Rosenbaum L. Post-traumatic stress disorder: a challenge for primary care—misunderstood and incognito. Brit J Gen Pract 2004;54(499):83-85. [doi: 10.1093/med/9780199383863.003.0011] [Medline: 14965384]
27. Verhaak PFM, Schellevis FG, Nuijen J, Volkers AC. Patients with a psychiatric disorder in general practice: determinants of general practitioners' psychological diagnosis. Gen Hosp Psychiatry 2006 Apr;28(2):125-132. [doi: 10.1016/j.genhosppsych.2005.11.001] [Medline: 16516062]
28. Dekkers AMM, Olff M, Näring GWB. Identifying persons at risk for PTSD after trauma with TSQ in the Netherlands. Community Ment Health J 2010 Feb;46(1):20-25 [FREE Full text] [doi: 10.1007/s10597-009-9195-6] [Medline: 19396545]
29. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
30. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
31. Connor KM. Assessment of resilience in the aftermath of trauma. J Clin Psychiatry 2006;67 Suppl 2:46-49 [FREE Full text] [Medline: 16602815]
32. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. Health Qual Life Outcomes 2011 Feb 04;9(8):1-18 [FREE Full text] [doi: 10.1186/1477-7525-9-8] [Medline: 21294858]
33. Mouthaan J, Sijbrandij M, Reitsma JB, Gersons BPR, Olff M. Comparing screening instruments to predict posttraumatic stress disorder. PLoS One 2014;9(5):e97183 [FREE Full text] [doi: 10.1371/journal.pone.0097183] [Medline: 24816642]
34. Brewin CR. Systematic review of screening instruments for adults at risk of PTSD. J Trauma Stress 2005 Mar;18(1):53-62. [doi: 10.1002/jts.20007] [Medline: 16281196]
35. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Arlington: American Psychiatric Publishing; 2013.
36. Price M, Sawyer T, Harris M, Skalka C. Usability evaluation of a mobile monitoring system to assess symptoms after a traumatic injury: a mixed-methods study. JMIR Ment Health 2016 Jan;3(1):e3 [FREE Full text] [doi: 10.2196/mental.5023] [Medline: 26753673]
37. Price M, Kuhn E, Hoffman JE, Ruzek J, Acierno R. Comparison of the PTSD checklist (PCL) administered via a mobile device relative to a paper form. J Trauma Stress 2015 Oct;28(5):480-483. [doi: 10.1002/jts.22037] [Medline: 26375277]
38. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

39. Rodriguez-Paras C, Tippey K, Brown E, Sasangohar F, Creech S, Kum H. Posttraumatic stress disorder and mobile health: app investigation and scoping literature review. JMIR Mhealth Uhealth Oct 26 2017;5(10):e156. [doi: 10.2196/mhealth.7318] [Medline: 29074470]
40. International Statistical Classification of Diseases and Related Health Problems: Tenth Revision, 2nd Edition. World Health Organization. 2004. URL: https://apps.who.int/iris/handle/10665/42980 [accessed 2020-09-02]
41. ARQ Centre of Expertise for the Impact of Disasters and Crises. Dutch Multidisciplinary Guideline for Post-Disaster Psychosocial Support. Diemen: ARQ Centre of Expertise for the Impact of Disasters and Crises; 2014.
42. Boeschoten MA, Bakker A, Jongedijk RA, Olff M. Life Events Checklist for DSM-5: Nederlandstalige versie [Dutch version]. ARQ Centre '45. 2014. URL: https://www.psychotraumadiagnostics.centrum45.nl/en [accessed 2020-09-02]
43. van der Meer CAI, Te Brake H, van der Aa N, Dashtgard P, Bakker A, Olff M. Assessing psychological resilience: development and psychometric properties of the English and Dutch version of the Resilience Evaluation Scale (RES). Front Psychiatry 2018;9:169 [FREE Full text] [doi: 10.3389/fpsyt.2018.00169] [Medline: 29867601]
44. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol 2004;55:803-832. [doi: 10.1146/annurev.psych.55.090902.141601] [Medline: 14744235]
45. Wood RTA, Griffiths MD, Eatough V. Online data collection from video game players: methodological issues. Cyberpsychol Behav 2004 Oct;7(5):511-518. [doi: 10.1089/cpb.2004.7.511] [Medline: 15667045]
46. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5). URL: https://www.ptsd.va.gov/professional/assessment/documents/PCL5_Standard_form.PDF [accessed 2020-09-02]
47. Eidhof MB, Ter Heide FJJ, van Der Aa N, Schreckenbach M, Schmidt U, Brand BL, et al. The dissociative subtype of PTSD interview (dsp-i): development and psychometric properties. J Trauma Dissociation 2019;20(5):564-581. [doi: 10.1080/15299732.2019.1597806] [Medline: 31132959]
48. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
49. van der Meer CAI, Bakker A, Schrieken BAL, Hoofwijk MC, Olff M. Screening for trauma-related symptoms via a smartphone app: the validity of Smart Assessment on your Mobile in referred police officers. Int J Methods Psychiatr Res 2017 Sep;26(3):1 [FREE Full text] [doi: 10.1002/mpr.1579] [Medline: 28948699]
50. Krüger-Gottschalk A, Knaevelsrud C, Rau H, Dyer A, Schäfer I, Schellong J, et al. The German version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): psychometric properties and diagnostic utility. BMC Psychiatry 2017 Nov 28;17(1):379 [FREE Full text] [doi: 10.1186/s12888-017-1541-6] [Medline: 29183285]
51. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
52. de Beurs E, van Dyck R, Marquenie L, Lange A, Blonk R. De DASS: een vragenlijst voor het meten van depressie, angst en stress. Gedragstherapie 2001;34(1):35-54. [doi: 10.1037/t69646-000]
53. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
54. Lee D. The convergent, discriminant, and nomological validity of the Depression Anxiety Stress Scales-21 (DASS-21). J Affect Disord 2019 Dec 01;259:136-142. [doi: 10.1016/j.jad.2019.06.036] [Medline: 31445339]
55. Keyes CLM. Mental health in adolescence: is America's youth flourishing? Am J Orthopsychiatry 2006 Jul;76(3):395-402. [doi: 10.1037/0002-9432.76.3.395] [Medline: 16981819]
56. Lamers SMA, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CLM. Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF). J Clin Psychol 2011 Jan;67(1):99-110. [doi: 10.1002/jclp.20741] [Medline: 20973032]
57. Keyes CLM, Wissing M, Potgieter JP, Temane M, Kruger A, van Rooy S. Evaluation of the mental health continuum-short form (MHC-SF) in Setswana-speaking South Africans. Clin Psychol Psychother 2008;15(3):181-192. [doi: 10.1002/cpp.572] [Medline: 19115439]
58. Muthén L, Muthén B. Mplus User's Guide. Eighth Edition. Los Angeles: Muthén & Muthén; 2017.
59. Cheung G, Rensvold R. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Eq Mod 2002;9(2):233-255 [FREE Full text] [doi: 10.1207/s15328007sem0902_5]
60. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Eq Mod 1991;6:1-55 [FREE Full text] [doi: 10.1080/10705519909540118]
61. Clark L, Watson D. Constructing validity: basic issues in objective scale development. Psychol Assess 1995;7(3):309 [FREE Full text] [doi: 10.1037/1040-3590.7.3.309]
62. Folkman S, Lazarus RS. Stress, Appraisal, and Coping. New York: Springer Publishing Company; 1984.
63. Olff M. Bonding after trauma: on the role of social support and the oxytocin system in traumatic stress. Eur J Psychotraumatol 2012;3:1 [FREE Full text] [doi: 10.3402/ejpt.v3i0.18597] [Medline: 22893838]

64. Kuhn E, Blanchard EB, Hickling EJ. Posttraumatic stress disorder and psychosocial functioning within two samples of MVA survivors. Behav Res Ther 2003 Sep;41(9):1105-1112 [FREE Full text] [doi: 10.1016/s0005-7967(03)00071-8] [Medline: 12914811]
65. Ozer EJ, Best SR, Lipsey TL, Weiss DS. Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. Psychol Bull 2003 Jan;129(1):52-73. [doi: 10.1037/0033-2909.129.1.52] [Medline: 12555794]
66. Palyo SA, Beck JG. Post-traumatic stress disorder symptoms, pain, and perceived life control: associations with psychosocial and physical functioning. Pain 2005 Sep;117(1-2):121-127 [FREE Full text] [doi: 10.1016/j.pain.2005.05.028] [Medline: 16099099]
67. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One 2014;9(2):e90311 [FREE Full text] [doi: 10.1371/journal.pone.0090311] [Medline: 24587318]
68. O'Donnell ML, Creamer M, Pattison P. Posttraumatic stress disorder and depression following trauma: understanding comorbidity. Am J Psychiatry 2004 Aug;161(8):1390-1396. [doi: 10.1176/appi.ajp.161.8.1390] [Medline: 15285964]
69. Te Brake H, Dückers M. Early psychosocial interventions after disasters, terrorism and other shocking events: is there a gap between norms and practice in Europe? Eur J Psychotraumatol 2013;4(1):19093 [FREE Full text] [doi: 10.3402/ejpt.v4i0.19093] [Medline: 23393613]
70. Te Brake H, Dückers M, De Vries M, Van Duin D, Rooze M, Spreeuwenberg C. Early psychosocial interventions after disasters, terrorism, and other shocking events: guideline development. Nurs Health Sci 2009 Dec;11(4):336-343. [doi: 10.1111/j.1442-2018.2009.00491.x] [Medline: 19909437]
71. Jacobs J, Oosterbeek M, Tummers LG, Noordegraaf M, Yzermans CJ, Dückers MLA. The organization of post-disaster psychosocial support in the Netherlands: a meta-synthesis. Eur J Psychotraumatol 2019;10(1):1544024 [FREE Full text] [doi: 10.1080/20008198.2018.1544024] [Medline: 30815232]
72. Australian Centre for Posttraumatic Mental Health. Australian Guidelines for the Treatment of Acute Stress Disorder and Posttraumatic Stress Disorder. Melbourne, Victoria: ACPMH; 2013.
73. Bisson JI, Weltch R, Maddern S, Shepherd JP. Implementing a screening programme for post-traumatic stress disorder following violent crime. Eur J Psychotraumatol 2010;1:1 [FREE Full text] [doi: 10.3402/ejpt.v1i0.5541] [Medline: 22893799]
74. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
75. Olff M. Sex and gender differences in post-traumatic stress disorder: an update. Eur J Psychotraumatol 2017 Jul 27;8(sup4):1351204. [doi: 10.1080/20008198.2017.1351204]
76. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
77. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
78. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]
79. Ennis N, Sijercic I, Monson CM. Internet-delivered early interventions for individuals exposed to traumatic events: systematic review. J Med Internet Res 2018 Nov 14;20(11):e280 [FREE Full text] [doi: 10.2196/jmir.9795] [Medline: 30429113]
80. e-mental health innovation and transnational implementation platform North West Europe (eMEN). Interreg NWE. URL: https://www.nweurope.eu/projects/project-search/e-mental-health-innovation-and-transnational-implementation-platform-north-west-europe-emen/ [accessed 2020-09-02]

Abbreviations

ANOVA: analysis of variance
CFA: confirmatory analysis
CFI: comparative fit index
DASS-21: Depression Anxiety Stress scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EFA: exploratory factor analysis
eMEN: e-Mental health innovation and transnational implementation platform North West Europe
GAF: Global Assessment of Functioning
ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision
LEC-5: Life Events Checklist for DSM-5
MHC-SF: Mental Health Continuum Short Form
MIRROR: Mobile Insight in Risk, Resilience and Online Referral
NICE: National Institute for Health and Care Excellence
PCL-5: PTSD Checklist for DSM-5
PTE: potentially traumatic event

PTSD: posttraumatic stress disorder
RES: Resilience Evaluation Scale
RMSEA: root mean square error of approximation
TENTS: The European Network for Traumatic Stress
TLI: Tucker-Lewis index
WLSMV: means and variance adjusted weighted least square

Edited by T Rashid Soron; submitted 30.04.20; peer-reviewed by A Teles, A Mavragani, C Benjet; comments to author 25.05.20; revised version received 08.07.20; accepted 17.07.20; published 25.09.20.

Please cite as:
van Herpen MM, Boeschoten MA, te Brake H, van der Aa N, Olff M
Mobile Insight in Risk, Resilience, and Online Referral (MIRROR): Psychometric Evaluation of an Online Self-Help Test
J Med Internet Res 2020;22(9):e19716
URL: http://www.jmir.org/2020/9/e19716/
doi:10.2196/19716
PMID:32975521

Original Paper

Nonprofessional Peer Support to Improve Mental Health: Randomized Trial of a Scalable Web-Based Peer Counseling Course

Samantha L Bernecker[1], PhD; Joseph Jay Williams[2], PhD; Norian A Caporale-Berkowitz[3], MA; Akash R Wasil[4], BA; Michael J Constantino[5], PhD

[1]Department of Health Care Policy, Harvard Medical School, Boston, MA, United States

[2]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[3]Department of Educational Psychology, University of Texas at Austin, Austin, TX, United States

[4]Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States

[5]Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA, United States

Corresponding Author:
Samantha L Bernecker, PhD
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA, 02115-5899
United States
Phone: 1 8145749625
Email: samantha.bernecker@gmail.com

Abstract

Background: Millions of people worldwide are underserved by the mental health care system. Indeed, most mental health problems go untreated, often because of resource constraints (eg, limited provider availability and cost) or lack of interest or faith in professional help. Furthermore, subclinical symptoms and chronic stress in the absence of a mental illness diagnosis often go unaddressed, despite their substantial health impact. Innovative and scalable treatment delivery methods are needed to supplement traditional therapies to fill these gaps in the mental health care system.

Objective: This study aims to investigate whether a self-guided web-based course can teach pairs of nonprofessional peers to deliver psychological support to each other.

Methods: In this experimental study, a community sample of 30 dyads (60 participants, mostly friends), many of whom presented with mild to moderate psychological distress, were recruited to complete a web-based counseling skills course. Dyads were randomized to either immediate or delayed access to training. Before and after training, dyads were recorded taking turns discussing stressors. Participants' skills in the *helper* role were assessed before and after taking the course: the first author and a team of trained research assistants coded recordings for the presence of specific counseling behaviors. When in the *client* role, participants rated the session on helpfulness in resolving their stressors and supportiveness of their peers. We hypothesized that participants would increase the use of skills taught by the course and decrease the use of skills discouraged by the course, would increase their overall adherence to the guidelines taught in the course, and would perceive posttraining counseling sessions as more helpful and their peers as more supportive.

Results: The course had large effects on most helper-role speech behaviors: helpers decreased total speaking time, used more restatements, made fewer efforts to influence the speaker, and decreased self-focused and off-topic utterances (*d*s=0.8-1.6). When rating the portion of the session in which they served as clients, participants indicated that they made more progress in addressing their stressors during posttraining counseling sessions compared with pretraining sessions (*d*=1.1), but they did not report substantive changes in feelings of closeness and supportiveness of their peers (*d*=0.3).

Conclusions: The results provide proof of concept that nonprofessionals can learn basic counseling skills from a scalable web-based course. The course serves as a promising model for the development of web-based counseling skills training, which could provide accessible mental health support to some of those underserved by traditional psychotherapy.

(J Med Internet Res 2020;22(9):e17164) doi:10.2196/17164

KEYWORDS

online learning; nonprofessional education; educational technology; computer-assisted instruction; social support; mental health; psychological stress; eHealth; internet

Introduction

Background

The mental health care system in the United States fails to meet the needs of millions of people, prompting numerous calls for *disruptive innovations* in mental health care delivery [1,2]. Several gaps in the current system point to the need for such innovations. First, many people with mental illness are unable to access treatment; the number of people with mental illness far outstrips available resources, and cost and other structural barriers are pervasive [3,4]. Second, others choose not to seek help because of negative beliefs about treatment [5,6]. Finally, the mental health system is not designed to address the adverse effects of subclinical symptoms and chronic stress that affect even those without diagnosable mental illnesses [7], and which increases the risk of future psychological and physical decline [8].

Self-guided digital technologies, including self-help apps and chatbots, have been proposed as solutions because of the advantages they provide in access and cost, but they are not a panacea, displaying several limitations [9,10]. Their reach is limited because people seeking mental health support typically prefer face-to-face over computerized therapy [11,12]. Their efficacy is limited because digital tools often fail to motivate and engage users [13,14]. They also currently lack the human-level intelligence required to address nuanced problems [15,16]. It appears that until realistic artificial intelligence is available, many people require human-delivered interventions to meet their preferences, engage them, and respond to their unique concerns. However, this raises the question of how human-delivered interventions could solve the problems with traditional treatments that digital interventions have been created to address—how can human-delivered interventions scale to reach an enormous number of people with mental illnesses, appeal to those who are not interested in professional care, and reduce the burden of subclinical symptoms and stress?

We propose one possibility for a human-delivered solution to address these needs: a *Crowdsourcing Mental Health* (CMH) model that leverages the benefits of technology to overcome treatment barriers while addressing limitations of technology by incorporating the important human element. In the proposed model, digital tools could be used to train nonprofessionals, who would then counsel their peers face-to-face. Even if it is less potent than traditional psychotherapy, such a scalable intervention could have considerable public health impact because of its greater reach [17], a possibility corroborated by survey research. In a survey of more than 500 internet users, 64% of respondents indicated that they would participate in reciprocal peer counseling using skills that they and a peer learned via a web-based course [18]. More than 50% of the respondents who stated that they would never seek psychotherapy or medication expressed willingness to try this model—an important indicator that some of those underserved by traditional treatments could benefit from reciprocal peer counseling.

Design Considerations for a Peer Counseling Program

We propose 3 features to include in the design of a nonprofessional peer counseling program if it is to meet the aforementioned gaps in traditional mental health care by scaling to meet demand, appealing to those who do not want to seek professional care, and treating subclinical symptoms and stress, all while incorporating human interaction. These features are (1) transdiagnostic applicability (ie, applicability regardless of diagnosis), (2) reciprocity between peers, and (3) scalability of training. In this section, we highlight the relevant literature from which these design considerations were derived. We discuss how these features can address the above gaps and provide additional benefits, and we describe how these features might be implemented.

Transdiagnostic Applicability

Applicability to a wide range of problems provides several advantages for nonprofessional peer counseling interventions. This could increase the appeal of the intervention to those who are reluctant to see a professional: if the intervention is appropriate regardless of whether one has received a diagnosis, participants would not need to identify themselves as having a mental illness or to see their symptoms as “severe enough” to merit professional treatment, which are among the most common reasons individuals choose not to seek care [5,6]. In addition, a broad intervention could address the growing number of individuals with impairing subclinical symptoms or chronic stress [7], which increases the risk for mental and physical health problems [8] in addition to the direct distress they cause. Finally, a domain-general intervention can be useful if individuals cannot receive accurate diagnosis (which is challenging in the absence of a professional) [19]; a simpler screening for the level of severity or appropriateness of peer counseling may be viable.

How might this transdiagnostic applicability be achieved? Among extant transdiagnostic treatments, supportive psychotherapy may be especially well-suited for peers with limited mental health training. In supportive psychotherapy, the therapist does not target a specific symptom but rather follows the support seeker’s lead while providing “reflection, empathic listening, encouragement, and [an opportunity] to explore and express ... experiences and emotions” [20]. Thus, support seekers can address whatever problems may arise, including psychological symptoms and “normal” stressors [21]. These techniques align well with what support seekers desire from nonprofessional social support, making supportive psychotherapy especially appropriate for peer delivery [22,23]. Furthermore, in contrast with many transdiagnostic treatments that require extensive training and supervision to implement with fidelity [24,25], supportive psychotherapy’s abbreviated

list of specific techniques may render it easier to learn, although this is ultimately an empirical question.

There is consistent evidence that supportive psychotherapy improves psychological symptoms. For example, in randomized controlled trials for depression, it has medium effects versus wait-list or no treatment (approximate *d*=0.6) [20,26]. Compared with treatments that directly target the symptoms or theorized root causes of a particular disorder, supportive psychotherapy does appear at a small disadvantage, with relative effects around *d*=−0.2, but there is some indication that this difference could be driven partly by publication bias or unequal dosages [20,27,28]. Indeed, several meta-analyses have failed to find differences between supportive psychotherapy and gold-standard cognitive behavioral therapy for generalized anxiety disorder [29,30]. In addition, several randomized trials have found no or minimal differences between supportive psychotherapy and directive or expressive treatments for a variety of other conditions, including borderline personality disorder [31], posttraumatic stress disorder [32], social anxiety disorder [33], generalized anxiety disorder [34], anorexia nervosa [35], personality disorders characterized by fearful behaviors [36], and comorbid chronic depression with alcohol dependence [37]. Consequently, some have argued that supportive psychotherapy should be regarded as a "therapy of choice" rather than a control condition [36,38].

This is not to deny that, in many cases, specific techniques that target symptoms or causes may increase the potency of treatment or may be necessary to achieve remission; for example, there is an increasing consensus that treatments that incorporate exposure are superior for anxiety disorders [39-41]. However, even if less powerful than such disorder-specific treatments, delivery of supportive psychotherapy skills on a large scale by laypeople could have a substantial public health impact, especially in cases where the alternative to supportive peer counseling is no treatment at all [17]. Determining the appropriate population for peer-delivered supportive psychotherapy techniques should be guided by future clinical trials, but we propose that this intervention may be a strong fit for any individual with subclinical distress and prodromal symptoms as well as for individuals with mild to moderate mental illness across a spectrum of disorders (eg, anxiety and related disorders, mood disorders, substance use disorders, and eating disorders) who would not otherwise seek treatment.

Reciprocity Between Peers

There are many advantages to making supportive psychotherapy delivered by nonprofessionals reciprocal, such that 2 members of a dyad both give and receive support, as opposed to unidirectional, such that one member takes on a patient role and the other a counselor role. A reciprocal model has a major advantage in scaling to meet demand. Unidirectional solutions such as task shifting to trained nonprofessionals would require a multiple-thousands-fold increase in employees delivering therapeutic services full time to treat all individuals with mental health difficulties [42,43]; in contrast, a reciprocal model does not demand a large change in the workforce. Instead, reciprocal peer counseling requires only a few hours of each person's leisure time, and each person is compensated via (1) receiving support in return and (2) the benefits of providing support to others. In a sense, this model crowdsources mental health care by dividing the enormous undertaking of treating mental illness into manageable tasks carried out by laypeople.

Reciprocal peer counseling may also appeal to those who would not seek professional assistance (or, indeed, request it from their friends) because of the threat to self-esteem associated with being a person who needs help or because of concerns about burdening others. Indeed, unidirectional support receipt is sometimes associated with negative mood, potentially because of these features [44]. In contrast, reciprocity of support maintains an egalitarian relationship, and the opportunity to act as a support provider can protect health and improve mood [45,46], in some cases even more than receiving support [47].

The involvement of a peer can also remedy a limitation of most web-based self-help programs, that is, nonadherence or withdrawal from the program. A recent meta-analysis of clinical trials of smartphone apps for treating anxiety and depression found that 26% of participants withdrew (closer to half when adjusting for publication bias); however, the inclusion of human interaction reduced dropout to close to 12% [48]. Interaction with another person can provide a sense of accountability [49,50]; indeed, in a reciprocal program, participants might be especially motivated to persist because in addition to promoting their own well-being, they know another person is benefitting from their involvement.

Finally, a reciprocal peer counseling model may attack a driver of psychological ill health at its root. The detrimental health and mortality effects of social isolation and loneliness are well established [51,52], and perceived social support protects against mental illnesses [53-55]. Reciprocal self-disclosure generates intimacy [56], so taking turns as helper and client could increase perceived social support in a manner that is not present in traditional psychotherapy. Thus, a peer counseling program using this format could improve psychological well-being through 2 classes of mechanisms: it could not only give participants an opportunity to address the sources of their distress but could also generate feelings of closeness and support.

Scalability of Training

To meaningfully address mental health care shortages, training for peer counselors must be widely accessible at scale. To achieve the required reach, training should have little or no monetary cost, should be available regardless of geographic location or population density, and should effectively train people with varying backgrounds and abilities.

Therefore, we suggest that the training should be available on the web in a self-directed format (although this does limit its use to individuals who have access to an internet-connected device; additional solutions are needed for those who lack such access). Crucially, to reach the required scale, web-based training should be primarily self-guided, rather than requiring a live instructor [57]. Otherwise, the number of human trainers available would act as the limiting factor in the number of people who could be served, and human involvement would drive costs. This approach is consistent with the evolution of massive open

online courses, which are increasingly taught in a self-paced format. However, it is far from guaranteed that a self-guided web-based course could effectively teach interpersonal skills, especially to nonprofessionals who may be experiencing psychological symptoms; as discussed below, the literature on web-based interpersonal skills training is limited. Consequently, such a course must be carefully designed, drawing on the science of learning and research on online pedagogy.

Research on Extant Web-Based Therapeutic Skills Training Programs

In this section, we briefly review the supporting evidence for web-based programs that have been created to teach related skills, and we explain how the proposed intervention differs from that work.

One group of existing web-based training programs includes courses for professionals in evidence-based psychotherapies [58] and distance education programs for graduate-level counselors [59]. These programs have succeeded in increasing knowledge, self-reported skill, and, more rarely, observed skill. However, these differ from the proposed peer counseling course in 2 consequential ways: (1) they often involve considerable instructor and student interaction through telecommunication, making them difficult to expand, and (2) they teach nondistressed groups that have self-selected for aptitude and interest in mental health care delivery [58].

Another handful of web-based peer support platforms train nondistressed volunteer listeners, but unlike our program, these platforms generally do not use evidence-based behavioral teaching techniques, and it is unknown if these trainings improve listeners' behavioral adherence to guidelines [60-62].

Finally, several web-based romantic relationship enhancement programs exist, some of which teach communication skills, and these often do target couples in distress. However, studies of these programs have only shown benefits for relationship outcomes and have rarely measured changes in interpersonal behaviors [63,64] (refer to the study by Braithwaite and Fincham [65] for an exception).

Across all these types of programs, rather than rigorously evaluating observed behavior, tests of teaching efficacy tend to rely on assessments of learners' *self-reported* perceived skills and book knowledge, which may be only weakly correlated with a learner's ability to implement skills in a real interpersonal interaction [57]. We address this limitation in this study.

This Study

Owing to these gaps in the literature, it remains unclear whether a training program fitting our design desiderata would be effective—in other words, it is unknown whether nonprofessionals, including people reporting moderate psychological distress, could learn peer counseling skills via a self-guided web-based course. To address this question, we developed a peer counseling program called Crowdsourcing Mental Health (CMH). CMH fulfills our design criteria: it is a reciprocal program that begins by teaching pairs of peers supportive psychotherapy skills via a self-guided, web-based course. Thus, it may have the potential to address the current limitations of mental health care systems around access, appeal, and treatment of subclinical symptoms. Of course, CMH and other similar peer counseling programs are far from the sole solution; they are unlikely to be appropriate for some of the most vulnerable or most ill or those who have specific limitations around technology use or peer interactions. However, reciprocal peer support programs can add strong value to a portfolio of novel mental health interventions to fill gaps in the current health care system.

In this study, we describe CMH's development and report on a randomized trial designed to test its efficacy in improving skill use, adherence to guidelines, and perceived helpfulness by evaluating users' performance in recorded CMH sessions and their postsession reactions. The primary research questions (RQs) were as follows:

- RQ 1: How much does the course change the use of specific helping skills? By estimating changes in the use of individual skills, we can differentiate skills that were effectively taught from those ineffectively taught, informing revision of specific course sections. We hypothesized that helpers would increase the use of 2 behaviors prescribed by the course, would decrease the use of 2 behaviors proscribed by the course, and would decrease in-session speaking time. (We also measured some common behaviors that were neither prescribed nor proscribed and had no strong hypotheses about changes in those.)
- RQ 2: Does the course improve helpers' overall adherence in delivering helping skills? We predicted an increase in adherence from pre- to posttraining.

As the primary goal of this study was to investigate the teaching effectiveness of the course and not its impact on mental health, participants were not required to meet with their peers after completing this study; consequently, the mental health effects of repeated peer interactions could not be determined. However, as a proxy measure of whether a reciprocal peer counseling intervention of this kind could produce mental health benefits, we assessed participants' perceptions of the short-term impact of using the skills during the in-laboratory counseling sessions, enabling us to address the following RQ:

- RQ 3: Does the talker's perception of session helpfulness increase after taking the course? We hypothesized that talkers would perceive the sessions after training as more productive and that they would feel closer to their peers—in other words, that changes would take place in the 2 proposed mechanisms of reciprocal peer counseling.

Methods

Course Design

In CMH, pairs of acquaintances take a web-based course that teaches helping skills, which are the focus of this investigation, as well as talking skills, which consist of guidelines drawn from the literature on coping and emotion regulation. Both participants learn both roles. Once each person has completed the web-based course on his or her own, the peers can then meet for mutual support sessions, taking turns in the helper and talker roles. To address the design consideration of transdiagnostic

applicability, CMH's helping skills parallel the skills of supportive psychotherapy [38], which are also the core skills taught in the dominant counselor training models [66,67]. These skills include taking a warm and nonjudgmental attitude, listening attentively without attempting to influence the speaker, and using techniques to elicit reflection and elaboration (eg, paraphrasing and asking open-ended questions).

The CMH course consisted of 10 lessons, 5 on taking the helper role and 5 on taking the talker role. (The talking skills lessons were included because CMH users may be therapy-naïve and have difficulty directing their own sessions. These lessons gave instructions on how to explore a stressor, describe emotions, and develop a coping plan. As talking performance was not the focus of this study, we do not discuss these lessons further.) Each of the helper lessons addressed 1 of the following 5 topics: focusing one's attention on the talker, taking an accepting and caring attitude, avoiding unhelpful attempts to influence the talker, restating (paraphrase and summary), and asking open-ended questions.

The success of skills training is dependent on the pedagogical methods used [68]. In some previous studies of web-based psychotherapy skills training, self-guided instruction (which scales more easily) has been found to be inferior to self-guidance plus videoconference role-play with an instructor [69-71]. Owing to the need to address the design consideration of the scalability of training, we considered such use of videoconference to be infeasible for a widely disseminated and low-cost course. Therefore, we carefully developed alternative training strategies, relying on extensive review of basic and applied research on learning and online education to identify ways we could maximize efficacy while minimizing human-delivered instruction.

Implementation of Behavior Modeling Training as a Teaching Method

Behavior modeling training (BMT) is the best-supported set of techniques for increasing the performance of interpersonal and other behavioral skills [72], and it has been used effectively to teach nonprofessionals basic counseling and active listening skills in face-to-face settings [73,74]. Consequently, BMT served as the pedagogical foundation for the course. BMT includes 4 components: learners receive a description of each skill (*instruction*); view other people performing skills (*modeling*); practice skills, often through role-play (*practice*); and receive performance feedback (*feedback*).

To make the course scalable, these 4 components needed to be translated into a primarily self-guided, web-based format. Instruction and modeling were relatively simple to implement and took the form of videos: audio instruction was accompanied by text and images, and diverse volunteer actors modeled the skills. In creating these portions of the course, we also drew on training techniques identified through basic and applied research from areas as diverse as knowledge acquisition [75], motor learning [76], and computer-assisted instruction [77].

As noted earlier, practice and feedback are more challenging to translate into a primarily self-guided format. To implement practice, the course simulated interpersonal interactions with increasing degrees of complexity and realism, beginning with lower-fidelity, simpler automated exercises, and progressing toward live interactions. This approach has dual benefits: first, it can scaffold learning rather than immediately forcing learners to juggle the stimuli and challenges of a face-to-face conversation [78], and second, minimizing human involvement improves convenience and scalability. In the CMH course, learners began by typing responses to video-recorded actors, then progressed to practicing 3 times over the phone with a minimally trained mentor, and finally held 3 in-person practice sessions with the peer whom they had selected as their partner in the intervention. The demands of the mentor role were designed to be extremely minimal (eg, reading from a script) so that when CMH is publicly launched, any individual who uses CMH could volunteer to mentor new learners, eliminating the resource limitations associated with requiring trained instructors. For this study, undergraduate research assistants served as the telephone mentors.

Feedback took the form of self-assessments because of the challenges of providing nuanced human-delivered or machine-coded feedback at scale [79,80] and the risk associated with an untrained peer providing inaccurate or anxiety-provoking feedback [81]. After each exercise, the learners answered a series of questions about whether they followed each instruction. By assessing granular behaviors, learners can identify behaviors to change in the future while minimizing the threat to self-esteem and ensuing negative affect that could impede learning [82]. They were not asked to give themselves a global evaluation because self-evaluations are more accurate when specific and objective tasks are assessed [83].

Course Development Process

The course content was written by the first author, using BMT as an organizing framework for teaching the set of behavioral skills from supportive psychotherapy. Manuals on teaching counseling skills and motivational interviewing were consulted [66,67,84], both to ensure that no relevant skills were missed and to inform the design of practice exercises. When generating examples for the modeling portion of the course, we attempted to represent individuals with diverse life experiences and demographic characteristics (eg, socioeconomic status, race, and age). The written course content was then reviewed by another clinical psychologist and an online education researcher and was read and pilot tested by 2 research assistants.

After the first round of revisions to the written materials, a digital prototype of the course was created, including creating instructional and modeling videos and interactive exercises. These materials were designed in keeping with research on e-learning [77] to optimize visuals, narration style, and other elements for educational efficacy. The videos were edited using Camtasia software (TechSmith) and hosted on the TechSmith website, which allows for embedding quiz questions within videos. All course materials were hosted on the web using Qualtrics Research Suite survey software, which enabled additional interactive exercises in a variety of formats (eg, multiple choice and short answer questions). This digital version of the course was then pilot tested with 5 volunteers. Final

revisions were made based on volunteers' feedback as well as observations of their performance.

Participants

Inclusion and Exclusion Criteria

We sought a sample with somewhat elevated distress through our recruiting methods (ie, by framing the program as a way to reduce stress), but we did not exclude participants with low distress because we (1) did not want to make it more difficult for participants to find eligible partners and (2) hoped that CMH may be a useful tool for prevention and personal growth even in the absence of current symptoms. Although we expected that this peer counseling model would be appropriate for those with more severe symptoms, we decided to limit initial testing of the course to those with milder distress for safety and ethical reasons. Therefore, we excluded individuals scoring more than 2 SDs above general population norms on the Brief Symptom Inventory (BSI) [85] or responding in the affirmative to the BSI item on suicidal thoughts.

We excluded individuals currently receiving psychotherapy, given that they already have access and willingness to seek care and, therefore, are not in CMH's highest-priority target population. We did not exclude those taking psychiatric medication because it may be a weaker indicator of access to care (eg, some people might be prescribed medication through a general practitioner without having access to specialist treatment).

Additional eligibility criteria included being aged 18 years or older; having access to an internet-connected computer; and being able to speak, read, and write in English.

Sampling and Recruitment Method

Participants were recruited from several medium-sized towns (population 20,000-40,000) in the Western Massachusetts region. The study was advertised via flyers, web-based classifieds, and announcements on listservs and in college student groups. Advertisements presented the program as an opportunity to learn skills to reduce stress and to develop closeness with another person.

Recruitment followed a multistep process in which a first participant was recruited and screened, and then that individual recruited a peer from their existing social network. The first participant was discouraged from selecting first-degree relatives, romantic partners, or individuals with whom their relationships were characterized by conflict or disagreement. However, to increase the external validity of the study, no potential peer pairings were forbidden. Of the 30 initial participants enrolled in the study, 29 (97%) participated with their first-choice peers and 1 (3%) participated with her second-choice peer.

The sample of 60 individuals (30 pairs) comprised adult community members (18/60, 30%) and full-time undergraduate students (42/60, 70%). Of the 60 individuals, 42 (70%) identified as women, 15 (25%) as men, and 3 (5%) as transgender or gender nonconforming. They were aged 18 to 62 (median 20.5) years. The most common racial and ethnic identities were White non-Hispanic or Latinx (35/60, 58%), East Asian (10/60, 17%), and White Hispanic or Latinx (5/60, 8%), with the remainder identifying as South Asian (4/60, 7%), multiracial (3/60, 5%), Black (2/60, 3%), and Native American Hispanic or Latinx (1/60, 2%).

Measures

Psychological Distress

To assess psychological symptoms, we administered the BSI [85]. The BSI is a 53-item measure on which respondents rate symptoms experienced within the past week on 9 mental illness dimensions, from which an index of total distress can be calculated. This measure was chosen because it assesses symptoms of a range of disorders and summarizes them in a single index, has strong psychometric properties, and has published norms for patient and nonpatient populations. In the sample of this study, the BSI showed strong internal consistency in all sessions (coefficient α range=.95-.97).

To assess perceived stress, we administered the 10-item Perceived Stress Scale (PSS-10) [86]. This measure was chosen because although scores are correlated with psychological symptoms, the construct of stress as measured by this scale is distinct from mental illness and predicts future symptoms above and beyond current symptom measures [86]. The PSS-10 has been validated in numerous studies, and published norms also exist. In the sample of this study, internal consistency was good at all time points (coefficient α range=.82-.88).

Coding System for Skill Performance

The performance of participants in the helper role was evaluated using a study-specific coding system based on the psychometrically established Helping Skills Scale [87]. This coding system was developed specifically to assess participants' use of the skills taught in the course (and avoidance of proscribed behaviors).

Conversational turns are segmented into sentence-like grammatical units, and each unit is coded as falling within a certain category. The system is not intended to capture all possible categories of verbal utterances, but instead codifies behaviors that are prescribed or proscribed in the CMH course or that are very common in social support interactions. The coding system includes 6 mutually exclusive categories: *restatement* and *open-ended question* (central CMH skills), *closed-ended question* (discouraged by the course), *self-disclosure* and *sympathy* (common response modes that are neither prescribed nor explicitly proscribed, although we regarded excessive self-disclosure as evidence of failure to focus on the talker), and *other*. The system also includes a nonmutually exclusive category called *influencing*. Any speech unit in which the helper attempts to problem solve or change the talker's emotional response (which is proscribed by the course) is coded as influencing, in addition to its classification in 1 of the 6 primary categories. The 8 outcome variables for RQ 1 were the total number of sentence units uttered and the proportion of speech units in each category (the 6 mutually exclusive categories plus influencing).

Although these proportions provide a detailed profile of how helper behaviors change, they do not reveal whether learners increase their overall adherence to the guidelines given in the

course. Therefore, to address RQ 2, we created a composite index of adherence derived from the coded speech units. Participants are awarded points for engaging in behaviors encouraged by the course and are docked points for proscribed behaviors (eg, they earn points if restatements form a high proportion of the session; points are subtracted depending on the number of units of advice giving). This scale has a theoretical range of −50 to +25.

The coding system was applied by the first author, who developed the system, and a team of 9 trained undergraduate research assistants, all of whom were blinded to session condition and time point. Psychometrics and training procedures are reported in Multimedia Appendix 1 [88-94].

Perceived Session Helpfulness

To address RQ 3, both participants in each dyad rated how helpful the sessions were to them using the Crowdsourcing Mental Health Session Reaction Scale (CSRS; see Multimedia Appendix 2 for the instrument), a modified version of the Revised Session Reaction Scale [88] that focuses on their experiences when they were in the talker role. The CSRS items loaded on 2 subscales: task reactions (6 items), which reflect progress toward the resolution of the problem through insight, emotional relief, or problem solving, and relationship reactions (3 items), which reflect feeling understood by, connected to, and supported by one's peer. Thus, this measure addresses both types of potential mechanisms of peer counseling: resolution of distress and increased perceptions of closeness and support. Both subscales have a theoretical range of 1 to 9. Internal consistency of each subscale was good at all laboratory visits (coefficient of α range=.86-.92 for task reactions and .84-.95 for relationship reactions). Multimedia Appendix 1 provides scale development details.

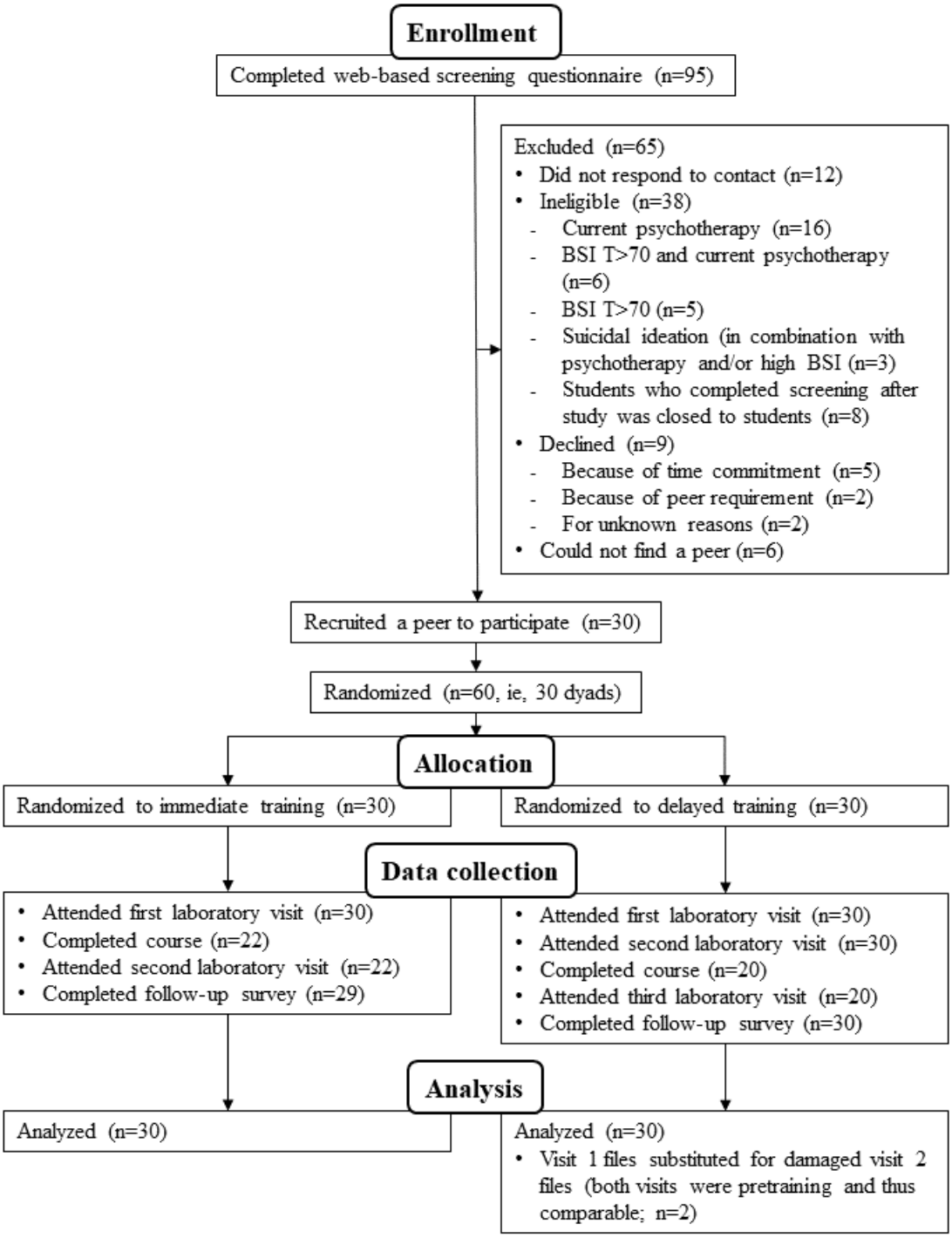
Procedure

All study procedures were approved by the University of Massachusetts Amherst institutional review board.

Study Design

This randomized experiment used a pretest-posttest wait-list controlled design to assess whether participants' behavior changed because of taking the course. Half of the dyads were randomized to an immediate training condition and half to a wait-list control (ie, delayed training) condition using a random number generator. The dyads in the immediate training condition were recorded while discussing stressors before and after completing the course over a 4-week period, whereas the dyads in the delayed training condition engaged in 2 stressor discussions separated by 4 weeks, then took the course for 4 weeks, and ultimately completed a final stressor discussion. Participants in both conditions were contacted weekly to address any questions or concerns. Figure 1 depicts the participants' flow through the study.

Figure 1. Study flow diagram. N denotes the total number of individuals, not the number of dyads. BSI: Brief Symptom Inventory.



The randomized wait-list controlled element of this study design enabled us to determine whether behavioral changes could be causally attributed to the course by evaluating between-group differences in behavior change from the first to second laboratory visit. Collecting data on *all* participants' pre- and postcourse behavior allowed us to obtain a more precise estimate of the magnitude of behavior change by analyzing data from all participants in a pre-post design.

Stressor Discussions

Stressor discussions were administered by trained research assistants according to a script and took place in treatment rooms in the university's clinical psychology training clinic, which provided an intimate, comfortable setting along with a means for nonintrusive video recording. At each laboratory visit, participants took turns talking and listening about stressors, taking 30 min each in the talker and helper roles. The order of turn-taking was determined by a coin flip. A careful procedure

for selecting stressors was used so that the severity of stressors was comparable across laboratory sessions: at visit 1, participants named 3 current stressors, rated their severity, and chose the second-most severe stressor; at subsequent visits, participants named 3 current stressors they had not previously discussed, rated their severity, and chose the stressor closest in severity to the stressor discussed previously.

In the precourse sessions, participants were told to disclose and respond as they would naturally. After taking the course, they were told to talk and respond using the skills they learned in the course; the instructions specified that they should use the skills "as they would when meeting outside of the lab rather than trying to impress anyone" to maximize ecological validity and reduce experimenter demand.

Participants were compensated for their time after each laboratory visit, US $50 for the precourse visits and US $70 for the postcourse visit. To minimize the impact of compensation on motivation to learn, the payment scheme was explained using language intended to encourage participants to construe payment as compensation for their laboratory visits, not for taking the course.

Data Analysis

All data analyses were planned a priori. We estimated 2 models to test the effects of the course on each outcome variable. First, to estimate the within-subjects magnitude of change from pre- to posttraining, we aggregated the pretraining and posttraining visits across conditions and tested the effect of time. Second, to establish whether changes could be attributed to the training (as opposed to, eg, repeated testing, maturation, and similar threats to internal validity) through a between-subjects analysis, we examined only the first 2 visits, testing the effects of time, condition, and their interaction to assess whether change from visit 1 to visit 2 was greater in the immediate training condition than in the delayed or wait-list condition. We used multilevel modeling to account for the nesting of time points within persons and the nesting of persons within dyads. As the limited number of data points would make such models unidentified, precluding maximum likelihood or related methods [95], we used Bayesian data analysis in the R package *brms* [89]. Bayesian data analysis produces a posterior distribution for each parameter that indicates the relative probability of all possible values in light of (1) the observed data and (2) a prior distribution that represents the possible values of the parameters as known or believed before data collection. Parameter estimates are typically summarized by the central tendency of the posterior (eg, the mean of the distribution) and the 95% credibility interval (CI), which is the range that contains 95% of the probability density of the posterior. When the 95% CI excludes 0 (or 1 in the case of odds ratios), one can conclude that an effect likely exists in the population. For all models, conservative priors were chosen, such that posterior distributions were influenced almost exclusively by the data. Unless explicitly stated otherwise, all analyses were planned a priori. More details, including model equations, are given in Multimedia Appendix 1.

Results

Participant Characteristics

Demographic and baseline clinical characteristics of the sample are provided in Table 1. The average participant's global psychological symptoms fell 1 SD above the general population (nonpatient) norms, indicating that a substantial proportion of participants were experiencing elevated distress.

Participants reported a variety of relationship types. Most (17/30, 57%) pairs were friends; 20% (6/30) of pairs were in a romantic relationship, 10% (3/30) of pairs were coworkers, 10% (3/30) of pairs were roommates, and 3% (1/30) of pairs consisted of a mother and daughter. One-third of the pairs had been acquainted for less than 1 year, and one-fourth of the pairs had known each other for more than 10 years (median relationship length=2.5 years).

Table 1. Baseline demographic and clinical characteristics.

Characteristics	Values
Age (years; n=60)	
Mean (SD)	24.6 (12.4)
Median	20.5
Gender (n=60), n (%)	
Woman	42 (70)
Man	15 (25)
Transman	2 (3)
Genderqueer woman	1 (2)
Race and ethnicity (n=60), n (%)	
White non-Hispanic or Latinx	35 (58)
East Asian	10 (17)
White Hispanic or Latinx	5 (8)
South Asian	4 (7)
Black	2 (3)
Native American, Hispanic or Latinx	1 (2)
Multiracial	3 (5)
Born outside the United States (n=60), n (%)	13 (22)
Nonnative English speakers (n=60), n (%)	8 (13)
Educational level (n=60), n (%)	
Some college education	41 (68)
Associate's or technical degree	1 (2)
Bachelor's degree	9 (15)
Some graduate or professional school	3 (5)
Graduate or professional degree	6 (10)
Marital status (n=60), n (%)	
Never married	51 (85)
Married	7 (12)
Separated or divorced	2 (3)
Household income (n=55), US $	
Mean (SD)	96,000 (86,000)
Median	80,000
Income/√ household members[a] (n=55), US $	
Mean (SD)	54,000 (43,000)
Median	42,000
Visit 1 Brief Symptom Inventory T score[b] (n=60)	
Mean (SD)	60.8 (9.3)
Median	61.5
Visit 1 Perceived Stress Scale-10 (n=60)	
Mean (SD)	18.7 (5.8)
Median	18
Ever in psychotherapy (n=60), n (%)	19 (32)

Characteristics	Values
Months in psychotherapy[c] (n=19)	
Mean (SD)	25.8 (31.7)
Median	10
Would consider psychotherapy (n=60), n (%)	52 (87)
Ever on psychiatric medication (n=60), n (%)	11 (18)
Months on psychiatric medication[c] (n=11)	
Mean (SD)	41.0 (52.4)
Median	18
Currently on psychiatric medication (n=60), n (%)	9 (15)
Would consider psychiatric medication (n=60), n (%)	44 (73)

[a]Income/√ household members is included to adjust total income for household size while accounting for economies of scale.

[b]In psychometrics, the T score refers to a normatively adjusted score with a mean of 50 and an SD of 10 (not to be confused with the *t* test statistic).

[c]Refers to total months of treatment over the course of the lifetime; these months were not necessarily one contiguous course of treatment.

Effects of the Course on Behaviors and Perceived Helpfulness

Means and SDs for each outcome variable are provided in Table 2. Effect sizes presented in the text represent within-person pre-post training changes. Table 3 displays the estimates for the coefficient of interest for each type of multilevel model. For the first model, the coefficient represents the change from pretraining to posttraining among all participants. For the second model, the coefficient represents the degree to which the change from visit 1 to visit 2 was greater in the immediate training condition than in the delayed training condition. When the 95% CI for this coefficient excludes 0 (or 1 for odds ratios), the change can be attributed to the training. The results for all other fixed and random effects are given in Multimedia Appendix 1.

Table 2. Descriptive statistics for outcome variables by visit and condition.

Outcome variable	Visit 1[a], mean (SD)	Visit 2[b], mean (SD)	Visit 3[c], mean (SD)
Total sentence units uttered			
Immediate training condition	239.3 (103.0)	112.6 (62.1)	—[d]
Delayed training condition	252.9 (113.6)	241.5 (77.7)	88.9 (64.2)
Percent restatement			
Immediate training condition	3.0 (3.9)	27.9 (25.8)	—
Delayed training condition	3.5 (3.2)	2.2 (4.0)	21.9 (17.4)
Percent influencing			
Immediate training condition	30.7 (17.6)	4.8 (6.8)	—
Delayed training condition	40.5 (21.3)	39.7 (17.0)	12.4 (18.2)
Percent open-ended questions			
Immediate training condition	2.9 (2.1)	6.1 (6.4)	—
Delayed training condition	2.1 (2.4)	3.7 (3.8)	12.7 (7.7)
Percent closed-ended questions			
Immediate training condition	16.1 (9.6)	13.9 (10.5)	—
Delayed training condition	9.1 (6.1)	10.9 (7.0)	18.1 (13.1)
Percent self-disclosure			
Immediate training condition	16.6 (13.3)	3.5 (8.2)	—
Delayed training condition	18.3 (13.1)	19.0 (14.9)	5.2 (9.6)
Percent sympathy			
Immediate training condition	16.1 (12.1)	15.3 (14.3)	—
Delayed training condition	13.4 (11.5)	12 (9.2)	8.1 (7.7)
Percent other			
Immediate training condition	45.4 (11.8)	33.4 (20.1)	—
Delayed training condition	53.5 (17)	52.3 (17.7)	34.1 (17.9)
Adherence score			
Immediate training condition	−17.8 (6.7)	4.1 (8.3)	—
Delayed training condition	−21.3 (8.3)	−22.8 (6.5)	0.9 (12.0)
CSRS[e] task reactions			
Immediate training condition	5.4 (1.5)	7.0 (1.5)	—
Delayed training condition	5.4 (1.5)	5.4 (1.5)	7.0 (1.6)
CSRS relationship reactions			
Immediate training condition	7.0 (1.4)	7.1 (1.8)	—
Delayed training condition	6.6 (1.6)	6.6 (1.6)	7.2 (1.9)

[a]n=60 at visit 1.

[b]n=22 for immediate training and n=28 for delayed training at visit 2.

[c]n=20 at visit 3 (delayed training condition only).

[d]Cells for the immediate training group are blank for visit 3 because the immediate training group only participated in two laboratory visits (see “Study Design” section).

[e]CSRS: Crowdsourcing Mental Health Session Reaction Scale.

Table 3. Results of Bayesian multilevel models estimating change from pretraining to posttraining among all participants (model 1) and the difference between the 2 conditions in change from the first to second laboratory visits (model 2).

Outcome variable	Pre-post training effect (model 1), estimate[a]	Pre-post training effect (model 1), 95% CI[b]	Visit×condition interaction (model 2), estimate	Visit×condition interaction (model 2), 95% CI
Total sentence units uttered[c]	−1.88	−2.21 to −1.54	−1.97	−2.47 to −1.46
Behavior categories[d]				
Restatement	16.44	7.92 to 34.81	28.50	6.89 to 119.1
Influencing	0.09	0.04 to 0.16	0.14	0.07 to 0.28
Open-ended questions	3.63	2.46 to 5.42	1.40	0.59 to 3.22
Closed-ended questions	1.15	0.84 to 1.55	0.63	0.36 to 1.07
Self-disclosure	0.05	0.01 to 0.16	0.09	0.03 to 0.25
Sympathy	0.86	0.59 to 1.23	1.08	0.59 to 1.95
Other	0.51	0.36 to 0.71	0.64	0.36 to 1.13
Adherence score[e]	1.64	1.39 to 1.9	1.90	1.48 to 2.32
Perceived session helpfulness[e]				
CSRS[f] task reactions	0.96	0.69 to 1.26	1.00	0.41 to 1.56
CSRS relationship reactions	0.28	−0.06 to 0.62	0.20	−0.38 to 0.77

[a]Estimate: mean of estimated posterior distribution.

[b]95% CI: 95% credibility interval.

[c]Count variable with Poisson link function; coefficients are in log units.

[d]Binomial or Bernoulli distributed variables with logistic link function; coefficients are in odds ratio units. For model 1, the estimate represents the relative odds of the behavior at posttraining compared with pretraining; for model 2, the estimate represents the degree to which the relative odds of a behavior at visit 2 compared with visit 1 were higher or lower in the group that underwent immediate training.

[e]Metric variables with identity link function; coefficients are in standardized units (ie, SDs).

[f]CSRS: Crowdsourcing Mental Health Session Reaction Scale.

RQ 1: How Much Does the Course Change the Use of Specific Helping Skills?

At baseline, participants' behaviors were typical of untrained supportive conversations. We observed signs of positive intentions and a lack of hostility (eg, criticism was rare, and advice and encouragement were common). However, helpers did not spontaneously display several other behaviors recommended by supportive psychotherapy guidelines. For example, participants spent more than one-third (34%) of the pretraining session, on average, trying to influence the talker through advice giving and related behaviors, and they delivered relatively few restatements or open-ended questions (approximately 3% on average for both categories; note that the averages reported in the text of this section represent mean values, aggregating across both the immediate and delayed training groups).

After training, we observed that participants changed their behavior to more closely match the supportive psychotherapy guidelines taught. As evidenced by the between-group comparisons, the course had strong effects on several of these baseline behaviors in line with our hypotheses. Helpers decreased their overall volume of speech (d=−1.5 for within-person change from pretraining to posttraining) from an average of 163 (SD 82) utterances per 30 min of discussion time before training to an average of 34 (SD 36) utterances posttraining. They increased their frequency of restatements (d=1.0); on average, restatements formed 3% (SD 4%) of the session at baseline and grew to 25% (SD 22%) posttraining. Helpers also decreased average attempts to influence the talker (d=−1.6) from 34% (SD 18%) of the session at baseline to 8% (SD 14%) posttraining. Taking the course decreased self-disclosure (d=−0.8) from 18% (SD 14%) to 4% (SD 9%) and speech behaviors in the *other* category (d=−0.8) from 48% (SD 15%) to 34% (SD 19%).

Evidence of an effect on open-ended questions was more equivocal: open-ended questions increased (d=0.8) from around 3% (SD 3%) to 9% (SD 8%), but the 95% CI for the visit by condition interaction included an odds ratio of 1, so one cannot claim with certainty that change was because of the course. There was no strong indication that participants changed closed-ended questions (d=0.2) or expressions of sympathy (d=−0.1).

RQ 2: Does the Course Increase Helpers' Overall Adherence to Helping Skills Guidelines?

Adherence scores greatly increased from an average of −20.0 (SD 7.0) at pretraining to +2.6 (SD 10.2) at posttraining (d=2.1). The between-group analysis showed that this change can be causally attributed to training.

RQ 3: Does the Talker's Perception of Session Helpfulness Increase After Taking the Course?

On the CSRS task reactions scale, participants indicated that they perceived more progress in developing insight and solving problems in their sessions after taking the course (d=1.1), going from an average score of 5.4 (SD 1.5) to an average score of 7.0 (SD 1.6). In contrast, there was no reliable evidence for change in the CSRS relationship reactions subscale, which represents feelings of understanding and support between peers (d=0.3); the average score was 6.8 (SD 1.5) pretraining and 7.1 (SD 1.9) posttraining.

Accounting for Attrition as a Potential Confound

Overall, 30% (9/30) of dyads withdrew, 4 of the 9 (44%) from the immediate training condition and 5 of the 9 (56%) from the delayed training condition (not a significant difference; χ^2_1=0.1; P=.79). The most endorsed reasons for attrition were difficulty finding time or motivation to work on the course, stress from the additional workload conferred by the course, and interference from unanticipated life events. There were no differences in psychological symptoms or stress between those who withdrew and those who did not, and there were no differences in demographic characteristics. Individuals who withdrew were more likely to report past psychotherapy (66.7% vs 16.7%; χ^2_1=12.3; P<.001) and current psychiatric medication use (33.3% vs 7.1%; χ^2_1=6.8; P=.02) than individuals who completed the study. There was also a marginally significant difference in household income (divided by the square root of the number of household members to adjust for household size and economies of scale), with those who withdrew coming from higher-income households (median US $61,500 for withdrawers and US $35,800 for completers; two-tailed $t_{25.4}$=1.94; P=.06).

In most trials, participants decide whether to withdraw of their own accord, raising the possibility that differences in outcome are due to self-selection rather than to the effects of the intervention. In this study, attrition from the study took place pairwise: if one participant wished to exit the study, that person's peer left as well; consequently, withdrawal from the study was not perfectly correlated with intention to remain in the study. All participants (regardless of whether they left the study prematurely) retrospectively rated on a 10-point Likert scale how much they had wanted to withdraw versus remain. Withdrawers indicated a greater desire to leave the study (mean 5.8, SD 2.1) than completers (mean 4.5, SD 2.1; d=0.60), although the difference failed to achieve statistical significance (P=.09). Thus, there is still some possibility of self-selection affecting the results, such that participants who withdrew from the study might have shown no skill improvement, attenuating effect sizes.

By statistically controlling for the desire to withdraw, one can potentially model the missing data mechanism so that the assumption of missingness at random is met, reducing or eliminating bias in effect estimates [96]. Therefore, we conducted a post-hoc analysis in which we re-estimated the models used to investigate changes from pre- to posttraining, now while controlling for desire to withdraw and the interaction between desire and time point. All of the 95% CIs for effects of motivation to withdraw included 0, and other coefficients remained similar, suggesting that attrition is unlikely to be a meaningful source of bias in the results. The detailed results are presented in Multimedia Appendix 1.

Discussion

Principal Findings

The goal of this study was to test the efficacy of a web-based course for teaching counseling skills to nonprofessionals, including those with elevated psychological symptoms. The course caused participants to change most of their helper speech behaviors in the hypothesized directions. Participants spoke less during a mock CMH session and they spent less time talking about themselves, suggesting that they learned to focus their attention on the talker. They increased their use of restatements and decreased their attempts to influence the talker. They also slightly increased their use of open-ended questions, although there was insufficient evidence that this increase was caused by taking the course, and there was no decrease in closed-ended questions. Overall, participants showed substantial increases in aggregate adherence scores. In addition, participants reported more progress in problem solving and insight during counseling sessions after taking the course, which may indicate that peer counseling using this model could improve mental health.

These findings provide cause for optimism that nonprofessionals can learn to deliver therapeutic ingredients via primarily self-directed web-based courses. This model—reciprocal peer delivery of techniques derived from supportive psychotherapy that are taught via a self-directed web-based course—has a variety of advantages that enable it to address gaps in traditional mental health care. First, it addresses practical barriers to treatment access because it does not require working with professionals (who often have limited availability), has no financial cost, and can be conducted in flexible times and places. Second, it addresses attitudinal barriers by not requiring participants to identify as mentally ill or see themselves as needing help (instead, they are in an egalitarian relationship). Third, it addresses gaps in the treatment of subclinical symptoms and distress by using a transdiagnostic treatment (supportive psychotherapy) that is appropriate even in the absence of a diagnosable mental illness. In addition to these gap-addressing features, it has the additional potential to increase feelings of intimacy and perceived social support and to provide the psychological benefits of delivering care in a way that is not present in psychotherapy with a professional.

The evidence suggests that supportive psychotherapy is efficacious for a variety of conditions, but it may not be as powerful as other treatments (eg, those that target specific symptoms of a disorder), especially if delivered by peers. However, even if such peer-delivered interventions are not as powerful as those delivered by professionals and even if only a subset of laypeople can learn the skills, disseminating therapeutic ingredients through nonprofessionals could improve public health by reaching large numbers of people who might not otherwise receive mental health support. One can imagine numerous permutations of peer-delivered interventions for the many settings where need is great and access or willingness to

use traditional psychotherapy is low. Continued research and development of web-based training programs such as CMH that use the reciprocal peer counseling model is warranted.

Despite the clear impact of the course on most behaviors and perceived helpfulness, this study also suggests that refinements might be needed to improve its efficacy and reach. Perhaps the largest concern is attrition: 30% of the participants chose to withdraw from the study. Although this value is comparable with dropout rates for psychotherapy trials [97,98] and trials of smartphone apps to treat anxiety and depression [48] and is low relative to the 80% to 90% rate often cited for massive open online courses [99], it suggests the course could be modified to increase motivation or decrease learner burden. Interestingly, individuals who withdrew from the study were no more psychologically distressed or symptomatic than those who continued, but they were more likely to have experience using professional mental health services and had marginally higher household income. They may have had greater access to or comfort with traditional treatment and thus felt a less pressing need to learn an alternative tool for mental health support. Regardless of their reasons for withdrawing, some participants clearly found the course burdensome; although the course had no financial cost, the version used in this study requires considerable time and effort. Maximizing scalability for CMH and related courses means minimizing the time and effort cost without compromising efficacy. Anecdotally, participants seemed to find the lessons on talking more onerous than the lessons on helping; thus, for future iterations of the course, we plan to reduce or eliminate the talking skills lessons and replace them with real-time, in-session topic prompts [100], in addition to using participant feedback to make the helper lessons more enjoyable.

Finally, reported feelings of interpersonal closeness and support assessed via the CSRS relationship reactions subscale remained stable (there was a small increase, but a zero increase was a credible value in the Bayesian models). This may be attributable to a ceiling effect: participants' ratings of their relationship-related perceptions in the first mock session were high. It is also likely that measurable changes in perceptions of support giving in relationships require more than one counseling session, especially in established relationships in which perceptions of the other person's supportive behavior may draw on information from numerous interactions. Nevertheless, future iterations of the course can draw on close relationships and communication research to identify more ways to foster feelings of closeness and support.

Generalizability

The study's sample was fairly culturally diverse, with approximately 40% of participants identifying as non-White (compared with about 25% of the US population), and more than 1 in 5 participants born outside the United States. Several were recent English language learners, and one of these informally commented that she found the English of the course accessible and useful for practicing her English skills. The success of the course with this sample suggests that it is at least effective in teaching people with diverse cultural and linguistic backgrounds. However, this does not mean that it will be successful in improving the mental health of individuals from all cultures, especially considering that culturally adapted psychotherapy is more effective [101] and that there are cultural differences in preferred and delivered social support styles [102,103].

In addition, the course's educational efficacy for individuals with less formal education or technology experience remains unknown because most participants were college educated or current students and, thus, may have been particularly well-equipped to learn from the course. Adaptations may be warranted for other populations. Fortunately, even if delivering these skills in this format to individuals with less education or comfort with technology is found to be impractical, CMH could still have a public health impact. It could, for example, be deployed with college students, addressing rising psychological distress and the shortage of mental health services on college campuses [104]. However, it would be ideal to make CMH accessible to as many individuals as possible; thus, further research must assess whether a redesign is needed to reach those without a college education.

Although we made efforts to make the study as ecologically valid as possible (eg, framing compensation as payment for study visits rather than completing the course; encouraging participants to speak "as [they] really would in everyday life, without trying to impress anybody"), it must be acknowledged that learners might engage with the course material differently or adhere less to the supportive psychotherapy skills if they are not monitored by research staff or financially incentivized to participate. This limitation to ecological validity highlights the need to make the course truly intrinsically motivating to facilitate adherence.

Limitations and Future Directions

In addition to some limitations to generalizability, the study's scope (ie, assessing the impact of the course on skill performance immediately after training) limits the conclusions that can be drawn. In particular, the mental health impact of applying CMH skills has not been rigorously investigated. As a proxy for the impact of the intervention, participants rated the perceived helpfulness of their sessions, but there is no guarantee that what participants viewed as helpful in the short run would have positive effects on psychological symptoms in the long run. Furthermore, because participants were not blinded to condition, the measured increase in perceived session helpfulness could be driven by a placebo effect or experimenter demand.

Furthermore, because the mental health effects of engaging in these peer counseling sessions were not assessed beyond the immediate postsession reaction, we could not thoroughly assess any harm or risks that could result from participating. For example, it is possible that peers could use these sessions as an opportunity to air grievances with each other or to gossip about mutual acquaintances, damaging their relationships. We attempted to mitigate this possibility by including instructions in the course that (1) discourage partnering with a peer with whom one has a contentious or familial relationship and (2) proscribe discussing a stressor that directly involves one's

partner (instead, we encouraged learners to find a neutral third party with whom to discuss such topics).

Peers could also disclose extremely troubling or traumatic material, or indications of risk such as suicidal thoughts, to which nonprofessionals rarely have the skills to respond. For the purposes of this study, participants were instructed not to discuss such material with their peers, but instead to contact a professional or the research staff if they needed to talk about such topics. (They were provided with extensive local referral information.) If the course were launched to the general public, additional safeguards would need to be put in place, such as easy access to crisis hotlines along with materials that have been empirically demonstrated to increase utilization of crisis services [105] and instructions for responding to a suicidal peer, among others.

A potentially more common risk is that peer counseling partners engage in co-rumination, a process that involves repetitive discussion and speculation about problems and has demonstrable negative as well as positive effects (ie, on relationships and mental health) [106-108]. This risk was one of the reasons we included instructions for the talker that encouraged talkers to identify proactive coping or problem-solving actions within a session or two rather than repeatedly rehashing problems. Despite the inclusion of these mitigations, it is impossible to perfectly control peer counseling behaviors, so harm could result from participating. The question is whether the benefits outweigh these risks and whether these harms would have occurred anyway in the absence of a formal peer counseling program (ie, friends may disclose distressing material or suicidal thoughts in everyday life, and the addition of a structured counseling program may not affect the frequency of such disclosures).

An additional limitation of this study is the absence of a follow-up timepoint to assess the durability of the training. It is possible that as they engage in repeated reciprocal peer counseling sessions over time, CMH users may forget the material or drift toward their typical interaction styles. Periodic self-assessments and booster training sessions may be needed to maintain skills over time.

Future work should aim to remedy the limitations of this investigation. As a first step, the course must be revised to reduce the time and effort required to complete it (eg, by replacing the talker lessons with in-session discussion prompts). Next, it is essential to assess the longer-term mental health effects as well as risks of harm that result from engaging in repeated CMH sessions with a peer. These studies should be conducted in naturalistic settings, to the degree possible, while maintaining some monitoring (eg, through regular assessment) for ethical reasons. These longer-term studies must also assess whether skills erode over time and whether such erosion can be prevented with self-assessment tools (to check whether one has followed guidelines) and/or booster training sessions.

To maximize its reach, CMH is also likely to require tailoring to specific populations (eg, individuals with less education or experience with technology, particular cultural groups, or others). The tailoring process can begin with qualitative research (eg, focus groups, interviews, and pilot testing) to drive initial revisions to the course, followed by experimental assessments of the updated course's effects on skills and the longer-term mental health effects. Importantly, early steps in this process may reveal that CMH's defining features do not adequately address the unique needs or barriers experienced by a particular group and that it is necessary to develop entirely different innovative interventions.

When the course is eventually launched at scale, frequent A/B testing (controlled experiments comparing two different versions of a website or software) can be used to fine-tune it to be as effective as possible, to tailor it further to specific populations, to eliminate exercises that do not increase skill, and to make other refinements.

Conclusions

This study demonstrates the feasibility of teaching empirically supported counseling skills to pairs of nonprofessionals via a highly scalable web-based course. Although the model may not be able to reach all populations, this study demonstrates the potential of the CMH model to fill important gaps in the current mental health care system. Further research and refinement are necessary to assess the mental health effects of the course and to ensure that it is effective for diverse groups. Our results underscore that reciprocal, peer-delivered interventions disseminated via web-based courses have the potential to fill gaps in mental health care, thus enabling evidence-based treatment ingredients to reach individuals who might otherwise not be served by the existing mental health care system.

Acknowledgments

The research reported in this publication was supported in part by a fellowship from the National Institute of Mental Health of the National Institutes of Health under award number F31MH103927 and grants from the Society for the Advancement of Psychotherapy, the Society for Psychotherapy Research, the American Psychological Association, and the University of Massachusetts Amherst. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or any other funding bodies. The authors wish to thank the many research participants, research assistants, and volunteers who made this study possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data analytic methods and results.
[DOCX File , 73 KB - jmir_v22i9e17164_app1.docx]

Multimedia Appendix 2
Crowdsourcing Mental Health Session Reaction Scale.
[DOCX File , 17 KB - jmir_v22i9e17164_app2.docx]

References

1. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
2. Rotheram-Borus MJ, Swendeman D, Chorpita BF. Disruptive innovations for designing and diffusing evidence-based interventions. Am Psychol 2012 Sep;67(6):463-476 [FREE Full text] [doi: 10.1037/a0028180] [Medline: 22545596]
3. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. Annu Rev Public Health 2008;29:115-129. [doi: 10.1146/annurev.publhealth.29.020907.090847] [Medline: 18348707]
4. World Health Organization. The Global Burden of Disease: 2004 Update. Geneva, Switzerland: World Health Organization; 2008.
5. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO world mental health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
6. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the national comorbidity survey replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
7. Cohen S, Janicki-Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Psychol 2012;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]
8. Seeman TE, McEwen BS, Rowe JW, Singer BH. Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. Proc Natl Acad Sci U S A 2001 Apr 10;98(8):4770-4775 [FREE Full text] [doi: 10.1073/pnas.081072698] [Medline: 11287659]
9. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
10. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
11. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
12. Renn BN, Hoeft TJ, Lee HS, Bauer AM, Areán PA. Preference for in-person psychotherapy versus digital psychotherapy options for depression: survey of adults in the US. NPJ Digit Med 2019;2:6 [FREE Full text] [doi: 10.1038/s41746-019-0077-1] [Medline: 31304356]
13. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided ehealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
14. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
15. Bohannon J. Artificial intelligence. The synthetic therapist. Science 2015 Jul 17;349(6245):250-251. [doi: 10.1126/science.349.6245.250] [Medline: 26185240]
16. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
17. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
18. Bernecker SL, Banschback K, Santorelli GD, Constantino MJ. A web-disseminated self-help and peer support program could fill gaps in mental health care: lessons from a consumer survey. JMIR Ment Health 2017 Jan 19;4(1):e5 [FREE Full text] [doi: 10.2196/mental.4751] [Medline: 28104578]
19. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]

20. Cuijpers P, Driessen E, Hollon SD, van Oppen P, Barth J, Andersson G. The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. Clin Psychol Rev 2012 Jun;32(4):280-291. [doi: 10.1016/j.cpr.2012.01.003] [Medline: 22466509]
21. Winston A, Rosenthal R, Pinsker K. Introduction to Supportive Psychotherapy. Arlington, VA: American Psychiatric Publishing; 2004.
22. Dunkel SC. Social support and cancer: findings based on patient interviews and their implications. J Soc Issues 1984;40(4):77-98. [doi: 10.1111/j.1540-4560.1984.tb01108.x]
23. Ingram KM, Betz NE, Mindes EJ, Schmitt MM, Smith NG. Unsupportive responses from others concerning a stressful life event: development of the unsupportive social interactions inventory. J Soc Clin Psychol 2001 Jun;20(2):173-207. [doi: 10.1521/jscp.20.2.173.22265]
24. Beidas RS, Kendall PC. Training therapists in evidence-based practice: a critical review of studies from a systems-contextual perspective. Clin Psychol (New York) 2010 Mar;17(1):1-30 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01187.x] [Medline: 20877441]
25. Rakovshik SG, McManus F. Establishing evidence-based training in cognitive behavioral therapy: a review of current empirical findings and theoretical guidance. Clin Psychol Rev 2010 Jul;30(5):496-516. [doi: 10.1016/j.cpr.2010.03.004] [Medline: 20488599]
26. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 2013;10(5):e1001454 [FREE Full text] [doi: 10.1371/journal.pmed.1001454] [Medline: 23723742]
27. Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. J Consult Clin Psychol 2003 Dec;71(6):973-979. [doi: 10.1037/0022-006X.71.6.973] [Medline: 14622072]
28. Braun SR, Gregor B, Tran US. Comparing bona fide psychotherapies of depression in adults with two meta-analytical approaches. PLoS One 2013;8(6):e68135 [FREE Full text] [doi: 10.1371/journal.pone.0068135] [Medline: 23840824]
29. Gould RL, Coulson MC, Howard RJ. Efficacy of cognitive behavioral therapy for anxiety disorders in older people: a meta-analysis and meta-regression of randomized controlled trials. J Am Geriatr Soc 2012 Feb;60(2):218-229. [doi: 10.1111/j.1532-5415.2011.03824.x] [Medline: 22283717]
30. Hall J, Kellett S, Berrios R, Bains MK, Scott S. Efficacy of cognitive behavioral therapy for generalized anxiety disorder in older adults: systematic review, meta-analysis, and meta-regression. Am J Geriatr Psychiatry 2016 Nov;24(11):1063-1073. [doi: 10.1016/j.jagp.2016.06.006] [Medline: 27687212]
31. Jørgensen CR, Freund C, Bøye R, Jordet H, Andersen D, Kjølbye M. Outcome of mentalization-based and supportive psychotherapy in patients with borderline personality disorder: a randomized trial. Acta Psychiatr Scand 2013 Apr;127(4):305-317. [doi: 10.1111/j.1600-0447.2012.01923.x] [Medline: 22897123]
32. Cottraux J, Note I, Yao SN, de Mey-Guillard C, Bonasse F, Djamoussian D, et al. Randomized controlled comparison of cognitive behavior therapy with Rogerian supportive therapy in chronic post-traumatic stress disorder: a 2-year follow-up. Psychother Psychosom 2008;77(2):101-110. [doi: 10.1159/000112887] [Medline: 18230943]
33. Lipsitz JD, Gur M, Vermes D, Petkova E, Cheng J, Miller N, et al. A randomized trial of interpersonal therapy versus supportive therapy for social anxiety disorder. Depress Anxiety 2008;25(6):542-553. [doi: 10.1002/da.20364] [Medline: 17941096]
34. Stanley MA, Beck JG, Glassco JD. Treatment of generalized anxiety in older adults: a preliminary comparison of cognitive-behavioral and supportive approaches. Behav Ther 1996;27(4):565-581. [doi: 10.1016/s0005-7894(96)80044-x]
35. McIntosh VV, Jordan J, Carter FA, Luty SE, McKenzie JM, Bulik CM, et al. Three psychotherapies for anorexia nervosa: a randomized, controlled trial. Am J Psychiatry 2005 Apr;162(4):741-747. [doi: 10.1176/appi.ajp.162.4.741] [Medline: 15800147]
36. Hellerstein DJ, Rosenthal RN, Pinsker H, Samstag LW, Muran JC, Winston A. A randomized prospective study comparing supportive and dynamic therapies. Outcome and alliance. J Psychother Pract Res 1998;7(4):261-271 [FREE Full text] [Medline: 9752637]
37. Markowitz JC, Kocsis JH, Christos P, Bleiberg K, Carlin A. Pilot study of interpersonal psychotherapy versus supportive psychotherapy for dysthymic patients with secondary alcohol abuse or dependence. J Nerv Ment Dis 2008 Jun;196(6):468-474. [doi: 10.1097/NMD.0b013e31817738f1] [Medline: 18552624]
38. Markowitz JC. What is supportive psychotherapy? Focus 2014 Jan;12(3):285-289. [doi: 10.1176/appi.focus.12.3.285]
39. Constantino MJ, Bernecker SL. Bridging the common factors and empirically supported treatment camps: comment on Laska, Gurman, and Wampold. Psychotherapy (Chic) 2014 Dec;51(4):505-509. [doi: 10.1037/a0036604] [Medline: 25419731]
40. Tolin DF. Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. Clin Psychol Rev 2010 Aug;30(6):710-720. [doi: 10.1016/j.cpr.2010.05.003] [Medline: 20547435]
41. Tolin DF. Beating a dead dodo bird: looking at signal vs noise in cognitive-behavioral therapy for anxiety disorders. Clin Psychol Sci Pract 2014 Dec 15;21(4):351-362. [doi: 10.1111/cpsp.12080]

42. Hoeft TJ, Fortney JC, Patel V, Unützer J. Task-sharing approaches to improve mental health care in rural and other low-resource settings: a systematic review. J Rural Health 2018 Dec;34(1):48-62 [FREE Full text] [doi: 10.1111/jrh.12229] [Medline: 28084667]
43. Patel V. Global mental health: from science to action. Harv Rev Psychiatry 2012;20(1):6-12 [FREE Full text] [doi: 10.3109/10673229.2012.649108] [Medline: 22335178]
44. Gleason ME, Iida M, Shrout PE, Bolger N. Receiving support as a mixed blessing: evidence for dual effects of support on psychological outcomes. J Pers Soc Psychol 2008 May;94(5):824-838 [FREE Full text] [doi: 10.1037/0022-3514.94.5.824] [Medline: 18444741]
45. Brown SL, Nesse RM, Vinokur AD, Smith DM. Providing social support may be more beneficial than receiving it: results from a prospective study of mortality. Psychol Sci 2003 Jul;14(4):320-327. [doi: 10.1111/1467-9280.14461] [Medline: 12807404]
46. Reblin M, Uchino BN. Social and emotional support and its implication for health. Curr Opin Psychiatry 2008 Mar;21(2):201-205 [FREE Full text] [doi: 10.1097/YCO.0b013e3282f3ad89] [Medline: 18332671]
47. Schwartz CE, Sendor RM. Helping others helps oneself: response shift effects in peer support. Soc Sci Med 1999 Jun;48(11):1563-1575. [doi: 10.1016/s0277-9536(99)00049-0] [Medline: 10400257]
48. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
49. McArthur D, Dumas A, Woodend K, Beach S, Stacey D. Factors influencing adherence to regular exercise in middle-aged women: a qualitative study to inform clinical practice. BMC Womens Health 2014 Mar 26;14:49 [FREE Full text] [doi: 10.1186/1472-6874-14-49] [Medline: 24666887]
50. Oussedik E, Foy CG, Masicampo EJ, Kammrath LK, Anderson RE, Feldman SR. Accountability: a missing construct in models of adherence behavior and in clinical practice. Patient Prefer Adherence 2017;11:1285-1294 [FREE Full text] [doi: 10.2147/PPA.S135895] [Medline: 28794618]
51. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul 27;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
52. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
53. Brugha TS, Weich S, Singleton N, Lewis G, Bebbington PE, Jenkins R, et al. Primary group size, social support, gender and future mental health status in a prospective study of people living in private households throughout Great Britain. Psychol Med 2005 May;35(5):705-714. [doi: 10.1017/s0033291704003903] [Medline: 15918347]
54. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago health, aging, and social relations study. Psychol Aging 2010 Jun;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]
55. Fuhrer R, Stansfeld S, Chemali J, Shipley M. Gender, social relations and mental health: prospective findings from an occupational cohort (Whitehall II study). Soc Sci Med 1999 Jan;48(1):77-87. [doi: 10.1016/s0277-9536(98)00290-1] [Medline: 10048839]
56. Laurenceau J, Barrett LF, Pietromonaco PR. Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. J Pers Soc Psychol 1998 May;74(5):1238-1251. [doi: 10.1037//0022-3514.74.5.1238] [Medline: 9599440]
57. Fairburn CG, Cooper Z. Therapist competence, therapy quality, and therapist training. Behav Res Ther 2011 Jun;49(6-7):373-378 [FREE Full text] [doi: 10.1016/j.brat.2011.03.005] [Medline: 21492829]
58. Jackson CB, Quetsch LB, Brabson LA, Herschell AD. Web-based training methods for behavioral health providers: a systematic review. Adm Policy Ment Health 2018 Jul;45(4):587-610 [FREE Full text] [doi: 10.1007/s10488-018-0847-0] [Medline: 29352459]
59. Murdock J, Williams A, Becker K, Bruce M, Young S. Online versus on-campus: a comparison study of counseling skills courses. J Hum Resour 2012;8(1):- [FREE Full text]
60. Baumel A. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. J Ment Health 2015;24(5):313-320. [doi: 10.3109/09638237.2015.1079308] [Medline: 26485198]
61. Baumel A, Schueller SM. Adjusting an available online peer support platform in a program to supplement the treatment of perinatal depression and anxiety. JMIR Ment Health 2016 Mar 21;3(1):e11 [FREE Full text] [doi: 10.2196/mental.5335] [Medline: 27001373]
62. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital peer-support platform (7cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
63. Doss BD, Cicila LN, Georgia EJ, Roddy MK, Nowlan KM, Benson LA, et al. A randomized controlled trial of the web-based OurRelationship program: effects on relationship and individual functioning. J Consult Clin Psychol 2016 Apr;84(4):285-296 [FREE Full text] [doi: 10.1037/ccp0000063] [Medline: 26999504]

64. Kalinka CJ, Fincham FD, Hirsch AH. A randomized clinical trial of online-biblio relationship education for expectant couples. J Fam Psychol 2012 Feb;26(1):159-164. [doi: 10.1037/a0026398] [Medline: 22309819]
65. Braithwaite SR, Fincham FD. Computer-based dissemination: a randomized clinical trial of ePREP using the actor partner interdependence model. Behav Res Ther 2011 Feb;49(2):126-131. [doi: 10.1016/j.brat.2010.11.002] [Medline: 21159326]
66. Hill C. Helping Skills: Facilitating Exploration, Insight, and Action. Fourth Edition. Washington, DC: American Psychological Association; 2014.
67. Ivey A, Ivey M, Zalaquett C. Intentional Interviewing and Counseling: Facilitating Client Development in a Multicultural Society. Ninth Edition. Boston, MA: Cengage Learning; 2017.
68. Herschell AD, Kolko DJ, Baumann BL, Davis AC. The role of therapist training in the implementation of psychosocial treatments: a review and critique with recommendations. Clin Psychol Rev 2010 Jun;30(4):448-466 [FREE Full text] [doi: 10.1016/j.cpr.2010.02.005] [Medline: 20304542]
69. Harned MS, Dimeff LA, Woodcock EA, Kelly T, Zavertnik J, Contreras I, et al. Exposing clinicians to exposure: a randomized controlled dissemination trial of exposure therapy for anxiety disorders. Behav Ther 2014 Nov;45(6):731-744 [FREE Full text] [doi: 10.1016/j.beth.2014.04.005] [Medline: 25311284]
70. Kobak KA, Wolitzky-Taylor K, Craske MG, Rose RD. Therapist training on cognitive behavior therapy for anxiety disorders using internet-based technologies. Cognit Ther Res 2017 Apr;41(2):252-265 [FREE Full text] [doi: 10.1007/s10608-016-9819-4] [Medline: 28435174]
71. Rakovshik SG, McManus F, Vazquez-Montes M, Muse K, Ougrin D. Is supervision necessary? Examining the effects of internet-based CBT training with and without supervision. J Consult Clin Psychol 2016 Mar;84(3):191-199. [doi: 10.1037/ccp0000079] [Medline: 26795937]
72. Taylor PJ, Russ-Eft DF, Chan DW. A meta-analytic review of behavior modeling training. J Appl Psychol 2005 Jul;90(4):692-709. [doi: 10.1037/0021-9010.90.4.692] [Medline: 16060787]
73. Hill CE, Lent RW. A narrative and meta-analytic review of helping skills training: time to revive a dormant area of inquiry. Psychotherapy (Chic) 2006;43(2):154-172. [doi: 10.1037/0033-3204.43.2.154] [Medline: 22122035]
74. Klein C. What Do We Know About Interpersonal Skills? A Meta-analytic Examination Of Antecedents, Outcomes, And The Efficacy Of Training. University of Central Florida. 2009. URL: https://stars.library.ucf.edu/etd/3950/ [accessed 2020-09-03]
75. Karpicke JD. Retrieval-based learning. Curr Dir Psychol Sci 2012 May 30;21(3):157-163. [doi: 10.1177/0963721412443552]
76. Weeks D, Anderson L. The interaction of observational learning with overt practice: effects on motor skill learning. Acta Psychol (Amst) 2000 May;104(2):259-271. [doi: 10.1016/s0001-6918(00)00039-1] [Medline: 10900708]
77. Clark R, Mayer R. E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. Fourth Edition. Hoboken, UK: Wiley; 2016.
78. Brydges R, Carnahan H, Rose D, Rose L, Dubrowski A. Coordinating progressive levels of simulation fidelity to maximize educational benefit. Acad Med 2010 May;85(5):806-812. [doi: 10.1097/ACM.0b013e3181d7aabd] [Medline: 20520031]
79. Kulkarni C, Wei KP, Le H, Chia D, Papadopoulos K, Cheng J, et al. Peer and self assessment in massive online classes. ACM Trans Comput-Hum Interact 2013 Dec 1;20(6):1-31. [doi: 10.1145/2505057]
80. Piech C, Huang J, Chen Z, Do C, Ng A, Koller D. MOOCS: What Have We Learned? In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: KDD'15; August 10-15, 2015; Sydney, Australia. [doi: 10.1145/2783258.2785465]
81. Shute VJ. Focus on formative feedback. Rev Educ Res 2008 Mar;78(1):153-189. [doi: 10.3102/0034654307313795]
82. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bulltin 1996 Mar;119(2):254-284. [doi: 10.1037/0033-2909.119.2.254]
83. Zell E, Krizan Z. Do people have insight into their abilities? A metasynthesis. Perspect Psychol Sci 2014 Mar;9(2):111-125. [doi: 10.1177/1745691613518075] [Medline: 26173249]
84. Rosengren D. Building Motivational Interviewing Skills: A Practitioner Workbook. New York, USA: Guilford Press; 2009.
85. Derogatis L. BSI, Brief Symptom Inventory: Administration, Scoring & Procedures Manual. Bloomington, MN: PsychCorp; 1993.
86. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: Claremont Symposium on Applied Social Psychology. Newbury Park, CA: Sage; 1988:31-67.
87. Hill C, O'Brien K. Helping Skills: Facilitating Exploration, Insight, and Action. Washington, DC: American Psychological Association; 2014.
88. Elliott R. The Revised Session Reaction Scale. Toledo, OH: University of Toledo; 1993.
89. Bürkner P. brms: an R package for Bayesian multilevel models using Stan. J Stat Soft 2017;80(1):1-28. [doi: 10.18637/jss.v080.i01]
90. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
91. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992 Nov;7(4):457-472 [FREE Full text] [doi: 10.2307/2246093]

92. Radford MN. MCMC using Hamiltonian dynamics. Handbook of Markov Chain Monte Carlo.: University of Toronto URL: https://arxiv.org/pdf/1206.1901.pdf [accessed 2020-09-03]
93. Rosseel Y. lavaan: an R Package for structural equation modeling. J Stat Soft 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
94. Stan Development Team. The Stan C++ Library. URL: https://mc-stan.org/ [accessed 2020-09-08]
95. McMahon JM, Pouget ER, Tortu S. A guide for multilevel modeling of dyadic data with binary outcomes using SAS PROC NLMIXED. Comput Stat Data Anal 2006 Aug;50(12):3663-3680 [FREE Full text] [doi: 10.1016/j.csda.2005.08.008] [Medline: 16926924]
96. Leon AC, Demirtas H, Hedeker D. Bias reduction with an adjustment for participants' intent to dropout of a randomized controlled clinical trial. Clin Trials 2007;4(5):540-547. [doi: 10.1177/1740774507083871] [Medline: 17942469]
97. Swift JK, Greenberg RP. A treatment by disorder meta-analysis of dropout from psychotherapy. J Psychother Integr 2014;24(3):193-207. [doi: 10.1037/a0037512]
98. Wierzbicki M, Pekarik G. A meta-analysis of psychotherapy dropout. Prof Psychol: Res Pract 1993;24(2):190-195. [doi: 10.1037/0735-7028.24.2.190]
99. Clow D. An overview of learning analytics. Teach High Educ 2013 Aug;18(6):683-695. [doi: 10.1080/13562517.2013.827653]
100. O'Leary K, Schueller S, Wobbrock J, Pratt W. 'Suddenly, We Got to Become Therapists for Each Other': Designing Peer Support Chats for Mental Health. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI 2018; April 21-26, 2018; Montréal, Canada. [doi: 10.1145/3173574.3173905]
101. Benish SG, Quintana S, Wampold BE. Culturally adapted psychotherapy and the legitimacy of myth: a direct-comparison meta-analysis. J Couns Psychol 2011 Jul;58(3):279-289. [doi: 10.1037/a0023626] [Medline: 21604860]
102. Burleson BR. The experience and effects of emotional support: what the study of cultural and gender differences can tell us about close relationships, emotion, and interpersonal communication. Pers Relat 2003 Mar;10(1):1-23. [doi: 10.1111/1475-6811.00033]
103. Dutton YE. Butting in vs being a friend: cultural differences and similarities in the evaluation of imposed social support. J Soc Psychol 2012;152(4):493-509. [doi: 10.1080/00224545.2011.642025] [Medline: 22822687]
104. Thielking M. A Dangerous Wait: Colleges Can't Meet Soaring Student Needs for Mental Health Care. STAT - Reporting From the Frontiers of Health and Medicine. 2017 Feb 06. URL: https://www.statnews.com [accessed 2017-04-10]
105. Jaroszewski AC, Morris RR, Nock MK. Randomized controlled trial of an online machine learning-driven risk assessment and intervention platform for increasing the use of crisis services. J Consult Clin Psychol 2019 Apr;87(4):370-379. [doi: 10.1037/ccp0000389] [Medline: 30883164]
106. Smith RL, Rose AJ. The 'cost of caring' in youths' friendships: considering associations among social perspective taking, co-rumination, and empathetic distress. Dev Psychol 2011 Nov;47(6):1792-1803 [FREE Full text] [doi: 10.1037/a0025309] [Medline: 21895362]
107. Stone LB, Hankin BL, Gibb BE, Abela JR. Co-rumination predicts the onset of depressive disorders during adolescence. J Abnorm Psychol 2011 Aug;120(3):752-757 [FREE Full text] [doi: 10.1037/a0023384] [Medline: 21500877]
108. Waller EM, Rose AJ. Adjustment trade-offs of co-rumination in mother-adolescent relationships. J Adolesc 2010 Jun;33(3):487-497 [FREE Full text] [doi: 10.1016/j.adolescence.2009.06.002] [Medline: 19616839]

Abbreviations

BMT: behavior modeling training
BSI: Brief Symptom Inventory
CI: credibility interval
CMH: Crowdsourcing Mental Health
CSRS: Crowdsourcing Mental Health Session Reaction Scale
PSS: Perceived Stress Scale
RQ: research question

Edited by G Eysenbach; submitted 27.11.19; peer-reviewed by T Muto, L Vizer; comments to author 16.01.20; revised version received 16.05.20; accepted 26.07.20; published 21.09.20.

Please cite as:
Bernecker SL, Williams JJ, Caporale-Berkowitz NA, Wasil AR, Constantino MJ
Nonprofessional Peer Support to Improve Mental Health: Randomized Trial of a Scalable Web-Based Peer Counseling Course
J Med Internet Res 2020;22(9):e17164
URL: http://www.jmir.org/2020/9/e17164/
doi:10.2196/17164
PMID:32955451

Original Paper

Influencing Factors of Continuous Use of Web-Based Diagnosis and Treatment by Patients With Diabetes: Model Development and Data Analysis

Chunhua Ju[1*], PhD; Shuangzhu Zhang[1*], MS

Business Administration College, Zhejiang Gongshang University, Zhejiang Province, China

[*]all authors contributed equally

Corresponding Author:
Shuangzhu Zhang, MS
Business Administration College
Zhejiang Gongshang University
18 Xuezheng Street, Hangzhou City, Zhejiang Province
Zhejiang Province, 310001
China
Phone: 86 15968138344
Email: zhangshuangzhu0917@126.com

Abstract

Background: The internet has become a major source of health care information for patients and has enabled them to obtain continuous diagnosis and treatment services. However, the quality of web-based health care information is mixed, which raises concerns about the credibility of physician advice obtained on the internet and markedly affects patients' choices and decision-making behavior with regard to web-based diagnosis and treatment. Therefore, it is important to identify the influencing factors of continuous use of web-based diagnosis and treatment from the perspective of trust.

Objective: The objective of our study was to investigate the influencing factors of patients' continuous use of web-based diagnosis and treatment based on the elaboration likelihood model and on trust theory in the face of a decline in physiological conditions and the lack of convenient long-term professional guidance.

Methods: Data on patients with diabetes in China who used an online health community twice or more from January 2018 to June 2019 were collected by developing a web crawler. A total of 2437 valid data records were obtained and then analyzed using correlation factor analysis and regression analysis to validate our research model and hypotheses.

Results: The timely response rate (under the central route), the reference group (under the peripheral route), and the number of thank-you letters and patients' ratings that measure physicians' electronic word of mouth are all positively related with the continuous use of web-based diagnosis and treatment by patients with diabetes. Moreover, the physician's professional title and hospital's ranking level had weak effects on the continuous use of web-based diagnosis and treatment by patients with diabetes, and the effect size of the physician's professional title was greater than that of the hospital's ranking level.

Conclusions: From the patient's perspective, among all indicators that measure physicians' service quality, the effect size of a timely response rate is much greater than those of effect satisfaction and attitude satisfaction; thus, the former plays an essential role in influencing the patients' behavior of continuous use of web-based diagnosis and treatment services. In addition, the effect size of electronic word of mouth was greater than that of the physician's offline reputation. Physicians who provide web-based services should seek clues to patients' needs and preferences for receiving health information during web-based physician-patient interactions and make full use of their professionalism and service reliability to communicate effectively with patients. Furthermore, the platform should improve its electronic word of mouth mechanism to realize its full potential in trust transmission and motivation, ultimately promoting the patient's information-sharing behavior and continuous use of web-based diagnosis and treatment.

(J Med Internet Res 2020;22(9):e18737) doi:10.2196/18737

KEYWORDS

online health communities; patient-doctor trust; ELM; trust theory; structural equation modeling; online continuous diagnosis and treatment

Introduction

Background

In recent years, health care in China has been confronting problems such as increasing numbers of patients with chronic diseases and intensified population aging. With the continuous increase of the number of "netizens" as well as the popularization of web-based health care services [1], the internet has become an essential channel for the dissemination of health care information, providing an opportunity to alleviate the abovementioned problems. Specifically, integration of the internet and health care services has established new communication channels for people to seek medical services on the internet, including health information searches, web-based inquiries, and web-based registration. By assembling high-quality medical resources from various places, the rich information flow allows precise matching between suitable medical service suppliers and patients in need, consequently aiding the balance of medical resource distribution to a certain extent [2]. Furthermore, through web-based inquiry, health counseling, health care activities, and health-themed education, these Web 2.0–based online health communities also provide information support and social support to patients [3,4], exerting positive impact on their self-management of their health and daily disease control through communication [5].

At present, prevention and treatment of diabetes is a major public health problem in China. Diabetes is a chronic and noncommunicable disease that is characterized by long duration, high disability rates, and a wide range of complications [6]. The treatment and control of diabetes is complicated, and long-term professional scientific management is necessary to achieve the goal of reducing acute and chronic complications [7]. Even in daily life, patients with diabetes may experience physical discomfort; obtaining health-related information through the internet can not only cater to patients' individual needs but can also help decrease costs and greatly increase the efficiency of informational retrieval [8]. By definition, online health communities are places where users can engage in activities such as knowledge sharing and member exchange with regard to health- or treatment-related issues [9]. Web-based inquiry is a vital driving force for the development of online health communities. Physician-patient inquiry is not only the main service provided by online health communities; it is also the most important web-based activity for patients. With guidance in treatment management, medication management, dieting management, etc., communication between patients, family members of patients, and attending physicians is becoming more efficient; this can help patients achieve self-management of their health. The literature on online health communities mainly focuses on knowledge sharing, web-based health information–searching behavior, physician-patient interactions, and web-based health management [10-13]. Few studies have investigated trust and continuous diagnosis and treatment in the context of web-based medical environments. Although there are exceptions, trust is often examined as one of many influencing factors; however, in-depth research of trust is lacking. For example, through questionnaire research, Deng et al [14] found that the credibility of websites, hospitals, and physicians, as well as perceived benefits and perceived risks, have significant effects on the trust of patients who use web-based services. Yi et al [15] designed experiments to examine health information searching behavior and found that the quality of evidence, expertise of the source, perceived information quality, and perceived risk significantly affected users' trust in network health information. Additionally, web-based inquiry breaks the patient's constraints in seeking medical care in the sense that that online health communities not only bring together physicians and patients from all over the country to realize mutual assistance regarding knowledge and emotion [16] but also help avoid the embarrassment of face-to-face communication about certain diseases [17].

Prior studies showed that 26% of adult internet users have browsed and viewed others' published experiences concerning health and medical care [18], and 16% of internet users are willing to find groups with the same health problems through the internet; among these, patients with chronic diabetes are even more inclined to search for disease information and interact with each other on the internet [16]. Therefore, from the perspective of the patient, by analyzing posting data on disease-related questions by patients with diabetes and the top-rated answers from physicians crawled from an online health community and surveyed with a questionnaire, in this paper, we develop a conceptual model to explore factors that influence the continuous use of web-based diagnosis and treatment by patients with diabetes based on the elaboration likelihood model (ELM) and trust theory. The patients' continuous use of web-based diagnosis and treatment in the present study refers to the patients' repeated behavior of using an online health community for inquiry, consultation, and help services. Overall, this study not only enriches the strand of empirical research on the web-based interactions, web-based inquiries, and health self-management of patients with diabetes but also has practical implications for the operation and development of online health communities.

Prior Literature

Literature on User Behavior in Online Health Communities

At present, the Chinese literature on user behavior in online health communities mainly comprises various types of behavior, such as information disclosure, information acquisition and searching, information sharing, information service usage and continuous usage, and social support behavior. With regard to the methods, questionnaire surveys and user interviews [19] are the most important approaches for studying user behavior in online health communities in China. However, considering that questionnaire data is subjective in nature and limited in quantity, objective data from the network of the online health community are used in research on user behaviors in these communities. For instance, the impact of user competition on health status was explored in the context of an online weight-loss community [20]. In another study, web-based data from the Good Doctor online health community was obtained using a web crawler, and multiple regression analysis was then conducted to examine the impact of physicians' and patients' behavior in online health communities on the knowledge exchange effect [21]. In addition,

techniques such as text mining and content analysis have been applied to the analysis of network data, thus expanding the research methods used to study user behavior in online health communities. For example, by applying text mining, users' question data can be captured based on the keyword *hypertension* and then analyzed; it was found that the informational needs of users in a hypertension health community were mainly concentrated on daily disease management, disease diagnosis and treatment, and the expectation of patients to receive emotional support from society [22]. Applying content analysis to subdivide group behavior into six types for a cancer-themed group on the QQ social media platform, the analysis of different types of behavior showed that the most important types of behavior were emotional support, knowledge sharing, and off-topic behavior [23].

In recent years, the application of social network analysis to research user behavior in online health communities has developed gradually. For instance, based on posting and replying data for half a year as well as on users' personal information in Tianmijiayuan, an online diabetes health community, Liu et al [24] used an exponential random graph model to explore how network structure and node attributions affect the establishment of users' reply networks. Zhai et al [16] conducted statistical analysis and social network analysis on user data from the Baidu Quitting Smoking Post Bar; they found that the user group was gradually decreasing in size and that the loss rate was also accelerating. In addition, social network analysis was used to study users' knowledge-sharing behavior [25] and information dissemination and interaction behavior [26] in online health communities.

Physician-Patient Interaction and Patients' Trust

Physicians are among the most important participants in online health communities. Due to the information asymmetry in online health communities, physicians' personal information, responses to patients' consultations, and electronic word-of-mouth reputation can effectively help patients distinguish between physicians at different professional levels to make efficient decisions. For example, some studies have found that a physician's electronic word-of-mouth reputation, efforts, and service price significantly influence the quantity of their medical inquiry, and the relationship between reputation and inquiry quantity is partially mediated by service price [27]. Moreover, for different diseases, the influencing factors of physicians' contribution behavior exert varying degrees of impact within different time lengths [28]. Concerning physician-patient interactions, existing research has mainly been conducted from the perspectives of knowledge exchange, physician-patient communication, and physician-patient trust. For example, according to knowledge exchange theory, the impact of behaviors of both physicians and patients (amount of knowledge exchanged, trust, cost, benefits, etc.) on the effectiveness of knowledge exchange was empirically verified [29]. The interaction between physicians and patients was studied from the aspects of the physicians' degree of activity, patients' visits, and patients' satisfaction [22]. By integrating trust factors with perceived benefit and perceived risk, the influencing factors of physician-patient trust in online health services were also explored based on the conceptual framework of web-based trust [14].

Trust is considered to be a basic factor in the formation of successful relationships [30]. In particular, recent studies have focused on the relationship between trust toward providers of products and services and customers' intentions to make web-based purchases [31]. Among various contexts, patients' trust can be defined as the patients' belief and expectation that a medical service provider will take actions that are beneficial to them when they lack the capability to supervise physicians [32]. Previous studies found that trust toward members impacts web-based participation behavior, such as seeking and providing information in focus groups [33]. Hospital rules and regulations as well as the physician's professional skills and service attitudes will have an impact on patient trust. These factors usually exert roles in the context of medical institutions, medical staff, and medical treatment situations [34]. At the same time, patients' trust will affect their own health [35]. In particular, information obtained from credible sources is often considered to be more useful and is treated as the basis for decision-making [36].

To conclude, it can be found that literature studies on users in online health communities mainly focus on users' relationship networks, users' behavior regarding health information, physician-patient interaction, and patients' trust. Additionally, some studies have addressed the subjects of methods for calculating similarity among virtual health community users [37], emotional expression of users in online health communities [38], a member's value co-creation model and its influencing factors [39], and the algorithms of sentiment analysis for user reviews on the internet [40].

Theoretical Basis and Hypotheses

Theory of the ELM

The ELM is a social psychology model. It is a theoretical model that was proposed by Petty and Cacioppo to explain users' attitudes towards persuasive information changes [41]. It is believed that there are two development routes for changes in personal attitudes: the central route and the peripheral route [41]. The difference between the two routes is mainly reflected in the type of energy input or by information processing differences regarding verbosity. The attitude of the central path mainly comes from a careful evaluation of the information available and the possible benefits of adopting this attitude. This path requires individuals to think critically about the arguments contained in a message and to examine the arguments' relative advantages and relevance. Then, the judgment of the target behavior is formed. Conversely, in the peripheral route, users mainly rely on tips about the target behavior to make a judgment, such as the number of existing users and information technology experts' approval.

As a model of persuasion, the ELM is widely used in social psychology, management, marketing, and other fields. Bhattacherjee et al [42] explored how information processing processes affect users' information technology based on detailed likelihood models and examined how long these effects last. Research finds that quality and information credibility affect user intentions. Filieri and McLeay [43] used a detailed

likelihood model to explore information quality, information accuracy, value-added information, and information relevance as influencing factors of the central path related to the timeliness of information and the marginal path related to the product level on the acceptance of information related to accommodation and tourism-related products by tourists. Since the ELM was proposed, many studies have used this model to analyze and understand the process and mechanism underlying users' information processing in various contexts, such as the user's knowledge adoption intention [44], the user's attitude and intention to technology acceptance [45], the consumer's initial letter for mobile banking [46], and the user's intention to use the information system [47].

Influencing Factors of Patients' Continuous Diagnosis-Treatment Behavior Under the Central Route

According to the theory of the ELM, when individuals have sufficient motivation and capability, they will think carefully about and make judgments on the quality of argument information; finally, they will form their attitude and behavior towards the information accordingly [48]. For a web-based inquiry service, the homepage of the registered physicians can display evaluations of the service quality by treated patients. These evaluations mainly include patients' rating of and satisfaction with the physician's web-based service attitude and service outcomes, respectively reflecting their levels of professionalism and medical technology. As a special commodity service, through web-based medical inquiry, patients often expect physicians to respond quickly to help them solve their personal problems in a timely manner. Usually, if the patients' satisfaction with the web-based inquiry service and the quality of a physician's web-based service are high, the risk perceived by the patients will be lower, and they will expect to obtain more professional and effective medical advice and better service. Research shows that informational interaction between consumers and sellers can promote consumers' trust in sellers and their behavioral intentions [49]. Therefore, a timely response rate, as a characteristic of web-based physician-patient interactions, may positively affect a patient's online treatment behavior. Therefore, based on the above analyses, the following hypotheses are proposed:

H1: Satisfaction with service attitude has a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

H2: Satisfaction with service outcome has a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

H3: Physicians' timely response rate has a positive impact on patients' behavior toward continuous use of web-based diagnosis and treatment.

H4: The effectiveness of physicians' advice has a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

Influencing Factors of Patients' Continuous Use of Web-Based Diagnosis and Treatment Under the Peripheral Route

Trust propensity varies depending upon an individual's personality, resulting in different degrees of trust in an environment [50]. Therefore, internet users' trust in a website is influenced by their personal trust propensity traits. Even for users in the same network, their different trust propensities can lead to different degrees of trust [51]. Studies have found that trust propensity will affect a user's acceptance of new things, their attitudes towards network communities, and the the degrees of interaction among community members in unfamiliar environments [52]. The higher the individual's trust propensity, the higher their degree of acceptance of new things or information and the greater their tendency to exchange, communicate, and interact with other community members. A reference group refers to individuals or groups that are closely related to an individual's evaluation, pursuits, or behavior [50]. These groups have an impact on users' behaviors, lifestyle, attitudes, etc. [53] When users face complex products that lack relevant information, they are more inclined to obtain information from reference groups. Hence, based on the above illustration, an individual's trust propensity and reference group can be classified as influencing factors under the peripheral route, and the following hypotheses are proposed:

H5: An individual patient's trust propensity has a positive effect on the patient's behavior toward continuous use of web-based diagnosis and treatment.

H6: The credibility of the reference group has a positive effect on patients' continuous behavior toward web-based diagnosis and treatment.

Trust Theory

Trust theory was originally proposed by Luhmann [54] to verify changes regarding users' long-term relationships. This theory has since been applied widely in various disciplines. Trust theory states that users have confidence in network providers, and their willingness to rely on them increases during the process of opinion adoption and purchase payment in the context of shopping on the internet. Existing literature has mainly examined factors that affect web-based consumption, such as product reputation, word of mouth, brand, reliability, etc., and we found that consumer trust is related to factors such as the seller's reputation, word of mouth, and consumer interest–related guarantee mechanisms [55]. Furthermore, product safety and reliability can influence a customer's purchase intention by affecting customer trust [56].

The Impact of a Physician's Offline Reputation on Patients' Continuous Use of Web-Based Diagnosis and Treatment

All registered physicians in online health communities are required to provide real-name authentication and must display rank information in terms of their professional skill (eg, resident or attending physician) and the hospitals where they work (eg, tier 1C). Generally, the hospital's ranking represents the medical level of the hospital as a whole. The rank of the physician represents their own professionalism. A higher professional title indicates that the physician has rich clinical experience and a

higher degree of professionalism. Therefore, in web-based medical inquiry services, patients believe that a physician with a higher rank who is affiliated with a higher-ranked hospital will have better professional performance and higher authority; therefore, they are more likely to choose that physician. As a matter of fact, a physician's title is indeed a good indicator of their performance and knowledge. This is usually qualified based upon comprehensive evaluations of a physician's academic background, work experience, and certificate of their professional qualification examination [57]. As a result, a physician's title can improve patients' recognition and trust of that physician and can thus promote patients' behavioral intentions. Based on the above illustration, the following hypotheses are proposed:

H7: The physician's professional title has a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

H8: The hospital's ranking has a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

The Impact of a Physician's Electronic Word of Mouth on Patients' Continuous Use of Web-Based Diagnosis and Treatment

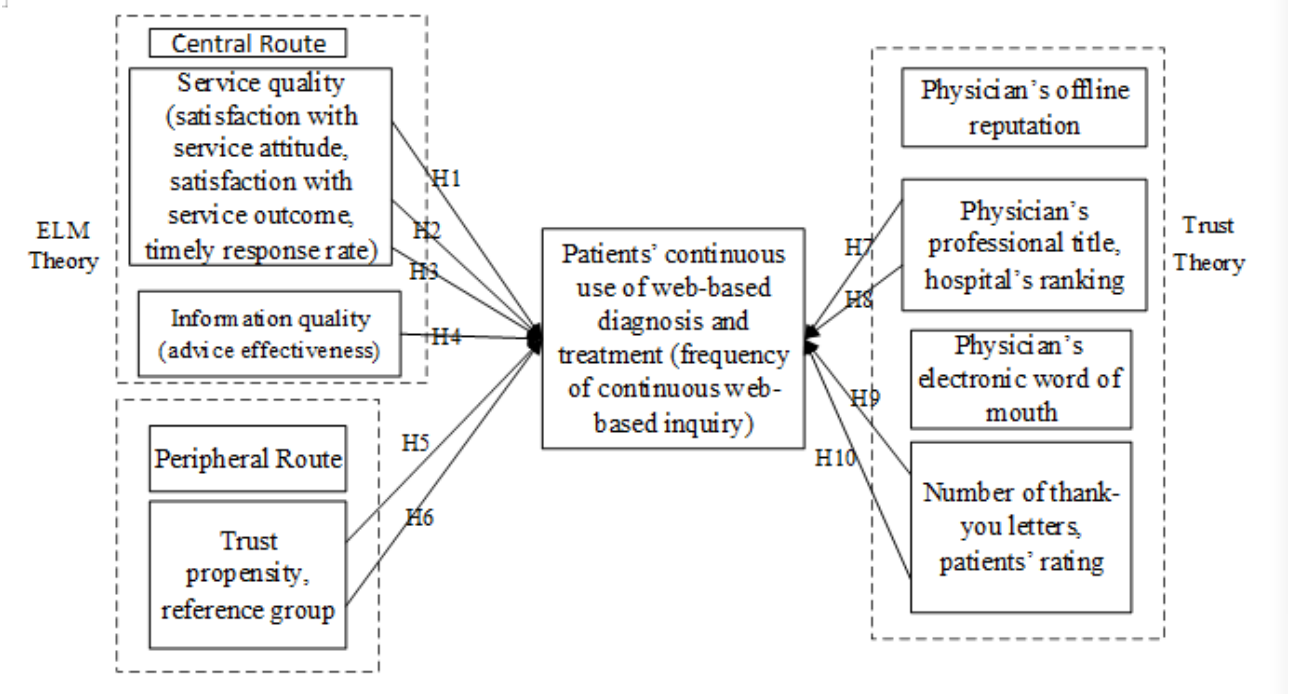
An online health community platform can capture various forms of feedback, such as the number of thank-you letters sent by patients and the patients' rating of the platform. Specifically, after receiving web-based inquiry services from physicians, patients usually give positive feedback such as thank-you letters and satisfaction ratings to express gratitude or recognize the physician's professionalism. This type of information is publicly shared on the internet and can improve other patients' understanding towards the physician, reducing the negative impact of asymmetric information between physicians and patients and helping patients with decision-making. Studies have found that on social media websites, electronic word of mouth can affect users' purchase intentions through trust transfer and then affect their purchase decisions [58,59]. Based on the above illustration, the following hypotheses are proposed:

H9: The number of thank-you letters has a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

H10: Ratings by other patients have a positive effect on patients' behavior toward continuous use of web-based diagnosis and treatment.

In summary, public information in a web-based medical community is the main basis for patients' perception and trust-building towards web-based physicians and determines whether they will choose their services. Therefore, in this study, after controlling for the effect of the platform on the choice of physician by patients with diabetes, we established a conceptual model (Figure 1) in which a patient's behavior toward continuous use of web-based diagnosis and treatment is conjointly influenced by factors through both the central route and peripheral route as well as by factors such as the physician's electronic word of mouth and offline reputation. Furthermore, patients with diabetes are considered as a whole in this study, and individual characteristics of the patients are not taken into consideration in this model.

Figure 1. Influencing factor model of behavior toward continuous web-based diagnosis and treatment for patients with diabetes. ELM: elaboration likelihood model.



Methods

Data Collection

In this study, the data were obtained from an online medical community, WeiYi [60], which archives information about patients' activities related to web-based diagnosis and treatment. Because patients form their perceptions toward physicians by browsing information on the website and then select a physician for medical inquiry, a web crawler was developed to collect the data. As noted, the focus of the present study is to analyze the influencing factors of the continuous use of web-based diagnosis and treatment by patients with diabetes. The crawler program was written in Python language, and data spanning January 2018 to June 2019 that documented patients with diabetes who had ≥2 medical inquiry records were collected. The data include basic information about physicians and hospitals, the patients' web-based inquiry procedures, and the patients' feedback. Furthermore, the patients' age and sex information were provided by WeiYi authorities. After matching the patients' ID from the two sources, 4100 pieces of raw data were obtained after excluding missing values. Then, the data information that measure the variables in the conceptual model were identified in a stepwise process to conduct the empirical analysis.

The independent variables were selected based on the conceptual model. The corresponding empirical indicators were selected according to the data provided by the website. The physicians' offline information (ie, professional title and hospital ranking level) represented the variable of the physicians' reputation. The indicators for the patients' central route processing were service quality and information quality, of which service quality was measured by the patient's satisfaction with the service attitude and service outcome as well as the timely response rate, while information quality was measured by the effectiveness of the physician's advice. The indicators for the patient's peripheral route processing were trust propensity and the reference group (ie, feasibility of information provided by the group). The indicators for a physician's electronic word-of-mouth reputation were the patients' rating and the number of thank-you letters. The index of service price was obtained directly based on the actual paid price.

The dependent variable is the goal of the research model. As mentioned, the patients' continuous use of web-based diagnosis and treatment in the present study refers to the patients' repeated behavior of using the online health community for inquiry, consultation, and help services. Therefore, patients with ≥2 inquiries were selected, and the total number of these patients' inquiries was measured as the dependent variable to explore its influencing factors.

Data Analysis

Data Preprocessing

Data that met descriptive statistical requirements were obtained through text mining and data cleaning, and data required for correlation analysis and regression analysis were obtained through data transformation. The data preprocessing steps included valid field data matching, character field assignment conversion, null value processing, descriptive statistical analysis, data transformation (normalization and standardization), and outlier rejection.

Data Cleaning

According to the variables operationalized in the definition, irrelevant original data were deleted. Then, data collected through the web crawler were matched and integrated with the patients' ID information. The rules of the character field assignment conversion were as follows. Physicians' professional titles were ranked as resident physician, attending physician, associate chief physician, and chief physician, corresponding to 1, 2, 3, and 4, respectively. Hospitals were ranked with 9 tiers, and the ranks of the hospitals were classified by searching their full names, in which tier 1C, tier 1B, tier 1A, tier 2C, tier 2B, tier 2A, tier 3C, tier 3B, and tier 3A corresponded to 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. The rules of null value processing were as follows. The null value was eliminated if one piece of information regarding the physician's professional level, hospital ranking level, patient rating score, or price was missing, as this would render the entire piece of information a meaningless outlier. However, if no thank-you letter was found, this was actually of practical significance and was assigned a value of 0, indicating that the physician had not received this type of feedback. In the text of the patient's review, their opinion of the physician's professionalism can be described in three ways: extremely professional, very professional, and professional, which were assigned values of 1, 2, and 3. Regarding the trust propensity, the credibility of information provided by patients (the reference group) and the effectiveness of physicians' advice, a value of 0 or 1 was given, for which 0 indicated low credibility and less effectiveness. After cleaning, 3200 pieces of data were obtained. A sample of the data is shown in Table 1.

Table 1. Sample of the data obtained for the study.

Physician's professional title[a]	Hospital's ranking level[b]	Thank-you letters, n	Patient's rating[c]	Satisfaction with service attitude, %	Satisfaction with service outcome, %	Timely response rate, %	Trust propensity[d]	Credibility of reference group[d]	Advice effectiveness[d]	Internet inquiries, n
4	9	20	5	98.8	97.6	67	1	1	1	2
4	9	9	5	100.0	100.0	55	0	1	1	2
3	9	8	5	100.0	78.0	91	1	1	1	2
3	9	0	5	89.0	100.0	100	0	1	1	2
3	9	10	5	100.0	98.6	69	1	0	1	2
2	9	90	4	98.6	100.0	64	1	1	1	3
3	9	40	5	100.0	99.0	100	1	1	0	3
4	9	65	5	97.9	98.8	100	1	1	1	2
4	8	38	5	98.8	99.0	63	1	0	1	4
3	9	40	3	98.0	96.0	66	1	0	0	2
2	9	60	5	99.0	98.8	66	1	0	1	5
2	9	46	4	99.0	98.8	72	0	1	1	2
2	9	97	5	99.0	99.0	78	1	1	1	3

[a]1: resident physician; 2: attending physician; 3: associate chief physician; 4: chief physician.

[b]1: tier 1C; 2: tier 1B; 3: tier 1A; 4: tier 2C; 5: tier 2B; 6: tier 2A; 7: tier 3C; 8: tier 3B; 9: tier 3A.

[c]Patients' rating: score of 1 to 5, where 1=poor and 5=excellent.

[d]Value of 0 or 1, where 0 is invalid and 1 is valid.

Data Transformation

Based on the descriptive statistical results, it was found that the magnitudes of the different variables varied greatly and that the variables showed obviously skewed distribution, excluding the physician's professional title, hospital's ranking level, patients' rating, effectiveness of advice, number of inquiries, and trust propensity and credibility of reference group information. Using square transformation and logarithmic transformation, all data were scaled to the same magnitude, and the skewness was corrected to a certain degree. Within the data boundary, the intervals of different variable indicators intersected. The lower bound of each CI is the difference between the average value and the sampling error, and the upper bound is the sum of the average value and the sampling error. Additionally, outliers that significantly exceeded the range of the variable were removed to reduce noise. Moreover, data that did not fall within the confidence interval were deleted. Finally, 2437 records were obtained for the correlation analysis and regression analysis.

Results

Demographic Characteristics

Among the 2437 patients surveyed, 1617 were female (66.4%). This may be because women are often required to care of family health and other responsibilities in addition to work; also, women tend to pay more attention to health information than men [61]. A total of 2060/2437 patients (52.3%) were 20 to 40 years of age. This may be due to specific limitations of the internet use of patients with diabetes. Specifically, diabetes is a chronic noncommunicable disease, and older people may not frequently consult physicians on the internet or ask their family members to perform online inquiries because of their rich personal experiences in consulting physicians and obtaining medicines. In the 2437 records, 1048 of the physicians (43.0%) are chief physicians or associate chief physicians, while 2368 hospitals (97.16%) were ranked 3A. All interviewees had two or more web-based interactions with physicians (see Table 2).

Table 2. Summary of the characteristics of the collected data records (N=2437), n (%).

Characteristic	Value
Gender	
Male	820 (33.7)
Female	1617 (66.4)
Age (years)	
20-30	786 (32.3)
31-40	1274 (52.3)
41-50	309 (12.7)
>55	68 (2.8)
Physician's professional title	
Resident physician	465 (19.1)
Attending physician	891 (36.6)
Associate chief physician	848 (34.8)
Chief physician	200 (8.2)
Other	30 (1.2)
Hospital's ranking level	
3A	2368 (97.2)
Other	69 (2.8)
Number of web-based consultations by the patient per year	
2	1680 (27.3)
3	376 (35.1)
4	205 (14.2)
>4	176 (4.5)

Spearman Correlation Analysis

Correlation analysis was conducted on all independent variables to effectively avoid the problem of multiple collinearity of independent variables during multiple regression analysis. Spearman correlation analysis was performed on the final data after cleaning because this method is more suitable for data sets with discontinuous variables. The analysis results showed three pairs of strong correlations. First, a strong correlation was observed between the patients' rating and the number of thank-you letters, with a correlation coefficient of 0.968. Second, a strong correlation was found between satisfaction with service outcomes and satisfaction with service attitude, with a correlation coefficient of 1; this is consistent with existing research [62]. Last, trust propensity and effectiveness of advice were strongly correlated, with a correlation coefficient of 1. The correlations of all other independent variables were within normal ranges (see Table 3).

Table 3. Spearman correlation analysis of all variables (significance level=.05), r.

Variable	Physician title	Hospital ranking level	Timely response rate	Satisfaction with service outcome	Satisfaction with service attitude	Thank-you letters	Patients' rating	Trust propensity	Reference group	Advice effectiveness
Physician title (*P*<.001)	1	0.040	-0.006	0.012	0.012	0.005	0.004	0.065	0.004	0.016
Hospital ranking level (*P*<.001)	0.040	1	-0.036	0.018	0.018	0.093	0.012	0.063	0.009	0.020
Timely response rate (*P*<.001)	–0.006	–0.036	1	0.089	0.089	0.087	0.056	0.593	0.026	0.023
Satisfaction with service outcome (*P*=.78)	0.012	0.018	0.089	1	1.000	0.548	0.387	0.048	0.017	0.032
Satisfaction with service attitude (*P*=.14)	0.012	0.018	0.089	0.089	1	0.548	0.387	0.163	0.018	0.032
Thank-you letters (*P*=.36)	0.005	0.093	0.087	0.548	0.548	1	0.968	0.029	0.032	0.014
Patients' rating (*P*=.74)	0.004	0.012	0.056	0.387	0.387	0.968	1	0.065	0.046	0.163
Trust propensity (*P*=.48)	0.065	0.063	0.593	0.048	0.163	0.029	0.065	1	0.034	1.000
Reference group (*P*=.04)	.004	0.009	0.026	0.017	0.018	0.032	0.046	0.034	1	0.008
Advice effectiveness (*P*=.74)	0.016	0.020	0.023	0.032	0.032	0.014	0.163	1.000	0.008	1

Regression Analysis

Linear regression analysis was performed on all variables, and the results showed that the R^2 value of the model is 0.916. However, when the significance level is .05, the *F* statistics of number of thank-you letters, satisfaction with service outcome, satisfaction with service attitude, patients' rating, trust propensity, and advice effectiveness are not significant, which indicates that serious multicollinearity is likely to occur. To eliminate collinearity, two groups of control experiments for the three sets of strongly correlated indicators in the correlation analysis results were established. In group A, the indicators of patients' rating, satisfaction with service attitude, and number of thank-you letters were retained. In group B, the indicators of trust propensity, satisfaction with service outcome, and advice effectiveness were retained. SPSS software (IBM Corp) was used to perform regression analysis on the normalized data set. The results are shown in Table 4.

The experimental results of the two groups were basically the same, indicating that the models were reliable. The R^2 values of both models were greater than 0.68, indicating that the independent variables can account for more than 68% of the change in the dependent variable. Therefore, the regression model had a good fit. The *F* statistic (α=.000, *P*<.005) showed that at the significance level of .05, the linear relationship of the regression model was significant; therefore, the regression model was explanatory for the hypotheses. Based on the model results, the final multiple regression model is as follows:

CQ_A = 0.004 * physician's professional title + 0.018 * hospital's ranking level + 0.096 * number of thank-you letters + 0.628 * timely response rate + 0.428 * patients' rating + 0.513 * reference group – 0.798 * satisfaction with service attitude

CQ_B = 0.006 * physician's professional title + 0.019 * hospital's ranking level + 0.693 * timely response rate + 0.518 * reference group + 0.002 * advice effectiveness – 0.912 * satisfaction with service outcome – 0.013 * trust propensity

Table 4. Results of the regression analyses.

Variable	Group A (R^2=0.689; significance level=.000)		Group B (R^2=0.692; significance level=.000)		All variables (R^2=0.916; significance level=.000)	
	Unstandardized coefficient	Significance level	Unstandardized coefficient	Significance level	Unstandardized coefficient	Significance level
Physician's professional title	0.004	0.000	0.006	0.000	0.472	0.000
Hospital's ranking level	0.018	0.000	0.019	0.000	0.220	0.000
Number of thank-you letters	0.096	0.000	N/A[a]	N/A	0.001	0.363
Timely response rate	0.628	0.000	0.693	0.000	0.144	0.000
Satisfaction with service outcome	N/A	N/A	-0.912	0.000	0.004	0.783
Satisfaction with service attitude	–0.798	0.000	N/A	N/A	–0.384	0.143
Patients' rating	0.428	0.000	N/A	N/A	0.000	0.736
Trust propensity	N/A	N/A	–0.013	0.000	–0.043	0.482
Reference group	0.513	0.000	0.518	0.000	–0.019	0.040
Advice effectiveness	N/A	N/A	0.002	0.000	–0.028	0.744

[a]N/A: not applicable.

Discussion

Main Findings

In this section, we discuss the hypotheses regarding the influencing factors of patients' continuous use of web-based diagnosis and treatment based on the above empirical results. Generally speaking, according to the analytical results, the timely response rate, reference group, and patients' rating are the three indicators with the largest positive regression coefficients, indicating that the timeliness of a physician's service has a great effect on a patient's choice and that other patients' rating scores and information reliability also greatly affect a patient's continuous use of web-based diagnosis and treatment. Furthermore, the coefficient of the hospital ranking level is very small; this may be related to the fact that patients generally choose physicians in hospitals that rank at tier 3a for web-based inquiry.

Hypotheses H1 through H6 can be discussed from the perspective of ELM theory. The physician's service quality (operationalized as satisfaction with service outcome, satisfaction with service attitude, and timely response rate) is an important factor that affects patients' continuous use of web-based diagnosis and treatment through the central route. The data show that attitude satisfaction and effect satisfaction have negative effects on patients' continuous use of web-based diagnosis and treatment, and the regression coefficients are large. The data also show that timely response rate has a significant positive effect on patients' continuous use of web-based diagnosis and treatment, while advice effectiveness presents a small effect. Therefore, the analytical results deny H1 and H2 and support H3 and H4. These findings can be explained in two ways. On the one hand, patients do not select physicians with high satisfaction ratings incautiously. On the other hand, in contrast with other diseases, diabetes is a disease that requires chronic long-term treatment; therefore, patients with diabetes may be more concerned about the quality of service from the perspective of response timeliness. For H5 and H6, the data show that the reference group and trust propensity have significant effects on patients' continuous use of web-based diagnosis and treatment through the peripheral route; the effect size of the reference group is greater than that of trust propensity, providing supportive evidence for both hypotheses. These results reveal that when patients lack motivation and capability to judge the information quality provided by physicians in online health communities, or when a clear view is not yet formed, a credible reference group is a very convincing indicator of the patients' decision. This is consistent with the conclusions in existing literature studies of the effects of reference groups on consumer behavior, lifestyle, self-concept development, and attitude [63].

From the perspective of trust theory, H7 through H10 can be discussed. The statistics show that a physician's professional title and hospital ranking level are positively related with patients' continuous use of web-based diagnosis and treatment, supporting H7 and H8. The significance of the relationship suggests that trust in a physician's offline reputation will influence trust in their web-based services. In other words, trust in offline medical services can be directly transferred to web-based medical services. However, the low coefficients further indicate that patients are less susceptible to a physician's offline characteristics after receiving multiparty information [62]. The statistics also show that the number of thank-you letters and the patients' rating are positively related with the number of uses and have a greater impact on patients' continuous use of web-based diagnosis and treatment, supporting H9 and H10. The significance of relationships suggests that patients care more about physicians' electronic world of mouth

because they believe that it is an indicator that condenses other patients' internet experiences. The finding that a physician's electronic word-of-mouth reputation can help patients make decisions is consistent with the research findings of Cao et al [64]. At the same time, this research also reflects that trust has an overall positive effect on the continuous use of web-based diagnosis and treatment by patients with diabetes, which demonstrates that patients' trust in web-based inquiry services is highly significant in predicting whether they will choose web-based consultation services.

Theoretical Contributions and Practical Implications

The present study makes several theoretical contributions.

First, few previous studies have comprehensively applied ELM theory and trust theory together to examine web-based inquiry services; also, these studies seldom take price and credibility of the trust source into consideration for analysis. By filling this research gap, our research contributes to new ideas regarding patients' behavior related to web-based diagnosis and treatment. We found that a physician's offline reputation has a relatively weak influence on patients' decision-making with regard to continuous use of web-based diagnosis and treatment. On the one hand, when using the web-based inquiry service, among the physicians and hospitals selected by patients, 55.4% of physicians are chief physicians or associate chief physicians and 98% of hospitals are in the top tier (3a). It can be concluded that the level of professional skill does not significantly influence patients' decision-making because this type of information is usually available to patients on the internet without distinction. On the other hand, unlike offline interactions with physicians, patients mainly make web-based inquiries for the purpose of obtaining disease-related information. After receiving information from multiple sources, they can make a wise decision as to which physician to choose. Furthermore, the acquisition of information is not restricted by region. Thus, the role of the physician's offline reputation is no longer significant.

Second, the core factor that influences patients' continuous use of web-based diagnosis and treatment is the physician's electronic word-of-mouth reputation. Patients can preliminarily perceive physicians' service attitudes and professional skill through the timeliness of their responses and their professional titles. Then, combined with other patients' ratings of physicians' receptive attitude and service efficacy, patients will establish their own trust toward the physicians, which will aid their subsequent decision-making. This finding suggests that patients place much more value on the timeliness of web-based inquiry. Previous studies have shown that the impact of positive evaluation on purchase intention is less than that of negative evaluation on customers' refusal to purchase [65], and both the quality and number of evaluations affect users' purchase intentions [66]. Our data illustrate that the average score of satisfaction with the physicians is quite high (98%), and only a few records show satisfaction below 60%. Taken together, these analyses reflect that the satisfaction scoring mechanism of the platform is not well designed, leading to a decrease in the reliability of the satisfaction score and likely consequently exerting a negative impact on patients' decision-making.

Building upon these findings, several practical implications can be summarized to improve actual service. First, in terms of physicians who are registered on the internet, to improve patients' rate of use of their services, it is key for the physicians to ensure the quality of their responses to patients' web-based inquiries and consultations as well as to increase their activity on the platform. Further, physicians are encouraged to provide complete supplementary information regarding their professional experiences. Second, in terms of web-based platform building, the feedback mechanism should be optimized. This is because patients pay more attention to the number of thank-you letters and ratings by other patients when making decisions. At the same time, the indicators that measure physician-patient interactions, such as timeliness of response, should be much more scientifically refined. The mechanism related to electronic word-of-mouth should be improved to reflect its role in trust transmission and encouragement. For instance, the control of satisfaction scoring should be strengthened to reduce or avoid adverse network behaviors such as click farming, ensuring the quality of web-based inquiry information. Third, in terms of the relationship between platforms, physicians, and patients, compared with the ranking of the hospital, the physician's professional title has a greater impact on patients' decision-making. This implies that the platform should strengthen and broaden its cooperation with patient-trusted physicians who offer high-quality services and that specific needs for behavior related to web-based diagnosis and treatment and the underlying psychological mechanisms for patients with different diseases should be further carefully considered.

Limitations

The present research is not without limitations. First, the data were obtained from a single web-based medical community, WeiYi, which decreases the generalizability of the findings. In future research, collection of data from multiple online health community platforms should be considered to verify the results. Second, this study is focused on factors that influence the behaviors of Chinese patients with diabetes regarding web-based diagnosis and treatment. In the future, similar studies should be carried out in western regions. Third, regarding the research methods, future studies should take various regression methods into consideration and conduct comparative analyses. For instance, in addition to the regression methods used in this study, baseline regression analysis containing all variables can also be applied, and follow-up studies on factors that moderate patients' behavior regarding web-based services can be conducted by introducing patients' experiences with web-based inquiries, involvement in social networks, etc.

Conclusions

In this study, we collaborated with the operators of an online health community by crawling backend log data and obtaining patients' basic information (age and gender) from the platform; the results of this study help provide a reference for how to optimize web-based inquiry services and enhance patients' continuous behavior related to diagnosis and treatment. As a necessary means to alleviate the problem of restricted offline medical resources, web-based inquiry services, which are not constrained by time and space, have improved the satisfaction

of patients with diabetes regarding their self-management of their health. However, the actual adoption rate of web-based inquiry services by patients with diabetes remains low. Applying ELM theory and trust theory, the results of our study show that the timeliness of physicians' responses to web-based inquiries is positively related to patients' continuous use of web-based diagnosis and treatment, leading to high intention of patients to continue to use web-based inquiry services; the reference group is conducive to patients' trust towards and continuous use of web-based inquiry services; and the effect sizes of hospital ranking levels and physicians' professional titles are small. Overall, the conceptual model showcases good explanatory power for predicting continuous use of web-based diagnosis and treatment by patients with diabetes, expanding the literature on mobile medical services. The findings provide important implications regarding ways to promote healthy development of online health communities and to help solve the problems of resource shortages and information asymmetry in the medical industry.

Acknowledgments

This project was funded by grants from the National Natural Science Foundation of China: Research on Consumer Credit Value Measurement Integrating Online Social Relationships in eCommerce (71571162). The data were collected with help from the administrator of the WeiYi platform.

Authors' Contributions

SZ and CJ refined the topics and methods at the initial stage of paper writing. Then, SZ conducted the statistical analysis and wrote the paper under the guidance of CJ. Both authors reviewed, revised, and approved the final draft.

Conflicts of Interest

None declared.

References

1. Meng Q. Research on the Application and Development of Internet and Medical Health. Beijing, China: Beijing Health Press; 2015:7.
2. Kim H, Mrotek A. A functional and structural diagnosis of online health communities sustainability: A focus on resource richness and site design features. Comput Hum Behav 2016 Oct;63:362-372. [doi: 10.1016/j.chb.2016.05.004]
3. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. Patient Educ Couns 2006 Aug;62(2):178-188. [doi: 10.1016/j.pec.2005.10.003] [Medline: 16406472]
4. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
5. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
6. Jiang Y, Wu TF, Tang X. Preliminary study on community standardized management model for diabetic patients. Article in Chinese. Cardiovascular and Cerebrovascular Disease Prevention and Treatment 2012;12(4):321-323 [FREE Full text]
7. Wang J, Wu H. Interpretation of the core contents of the international guideline for the diagnosis and treatment of diabetes mellitus in traditional Chinese Medicine. Article in Chinese. Beijing Journal of Traditional Chinese Medicine 2018;37(8):730-732. [doi: 10.16025/j.1674-1307.2018.08.010]
8. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decis Support Syst 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
9. Zhang X, Xia H, Chen X, Hou D. A Study of Factors Affecting the Information Credibility of Online Health Communities from an Elaboration Likelihood Model Perspective. Article in Chinese. Library and Information Service 2015;59(22):104-196. [doi: 10.13266/j.issn.0252-3116.2015.22.014]
10. Zhao J, Wang T, Fan X. Patient value co-creation in online health communities. J Serv Manag 2015 Mar 16;26(1):72-96. [doi: 10.1108/josm-12-2013-0344]
11. Lin CA, Atkin DJ, Cappotto C, Davis C, Dean J, Eisenbaum J, et al. Ethnicity, digital divides and uses of the Internet for health information. Comput Hum Behav 2015 Oct;51:216-223. [doi: 10.1016/j.chb.2015.04.054]
12. MA Cheng-yu. Online health community Medical treatment model Doctor-patient interaction behavior. Article in Chinese. Chinese Journal of Health Policy 2016;9(11):65-69. [doi: 10.3969/j.issn.1674-2982.2016.11.011]
13. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov 28;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
14. Deng ZH, Hong Z. An Empirical Study of Patient-physician Trust Impact Factors in Online Healthcare Services. Article in Chinese. J Manage Sci 2017;30(1):43-52. [doi: 10.3969/j.issn.1672-0334.2017.01.004]

15. Yi MY, Yoon JJ, Davis JM, Lee T. Untangling the antecedents of initial trust in Web-based health information: The roles of argument quality, source expertise, and user perceptions of information quality and risk. Decis Support Syst 2013 Apr;55(1):284-295. [doi: 10.1016/j.dss.2013.01.029]
16. Liu X, Wang L, Li J, Zhang P. Research on Mechanisms of User Replying Behaviors in Online Health Communities. Article in Chinese. J Manage Sci 2017;1:62-72. [doi: 10.3969/j.issn.1672-0334.2017.01.006]
17. Asiri E, Khalifa M, Shabir S, Hossain MN, Iqbal U, Househ M. Sharing sensitive health information through social media in the Arab world. Int J Qual Health Care 2017 Feb 01;29(1):68-74. [doi: 10.1093/intqhc/mzw137] [Medline: 28003369]
18. Fox S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2020-09-18]
19. Wang W, Xie Y, Liu K. Research on Virtual Health Community Users Willingness Based on Grounded Theory. Article in Chinese. Information and Documentation Services 2017;38(3):75-82. [doi: 10.3969/j.issn.1002-0314.2017.03.011]
20. Song X. Research on Social Relationship and Competition Behavior of Patient in Online Health. Dissertation in Chinese. Harbin Institute of Technology. 2015. URL: http://d.wanfangdata.com.cn/thesis/D751783 [accessed 2020-09-18]
21. Fan X, Ai S. An Empirical Study on the Relationship between Online Medical Community Participants' Behaviors and Knowledge Exchange Effect. Article in Chinese. Information Magazine 2016;35(7):173-178. [doi: 10.3969/j.issn.1002-1965.2016.07.029]
22. Deng S, Liu J. Journal of Information Resources Management 2016;6(3):25-33. [doi: 10.13365/j.jirm.2016.03.025]
23. Zhou JJ. Examining the Users' Participation Behavior Types in Online Health Communities: A Qualitative Approach. Article in Chinese. Journal of Management Case Studies 2016;9(2):A [FREE Full text]
24. Liu X, Wang L, Li J, Zhang P. Research on Mechanisms of User Replying Behaviors in Online Health Communities. Article in Chinese. J Manage Sci 2017;30(1):62-72. [doi: 10.3969/j.issn.1672-0334.2017.01.006]
25. Wu J, Zhou LS. The Study of Knowledge Sharing Network and Users' Knowledge Interaction in Online Health Community. Article in Chinese. Information Science 2017;35(3):144-151.
26. Wu J, Shi L. Study of the User Interaction Behavior in Online Health Community Based on Social Network Analysis. Article in Chinese. Information Science 2017;35(7):120-125 [FREE Full text]
27. Liu XX. The Impact of Online Doctor Reputation and Doctor Effort on Consultation Amount. Dissertation in Chinese. Harbin Institute of Technology. 2014. URL: http://d.wanfangdata.com.cn/thesis/D592727 [accessed 2015-08-17]
28. Li Y. Study on the Factors of Doctors' Contribution Behavior of Online Medical Community. Dissertation in Chinese. Harbin Institute of Technology. 2015. URL: http://d.wanfangdata.com.cn/thesis/D754171 [accessed 2016-05-05]
29. MA CY. An empirical study on the interactive behavior of physicians and patients in online health community: A case study of Haodf.com. Chinese Journal of Health Policy 2016;9(11):65-69. [doi: 10.3969/j.issn.1674-2982.2016.11.011]
30. Morgan RM, Hunt SD. The Commitment-Trust Theory of Relationship Marketing. J Mark 1994 Jul;58(3):20. [doi: 10.2307/1252308]
31. Wu J, Chen Y, Chung Y. Trust factors influencing virtual community members: A study of transaction communities. J Bus Res 2010 Sep;63(9-10):1025-1032. [doi: 10.1016/j.jbusres.2009.03.022]
32. Piao J, Sun FC. Doctor-patient trust under the perspective of doctors and patients. Article in Chinese. Chinese Medical Ethics 2013;26(6):772.
33. Ridings CM, Gefen D, Arinze B. Some antecedents and effects of trust in virtual communities. J Strategic Inf Sys 2002 Dec;11(3-4):271-295. [doi: 10.1016/S0963-8687(02)00021-5]
34. Li Y, Wang X, Guo R. Study of Doctor-Patient Trust and Influence Factors Based on the View of Patients. Article in Chinese. Chinese Hospital Management 2015;35(11):56-58.
35. Dong E. Empirical study of the impact of patients' trust on health outcomes. Article in Chinese. Chinese Journal of Hospital Administration 2012;28(5):355-359. [doi: 10.3760/cma.j.issn.1000-6672.2012.05.012]
36. Sussman SW, Siegal WS. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption. Inf Sys Res 2003 Mar;14(1):47-65. [doi: 10.1287/isre.14.1.47.14767]
37. Fang A, Li YZ. Computing Algorithm of User Similarity in Virtual Medical Communities. Article in Chinese. Journal of Medical Intelligence 2011;32(3):48-51. [doi: 10.3969/j.issn.1673-6036.2011.03.011]
38. Lu Y. Research on Text Mining in Online Health Community. Dissertation in Chinese. Shanghai Jiao Tong University. 2013. URL: http://d.wanfangdata.com.cn/thesis/Y2497641 [accessed 2015-01-04]
39. Zhao J. Value Cocreation in Online Health Community. Wuhan, China: Wuhan University Press; 2014:145-156.
40. Ma Y, Fu H, Jiang Y. Virtual community: valuable study area in health education in information age. Article in Chinese. Chinese Journal of Health Education 2008;24(4):296-297. [doi: 10.3969/j.issn.1002-9982.2008.04.020]
41. Petty RE, Cacioppo JT, Goldman R. Personal involvement as a determinant of argument-based persuasion. J Pers Soc Psychol 1981;41(5):847-855. [doi: 10.1037//0022-3514.41.5.847]
42. Bhattacherjee A, Sanford C. Influence Processes for Information Technology Acceptance: An Elaboration Likelihood Model. MIS Q 2006 Dec 12;30(4):805-873. [doi: 10.2307/25148755]
43. Filieri R, McLeay F. E-WOM and Accommodation. J Travel Res 2013 Mar 25;53(1):44-57. [doi: 10.1177/0047287513481274]

44. Sussman SW, Siegal WS. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption. Inf Syst Res 2003 Mar;14(1):47-65. [doi: 10.1287/isre.14.1.47.14767]
45. Wang X, Pang S, Luo W. The Establishment and Enlightenment of Consumers' Initial Trust Model of Mobile Banking. Article in Chinese. Modern Management Science 2009;7:98-99. [doi: 10.3969/j.issn.1007-368X.2009.07.036]
46. Li CL. Persuasive messages on information system acceptance: A theoretical extension of elaboration likelihood model and social influence theory. Comput Hum Behav 2013 Jan;29(1):264-275. [doi: 10.1016/j.chb.2012.09.003]
47. Bandura A. Social Foundation of Thought and Action. Englewood Cliffs, NJ: Prentice-Hall; 1986:32-36.
48. Kanjana J. A study on Consumers' Using Behavior of online Travelling Website Based on ELM. Article in Chinese. Chinese Journal of Management 2016;6:889-898. [doi: 10.3969/j.issn.1672-884x.2016.06.012]
49. Yu X, Xu Z, Guo W. Research on the Influence of Online Product Information Presentation on Customers' Behavior Intention. Article in Chinese. Information Studies: Theory and Application 2017 Mar 10:80-84. [doi: 10.16353/j.cnki.1000-7490.2017.10.015]
50. Zhang C, Wang X, Wang C. Research on Influencing Factors of Continuous Information Sharing Behavior of Users in Network Communities. Article in Chinese. Information and Documentation Services 2019;40(3):53-62. [doi: 10.12154/j.qbzlgz.2019.03.009]
51. Ye Y. Simulation Study on Consumer Confidence in Mobile E-Commerce. Article in Chinese. Computer Simulation 2017;34(8):247-250. [doi: 10.3969/j.issn.1006-9348.2017.08.051]
52. Yang Y. Study on the relationship between the shopping guide sharing community trust and network consumption willingness. Dissertation in Chinese. School of Information Management, Central China Normal University. 2014. URL: http://d.wanfangdata.com.cn/thesis/Y2556178 [accessed 2014-09-25]
53. Koo C, Chung N, Nam K. Assessing the impact of intrinsic and extrinsic motivators on smart green IT device use: Reference group perspectives. Int J Inf Manag 2015 Feb;35(1):64-79. [doi: 10.1016/j.ijinfomgt.2014.10.001]
54. Luhmann N, Qu T, Li Q. Trust: A Simplified Mechanism of Social Complexity. Chinese translation. Shanghai, China: Shanghai People's Publishing House; 2005.
55. Li Q, Ruan Y. Impacts of Reputation, Consumer Guarantee System and Online Reviews on Online Customers' First Purchase Intention. Article in Chinese. Economic Survey 2014;31(4):98-103. [doi: 10.15931/j.cnki.1006-1096.2014.04.045]
56. Jiang JH, Chen ZW, Ren JF. Empirical Study on the Effects of Perceived Reputation on Purchase Willing in C2C Electronic Market. Article in Chinese. Soft Science 2011;25(6):130-134. [doi: 10.3969/j.issn.1001-8409.2011.06.028]
57. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electron Commer Res Appl 2015 May;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
58. Liu H. Internet word-of-mouth effects vary from person to person? —The role of individual unique needs in online word-of-mouth influence consumer decision-making. Article in Chinese. Commercial Research 2014(2):97-104. [doi: 10.3969/j.issn.1001-148X.2014.02.015]
59. See-To EW, Ho KK. Value co-creation and purchase intention in social network sites: The role of electronic Word-of-Mouth and trust – A theoretical analysis. Comput Hum Behav 2014 Feb;31:182-189. [doi: 10.1016/j.chb.2013.10.013]
60. WeiYi. Webpage in Chinese. URL: https://www.guahao.com/ [accessed 2020-09-10]
61. Chen W, Huang X, Qi GQ, Bao A, Deng C. Study on Influencing Factors of Women's Online Health Information Sharing Behavior. Article in Chinese. Medicine and Society 2009;32(10):38-41. [doi: 10.13723/j.yxysh.2019.10.010]
62. Lu Q, Li Y, Chen J, Li B. Study on the Influencing Factors of the Patient?s Selection of Doctors in Online Health Care Community. Article in Chinese. Library and Information Service 2019;63(8):011-095. [doi: 10.13266/j.issn.0252-3116.2019.08.011]
63. Cao Y, Guo D. Patients' Behavior of Selecting Physicians in Online Health Community Based on Trust Perspective: Taking the Website of www.haodf.com as an Example. Article in Chinese. Information Studies: Theory and Application 2018 Sep 16;41(9):96-101. [doi: 10.16353/j.cnki.1000-7490.2018.09.016]
64. Kim H, Kim T, Shin SW. Modeling roles of subjective norms and eTrust in customers' acceptance of airline B2C eCommerce websites. Tour Manag 2009 Apr;30(2):266-277. [doi: 10.1016/j.tourman.2008.07.001]
65. Jin L. The Effects of Online WOM Information on Consumer Purchase Decision: An Experimental Study. Article in Chinese. Economic Management 2007(22):36-42.
66. Chen Z, Ping Z. An Experimental Study on the Effects of Mobile Shopping Reviews on Consumer Purchasing Intention. Article in Chinese. Research on Economics and Management 2016;37(6):122-129. [doi: 10.13502/j.cnki.issn1000-7636.2016.06.016]

Abbreviations

ELM: elaboration likelihood model

Edited by G Eysenbach; submitted 16.03.20; peer-reviewed by G Lim, H Chen; comments to author 04.07.20; revised version received 19.07.20; accepted 06.08.20; published 28.09.20.

Please cite as:
Ju C, Zhang S
Influencing Factors of Continuous Use of Web-Based Diagnosis and Treatment by Patients With Diabetes: Model Development and Data Analysis
J Med Internet Res 2020;22(9):e18737
URL: http://www.jmir.org/2020/9/e18737/
doi:10.2196/18737
PMID:32771982

Original Paper

An Ovarian Reserve Assessment Model Based on Anti-Müllerian Hormone Levels, Follicle-Stimulating Hormone Levels, and Age: Retrospective Cohort Study

Huiyu Xu[1,2,3,4*], PhD; Li Shi[1,2,3,4*], MA; Guoshuang Feng[5*], PhD; Zhen Xiao[6], BA; Lixue Chen[1,2,3,4], BA; Rong Li[1,2,3,4], MD, PhD; Jie Qiao[1,2,3,4], MD, PhD

[1]Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

[2]Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing, China

[3]National Clinical Research Center for Obstetrics and Gynecology, Beijing, China

[4]Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing, China

[5]Big Data and Engineering Research Center, Beijing Children's Hospital, Capital Medical University, Beijing, China

[6]Zhejiang Provincial People's Hospital, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Rong Li, MD, PhD
Center for Reproductive Medicine
Department of Obstetrics and Gynecology
Peking University Third Hospital
49 Huayuan North, Haidian District
Beijing,
China
Phone: 86 1082266836
Email: roseli001@sina.com

Abstract

Background: Previously, we reported a model for assessing ovarian reserves using 4 predictors: anti-Müllerian hormone (AMH) level, antral follicle count (AFC), follicle-stimulating hormone (FSH) level, and female age. This model is referred as the AAFA (anti-Müllerian hormone level–antral follicle count–follicle-stimulating hormone level–age) model.

Objective: This study aims to explore the possibility of establishing a model for predicting ovarian reserves using only 3 factors: AMH level, FSH level, and age. The proposed model is referred to as the AFA (anti-Müllerian hormone level–follicle-stimulating hormone level–age) model.

Methods: Oocytes from ovarian cycles stimulated by gonadotropin-releasing hormone antagonist were collected retrospectively at our reproductive center. Poor ovarian response (<5 oocytes retrieved) was defined as an outcome variable. The AFA model was built using a multivariable logistic regression analysis on data from 2017; data from 2018 were used to validate the performance of AFA model. Measurements of the area under the curve (AUC), sensitivity, specificity, positive predictive value, and negative predicative value were used to evaluate the performance of the model. To rank the ovarian reserves of the whole population, we ranked the subgroups according to the predicted probability of poor ovarian response and further divided the 60 subgroups into 4 clusters, A-D, according to cut-off values consistent with the AAFA model.

Results: The AUCs of the AFA and AAFA models were similar for the same validation set, with values of 0.853 (95% CI 0.841-0.865) and 0.850 (95% CI 0.838-0.862), respectively. We further ranked the ovarian reserves according to their predicted probability of poor ovarian response, which was calculated using our AFA model. The actual incidences of poor ovarian response in groups from A-D in the AFA model were 0.037 (95% CI 0.029-0.046), 0.128 (95% CI 0.099-0.165), 0.294 (95% CI 0.250-0.341), and 0.624 (95% CI 0.577-0.669), respectively. The order of ovarian reserve from adequate to poor followed the order from A to D. The clinical pregnancy rate, live-birth rate, and specific differences in groups A-D were similar when predicted using the AFA and AAFA models.

Conclusions: This AFA model for assessing the true ovarian reserve was more convenient, cost-effective, and objective than our original AAFA model.

(J Med Internet Res 2020;22(9):e19096) doi:10.2196/19096

KEYWORDS

ovarian reserve; poor ovarian response; AMH; AFC; FSH; logistic regression

Introduction

The antral follicle count (AFC) is the number of follicles <8 mm in diameter in early gonadotropin-dependent follicular growth. It has been widely accepted that the pool of primordial follicles in the ovary—the ovarian reserve—is related to the number of growing antral follicles. Thus, in theory, the AFC reflects the remaining ovarian follicle pool [1-3]. However, obtaining an accurate AFC demands a time- and resource-consuming ultrasound examination by a skilled transvaginal sonography specialist. The lack of standardization in AFC measurements [4], AFC changes through the menstrual cycle, contraceptive use [5], and the sensitivity and resolution of transvaginal sonography equipment are all confounding factors making the reliable assessment of AFC difficult.

We have previously published a model for estimating ovarian reserves, using 4 predictors: anti-Müllerian hormone (AMH) level, the AFC, follicle-stimulating hormone (FSH) level, and age. This model was named as the AAFA (anti-Müllerian hormone level–antral follicle count–follicle-stimulating hormone level–age) model [6]. With the development of accurate AMH assays [7,8], the level of this hormone might replace the use of AFC in the measurement of ovarian reserve, avoiding the complexity, cost, and interobserver variation in the AFC [9,10]. Here, we aimed to explore the possibility of establishing a model for assessing a true ovarian reserve using the 3 predictors: AMH levels, FSH levels, and age. This model is referred to as the AFA (anti-Müllerian hormone level–follicle-stimulating hormone level–age) model. If the performance of the AFA model without using the AFC is only slightly worse or even similar to the 4-predictor AAFA model, it might be of better clinical significance, especially in physical examination centers or third-party clinical laboratories, which cannot perform AFC measurements by transvaginal sonography.

Methods

Subjects

This was a retrospective observational cohort study using the same dataset as in our previous study [6]. Briefly, data from 2017 to 2018 were selected according to the inclusion and exclusion criteria. In total, we selected 1523 oocytes from ovarian cycles stimulated by a gonadotropin-releasing hormone (GnRH) antagonist 2017 and 3273 oocytes, from 2018. The first and second stimulation cycles were included as described by Xu et al [6], and there were no strict restrictions on the women's age or body mass index. Diseases potentially related to defects in follicular development were excluded, including ovarian cysts, previous ovarian surgery, polycystic ovarian syndrome, previous metabolic or endocrinological diseases, previous tuberculosis, chromosomal abnormalities, and women with pregnancies within the previous 3 months. The need for informed consent by the patients was waived, and institutional review board approval was not needed for the de-identified data in this retrospective analysis, as per the Declaration of Helsinki [11].

Sampling and Endocrine Assays

Venous blood samples were drawn, and the sample tubes were immediately inverted 5 times to facilitate thorough blood clotting. Serum was collected by centrifugation and used for endocrine assessment. The circulating FSH level was measured on menstrual cycle day 2, and the circulating AMH level was measured on any day of the menstrual cycle. Serum FSH measurements were performed using a Siemens Immulite 2000 immunoassay system (Siemens Healthcare Diagnostics). The quality controls used for the FSH assay were Lypocheck Immunoassay Plus Control, Trilevel, catalog number 370, lot number 40340 (Bio-Rad Laboratories). Serum AMH concentrations were measured by an ultrasensitive 2-site enzyme-linked immunosorbent assay (Ansh Laboratories), using quality controls supplied within the kits. The coefficients of variation for each assay were indicated previously [6].

Statistical Analysis

In this study, poor ovarian response with <5 oocytes retrieved was defined as an outcome variable. The predictor variables were age and basal serum FSH and AMH concentrations. A multivariable logistic regression analysis was performed to construct a predictive model for poor ovarian response to stimulation using 2017 data; the data from 2018 were used to validate the performance of that model. Measurements of the area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predicative value (NPV) were used to evaluate the predictive models. The main effect of each predicting variable measures the variation over the distribution of x_j in the mean poor ovarian response. Venn diagrams were used to compare the differences between the AAFA and AFA models.

To rank the ovarian reserve of the whole population, we ranked subgroups according to the predicted probability of poor ovarian response and further divided the 60 subgroups into 4 groups A-D, according to cut-off values consistent with our established AAFA model [6]. Analyses were conducted using SAS JMP Pro (version 14.2; SAS Institute), and $P<.05$ was considered statistically significant.

Results

We previously established an AAFA model, using the 4 predictors of AMH, AFC, FSH, and age [6]. We used this to classify the study population into 4 subgroups: A, B, C and D.

However, the lack of standardization in AFC measurements [4] makes the reliable assessment of AFC very difficult. Moreover, the AFCs were well correlated with AMH, FSH, or age [12,13], implying collinearity. Therefore, we sought to explore the possibility of establishing a model for assessment of the true

ovarian reserve using only 3 predictors—AMH, FSH, and age (ie, the AFA model)—instead of the previous 4-predictor AAFA model [6].

Multivariable Logistic Regression to Build a Predictive Model for Poor Ovarian Response Using the 2017 Data

Basic characteristics of the treatment cycles are shown in Table 1.

As in our previous study, we transformed the 3 continuous variables of age, AMH, and FSH into categorical variables. The data used here were exactly the same as those from 2017, when we built our AAFA model [6]. The cut-off values of each predictor in both AFA and AAFA models are listed in Table 2.

Table 1. Basic characteristics of treatment cycles.

Characteristics	2017 (n=1523)	2018 (n=3273)
Age (years), Mean (SD)	33.4 (5.3)	32.7 (4.8)
Basal FSH[a] (IU/L), Mean (SD)	7.5 (3.3)	7.2 (3.1)
AMH[b] (ng/mL), Median (IQR)	2.2 (1.1-4.0)	2.7 (1.2-4.8)

[a]FSH: follicle-stimulating hormone.

[b]AMH: anti-Müllerian hormone.

Table 2. Comparison of grouping criteria of the AFA and AAFA models.

Grouping criteria	AFA[a] model groups					AAFA[b] model groups	
	0	1	2	3	4	0	1
AMH[c] (ng/mL)	<0.5	0.5 to <1	1 to <1.5	1.5 to <2	≥2	≥1.2	<1.2
Basal FSH[d] (IU/L)	<6.5	6.5 to <8.5	8.5 to <10.5	≥10.5	N/A	≤8	>8
Age (years)	≤30	>30 to 40	>40	N/A	N/A	≤35	>35
AFC[e]	N/A	N/A	N/A	N/A	N/A	≥8	<8

[a]AFA: Anti-Müllerian hormone level–Follicle-stimulating hormone level–Age.

[b]AAFA: Anti-Müllerian hormone level–Antral follicle count–Follicle-stimulating hormone level–Age.

[c]AMH: anti-Müllerian hormone.

[d]FSH: follicle-stimulating hormone.

[e]AFC: antral follicle count.

The transformed categorical variables were then analyzed using multivariable logistic regression. The main effects that each independent variable exerted in this model were AMH (85.2%), followed by FSH (6.8%), and age (2.8%). Thus, we have named this model as AFA based on the order of the main effects of each predictor. The odds ratios of each predictor are indicated in Table 3.

Table 3. The odd ratios of each predictor in the AFA model.

Predictors	OR (95% CI)	*P* value
Intercept	1.753 (1.177-2.611)	<.001
Categorical age (1 vs 0)	2.239 (1.344-3.731)	.006
Categorical age (2 vs 0)	1.863 (1.270-2.734)	.002
Categorical FSH[a] (1 vs 0)	2.300 (1.438-3.681)	.002
Categorical FSH (2 vs 0)	3.594 (2.333-5.538)	.001
Categorical FSH (3 vs 0)	24.641 (14.997-40.488)	<.001
Categorical AMH[b] (0 vs 4)	11.431 (7.281-17.945)	<.001
Categorical AMH (1 vs 4)	5.010 (3.167-7.928)	<.001
Categorical AMH (2 vs 4)	2.211 (1.259-3.882)	<.001
Categorical AMH (3 vs 4)	1.753 (1.177-2.611)	.006

[a]FSH: follicle-stimulating hormone.

[b]AMH: anti-Müllerian hormone.

Comparing the Performances of the AFA and AAFA Models

To further evaluate the performance of this AFA model, we calculated the AUC, sensitivity, specificity, PPV, and NPV in the training set (2017 data) and the validation set (2018 data) as indicated in Table 4.

A calibration plot was drawn to evaluate the calibration performance of the AFA model in the training set and validation set (Multimedia Appendix 1). The performance of the AAFA model in the validation set (2018 data) is indicated in Table 4. A comparison shows that the AUCs of the AFA and AAFA models for the same validation set are similar at 0.853 (95% CI 0.841-0.865) and 0.850 (95% CI 0.838-0.862), respectively. The specificity, sensitivity, PPV, and NPV are also indicated in Table 4. The AUC between AFA model and AAFA model was tested with DeLong test. The difference of the 2 models in AUC level is 0.009 (95% CI –0.004 to 0.022), indicating no significant difference.

Table 4. The performance of AFA model.

Performance indicators	AFA[a] model		AAFA[b] model
	Training set, OR (95% CI)	Validation set, OR (95% CI)	Validation set, OR (95% CI)
AUC[c]	0.860 (0.843-0.877)	0.853 (0.841-0.865)	0.850 (0.838-0.862)
Sensitivity	0.511 (0.456-0.566)	0.489 (0.445-0.533)	0.519 (0.475-0.563)
Specificity	0.940 (0.925-0.952)	0.949 (0.940-0.957)	0.930 (0.920-0.939)
PPV[d]	0.688 (0.626-0.744)	0.633 (0.585-0.680)	0.570 (0.525-0.615)
NPV[e]	0.881 (0.862-0.897)	0.911 (0.901-0.922)	0.915 (0.904-0.925)

[a]AFA: Anti-Müllerian hormone level–Follicle-stimulating hormone level–Age.

[b]AAFA: Anti-Müllerian hormone level–Antral follicle count–Follicle-stimulating hormone level–Age.

[c]AUC: area under the curve.

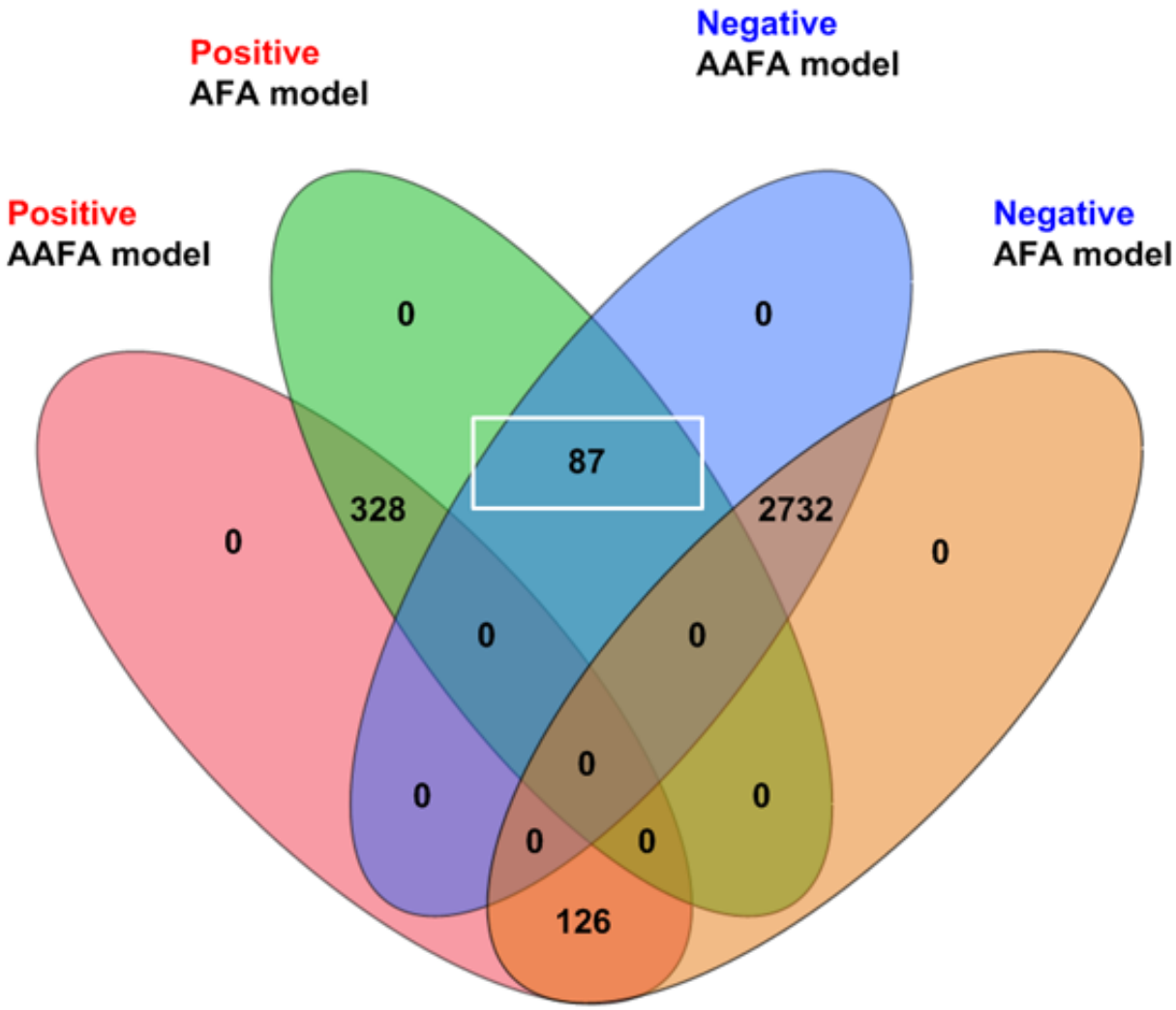
[d]PPV: positive predictive value.

[e]NPV: negative predictive value.

The numbers of overlapping and nonoverlapping cases in the predicted positive (poor ovarian response) and negative estimated by the 2 models are shown in Figure 1.

There were 328 positive (poor responders) and 2732 negative (nonpoor responders) cases overlapping in the 2 models. The 2 models had 93.5% (3060/3273) overlapping positive and negative cases in the 2018 validation set.

Figure 1. Comparison of the performances of the AFA and AAFA models in the 2018 validation data.



Ranking the Ovarian Reserve Based on the Predicted Probability of a Poor Ovarian Response

We previously ranked the ovarian reserve of the whole population according to the predicted probability of a poor ovarian response [6], given that the number of oocytes retrieved is closely related to the number of primordial follicles in the ovarian cortex [14-16]. In this study, we used the same method to rank the ovarian reserve according to the predicted probability of a poor ovarian response calculated using the AFA model. The 60 groups were further divided into 4 subgroups: A, B, C, and D (Multimedia Appendix 2).

The order of ovarian reserve from adequate to poor followed the order of predicted probability of a poor ovarian response from low to high. Women with a predicted probability of more than 50% were classified into the population with diminished ovarian reserve (namely, group D that includes subgroups 43-60), as shown in Multimedia Appendix 2. The actual incidences of poor ovarian response, clinical pregnancy rate per starting cycle, clinical pregnancy rate per embryo transfer cycle, live-birth rate per starting cycle, and live-birth rate per embryo transfer cycle (with 95% CIs) are also indicated in Table 5.

Table 5. The clinical pregnancy rate and live-birth rate in the 4 ovarian reserve groups.

Ovarian reserve group and model	Actual incidence of poor ovarian response (95% CI)	CP[a] per starting Cycles (95% CI)	CP per ET[b] cycles (95% CI)	LB[c] per starting Cycles (95% CI)	LB per ET cycles (95% CI)
A					
AFA[d]	0.037 (0.029-0.046)	0.217 (0.199-0.235)	0.454 (0.424-0.486)	0.176 (0.160-0.193)	0.368 (0.339-0.399)
AAFA[e]	0.038 (0.030-0.046)	0.212 (0.195-0.229)	0.439 (0.409-0.469)	0.174 (0.158-0.190)	0.360 (0.331-0.388)
B					
AFA	0.128 (0.099-0.165)	0.165 (0.123-0.205)	0.298 (0.242-0.361)	0.138 (0.108-0.175)	0.249 (0.197-0.309)
AAFA	0.139 (0.101-0.177)	0.210 (0.166-0.254)	0.370 (0.300-0.439)	0.167 (0.126-0.207)	0.294 (0.228-0.359)
C					
AFA	0.294 (0.250-0.341)	0.157 (0.124-0.197)	0.277 (0.221-0.340)	0.142 (0.110-0.180)	0.249 (0.196-0.310)
AAFA	0.362 (0.308-0.415)	0.140 (0.101-0.179)	0.402 (0.309-0.495)	0.124 (0.087-0.161)	0.355 (0.265-0.446)
D					
AFA	0.624 (0.577-0.669)	0.135 (0.105-0.171)	0.289 (0.229-0.356)	0.080 (0.057-0.110)	0.170 (0.124-0.229)
AAFA	0.571 (0.525-0.616)	0.126 (0.095-0.156)	0.268 (0.208-0.327)	0.077 (0.053-0.102)	0.164 (0.114-0.213)

[a]CP: clinical pregnancy.

[b]ET: embryo transfer.

[c]LB: live birth.

[d]AFA: Anti-Müllerian hormone level–Follicle-stimulating hormone level–Age.

[e]AAFA: Anti-Müllerian hormone level–Antral follicle count–Follicle-stimulating hormone level–Age.

Comparing Specific Differences Between the AFA and AAFA Models in Groups A-D

Figure 2 displays the specific differences between the AFA and AAFA models in classifying the whole population into groups A-D.

The horizontal axis includes the 3273 cases in the 2018 validation data. The 2 models did not show a 3-level difference; that is, there was no case classified as A (good ovarian reserve) by the AAFA model but as D (diminished ovarian reserve) by the AFA model. In addition, most cases were classified into the same groups by both models. However, there were differences for some cases. We focus on 3 groups having 2-level differences defined by the AFA or AAFA models, as shown by the red, green, and purple arrows in Figure 2. The same colors are used to indicate those 3 groups in Figure 3.

Figure 2. Specific differences between the AAFA and AFA models in A-D groups.

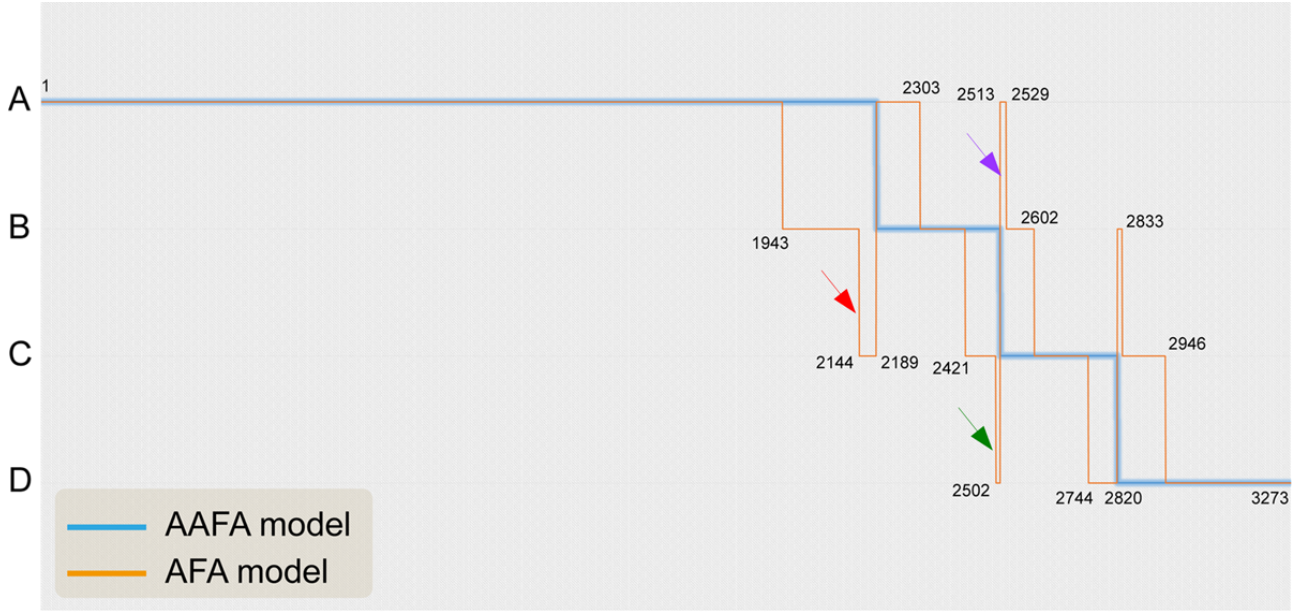


Figure 3. Specific differences between the AAFA and AFA models in A-D groups (tabular form).

AAFA classification	AFA classification	Number of samples	Specific groups (1-60 groups)
A	A	1942	1-11
	B	201	12-15, 17-19, 21-26, 28-29
	C	45	30, 34, 35, 41
B	A	114	2, 5-10
	B	118	12-15, 17-24, 26-29
	C	81	30, 32-37, 39, 41-42
	D	11	43, 45, 46, 53
C	A	16	10
	B	73	14, 16-17, 21-22, 24, 26-27, 29
	C	142	30-34, 36-42
	D	76	43-50, 52-54, 56-57, 59
D	B	13	21, 26, 28
	C	113	30, 32-35, 38-42
	D	328	43-45, 47-60

The raw data and the corresponding predicted probability of a poor ovarian response in the 2 models are listed in Multimedia Appendix 3. The actual incidences of poor ovarian response in the 3 subgroups were 4/45 (red), 5/11 (green), and 1/16 (brown). These results suggest that for the red subgroup, the AAFA classification might be closer to the actual incidence of poor ovarian response (4/45). Thus, these cases should have been placed in group A, rather than in the group C. However, for the purple subgroup with a poor ovarian response incidence of 1/16, the group A classified by AFA model might be more suitable. For the green subgroup with a poor ovarian response incidence of 5/11, not group B by the AAFA model or group D by the AFA model, but group C is more appropriate with its predicted probability of 30% to 50%. For groups having 1-level differences, specific cohorts are shown in Figure 2, Figure 3, and Multimedia Appendix 3.

Discussion

We previously established our AAFA model to assess ovarian reserve based on AMH, AFC, FSH, and age [6]. However, standardization of the AFC has long been difficult for fertility clinics worldwide. In this study, using the same 2018 validation data without the AFC predictor, the AFA model showed similar performance as that of the AAFA model, with an AUC of 0.853 (95% CI 0.841-0.865) vs 0.850 (95% CI 0.838-0.862) for the AAFA model. Since it does not require the AFC, the applicability and cost-effectiveness of the AFA model is better than the AAFA model. Thus, a large number of first- and second-tier hospitals, physical examination centers, or third-party clinical laboratories, which cannot conduct AFC tests, can now assess ovarian reserve using our AFA model.

There were no large (3-level) differences, in that no subject was classified into the A group by the AAFA model and the D group by the AFA model (Figure 2 and Multimedia Appendix 3). There were at most 2 levels of difference, as shown in Figure 2, indicated in red, green, and purple. After referring to the actual rate of poor ovarian response in these groups, we came to the conclusion that the 2 models have their own benefits and can complement each other in assessing ovarian reserve. Integration of these 2 models might give infertility clinics more individualized recommendations before starting controlled ovarian stimulation.

The global infertility rate is increasing, affecting about 1 in 7 couples [17]. A large proportion of women worldwide choose to delay having their first child for pursuit of opportunities to improve their education and workforce participation. It has long been acknowledged that fertility (the ability to establish a clinical pregnancy) decreases with increasing female age. Thus, the prevalence of infertility is increasing worldwide due to the postponement of childbearing. However, many women of reproductive age are not aware of the existing large heterogeneity in ovarian reserve for the same age [18]. In response to the increasing of infertility rate, to achieve a successful pregnancy, an increasing number of couples seek for assisted reproductive treatment. However, not all couples will benefit from it, as the beneficial effect of assisted reproductive treatment is limited in women with diminished ovarian reserve or in women with premenopause [19,20]. If women with potential diminished ovarian reserve could evaluate their ovarian reserve status earlier, it might be possible to avoid the subsequent infertility problem. Our new AFA model provides better means for assessing ovarian reserve, so that women of childbearing age, especially those who hesitate to start a family, might be able to evaluate their ovarian reserve in time.

The circulating AMH concentration is well-correlated with the AFC, and it is considered to be the best predictor for an ovarian response [3,14,21,22]. However, it should be noted that AMH concentrations and AFC are not necessarily linked. The term “ovarian reserve” refers to the number of primordial follicles remaining in the ovarian cortex. AMH is secreted by immature granulosa cells in the gonadotropin-independent phase of follicular development, while the AFC reflects the later gonadotropin-dependent phase. For example, in patients with hypogonadotropic hypogonadism, AFC is undetectable because of the extremely low level of FSH, but such young patients can have a sufficient ovarian reserve, manifested by normal AMH levels and good pregnancy outcomes when undergoing assisted reproductive technology. In addition, some patients exhibit a diminished ovarian reserve and low AMH concentrations but have a satisfactory AFC. AMH gene knockout mice might help us to understand the underlying mechanisms in such patients. In these mice, diminished ovarian reserve induced by the absence of AMH leads to accelerated follicular activation and an increase in the AFC in 4-month-old AMH-null mice (young adult) [23]. Therefore, it is possible that the AMH concentration is a more accurate measure of the actual ovarian reserve than the AFC. Furthermore, the main effect of AMH level was 62.0% in our AAFA model, and 85.2% in the AFA model, meaning that this hormone is the best predictor of ovarian reserve among the existing indicators.

The relationship between AMH concentration and pregnancy outcomes has been investigated extensively [14,24-27]. Fertility is defined as the natural capability to establish a clinical pregnancy [28]. The most accepted predictor for fertility is the ovarian reserve. Within a certain range, the number of primordial follicles does not correlate well with fertility [6], but when the number falls below a certain threshold, as in the case of diminished ovarian reserve defined by our AAFA [6] or AFA models (Table 5), female fertility declines significantly. This might explain the relatively weak relationship between fertility and ovarian reserve. There is a large variation in the number of granulosa cells needed to maintain at least 1 healthy oocyte; however, if there are too few granulosa cells to support at least 1 healthy oocyte, pregnancy is not possible.

There were some limitations to our study. First, it had a retrospective and nonrandomized design. However, as one of the largest reproductive centers in China, there is no strict limit on our selection of patients, thus helping avoid selection bias among our study population. Therefore, our AFA model is relevant to daily clinical practice. Second, our AFA model divides the population into 60 subgroups (3×4×5) rather than the 16 subgroups in the AAFA model. Thus, the sample sizes in our groups were relatively small, such as the 20th group (Multimedia Appendix 2) with only 1 case. We aim to include more samples in the future to verify and improve our formula used in the AFA model. Our last concern is that the positive rate predicted by the validation set (2018 data) is lower than the training set (2017 data), which may be induced by the lower rate of actual poor responders in 2018 data (315/1523 in 2017 data vs 499/3273 in 2018 data). Although the predicted positive rate of the validation set is low, considering the similarity of the AUC of the training set and the validation set, and the main purpose of our research, which is to classify the whole population into more groups according to the predicted probability of poor ovarian response, we believe that the AFA model is satisfactory and comparable to AAFA model. For subsequent related software, we will also integrate the AFA model, the AAFA model, and the actual rate of poor ovarian response in each subgroup together to further optimize the algorithm of this ovarian reserve assessment–related software.

Acknowledgments

We thank James Cummins, PhD, from Liwen Bianji, Edanz Group China, for editing the English text of a draft of this manuscript. This study was supported by the National Key Research and Development Program of China (Grant No. 2016YFC1000201, 2018YFC1002104, 2018YFC1002106, 2016YFC1000302); National Natural Science Foundation of China (Grant No. 81300373, 81771650); the Capital Health Research and Development of Special Project (Grant No. 2018-1-4091); Program for the Innovative Research Team of Yunnan, China (Grant No. 2017HC009); and Major National R & D Projects of China (Grant No. 2017ZX09304012-012).

Conflicts of Interest

None declared.

Multimedia Appendix 1
A calibration plot of the AFA model in the training set and validation set.
[PNG File , 36 KB - jmir_v22i9e19096_app1.png]

Multimedia Appendix 2
Ranking the ovarian reserve into 4 subgroups of A-D based on predicted probability of poor ovarian response in validation data.
[DOCX File , 25 KB - jmir_v22i9e19096_app2.docx]

Multimedia Appendix 3
Raw data and corresponding predicted groups and subgroups classified by the AFA or AAFA models.
[XLSX File (Microsoft Excel File), 443 KB - jmir_v22i9e19096_app3.xlsx]

References

1. de Carvalho BR, Rosa e Silva ACJDS, Rosa e Silva JC, dos Reis RM, Ferriani RA, Silva de Sá MF. Ovarian reserve evaluation: state of the art. J Assist Reprod Genet 2008 Jul;25(7):311-322 [FREE Full text] [doi: 10.1007/s10815-008-9241-2] [Medline: 18679790]
2. Broer SL, Mol BWJ, Hendriks D, Broekmans FJ. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril 2009 Mar;91(3):705-714. [doi: 10.1016/j.fertnstert.2007.12.013] [Medline: 18321493]
3. Broer SL, Dólleman M, Opmeer BC, Fauser BC, Mol BW, Broekmans FJM. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update 2011;17(1):46-54. [doi: 10.1093/humupd/dmq034] [Medline: 20667894]
4. Broekmans FJ, de Ziegler D, Howles CM, Gougeon A, Trew G, Olivennes F. The antral follicle count: practical recommendations for better standardization. Fertil Steril 2010 Aug;94(3):1044-1051. [doi: 10.1016/j.fertnstert.2009.04.040] [Medline: 19589513]
5. Somunkiran A, Yavuz T, Yucel O, Ozdemir I. Anti-Müllerian hormone levels during hormonal contraception in women with polycystic ovary syndrome. Eur J Obstet Gynecol Reprod Biol 2007 Oct;134(2):196-201. [doi: 10.1016/j.ejogrb.2007.01.012] [Medline: 17335955]
6. Xu H, Feng G, Wang H, Han Y, Yang R, Song Y, et al. A novel mathematical model of true ovarian reserve assessment based on predicted probability of poor ovarian response: a retrospective cohort study. J Assist Reprod Genet 2020 Apr;37(4):963-972. [doi: 10.1007/s10815-020-01700-1] [Medline: 32318905]
7. Li H, Wong B, Ip W, Yeung W, Ho P, Ng E. Comparative evaluation of three new commercial immunoassays for anti-Müllerian hormone measurement. Hum Reprod 2016 Dec;31(12):2796-2802. [doi: 10.1093/humrep/dew248] [Medline: 27702835]
8. Ferguson JM, Pépin D, Duru C, Matejtschuk P, Donahoe PK, Burns CJ. Towards international standardization of immunoassays for Müllerian inhibiting substance/anti-Müllerian hormone. Reprod Biomed Online 2018 Nov;37(5):631-640 [FREE Full text] [doi: 10.1016/j.rbmo.2018.08.012] [Medline: 30241771]
9. Indran IR, Huang Z, Khin LW, Chan JK, Viardot-Foucault V, Yong EL. Simplified 4-item criteria for polycystic ovary syndrome: A bridge too far? Clin Endocrinol (Oxf) 2018 Aug;89(2):202-211. [doi: 10.1111/cen.13755] [Medline: 29851127]
10. Eilertsen TB, Vanky E, Carlsen SM. Anti-Mullerian hormone in the diagnosis of polycystic ovary syndrome: can morphologic description be replaced? Hum Reprod 2012 Aug;27(8):2494-2502. [doi: 10.1093/humrep/des213] [Medline: 22693172]
11. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]

12. Nyström A, Mörse H, Nordlöf H, Wiebe K, Artman M, Øra I, et al. Anti-müllerian hormone compared with other ovarian markers after childhood cancer treatment. Acta Oncol 2019 Feb;58(2):218-224. [doi: 10.1080/0284186X.2018.1529423] [Medline: 30558460]
13. Baker VL, Gracia C, Glassner MJ, Schnell VL, Doody K, Coddington CC, et al. Multicenter evaluation of the Access AMH antimüllerian hormone assay for the prediction of antral follicle count and poor ovarian response to controlled ovarian stimulation. Fertil Steril 2018 Aug;110(3):506-513.e3. [doi: 10.1016/j.fertnstert.2018.03.031] [Medline: 29960708]
14. Xu H, Zeng L, Yang R, Feng Y, Li R, Qiao J. Retrospective cohort study: AMH is the best ovarian reserve markers in predicting ovarian response but has unfavorable value in predicting clinical pregnancy in GnRH antagonist protocol. Arch Gynecol Obstet 2017 Mar;295(3):763-770. [doi: 10.1007/s00404-016-4274-8] [Medline: 28012077]
15. Broer S, van Disseldorp J, Broeze K, Dolleman M, Opmeer BC, Bossuyt P, IMPORT study group. Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach. Hum Reprod Update 2013;19(1):26-36. [doi: 10.1093/humupd/dms041] [Medline: 23188168]
16. Qiao J, Wang Z, Feng H, Miao YL, Wang Q, Yu Y, et al. The root of reduced fertility in aged women and possible therapentic options: current status and future perspects. Mol Aspects Med 2014 Aug;38:54-85. [doi: 10.1016/j.mam.2013.06.001] [Medline: 23796757]
17. Talmor A, Dunphy B. Female obesity and infertility. Best Pract Res Clin Obstet Gynaecol 2015 May;29(4):498-506. [doi: 10.1016/j.bpobgyn.2014.10.014] [Medline: 25619586]
18. Daan NM, Fauser BC. Menopause prediction and potential implications. Maturitas 2015 Nov;82(3):257-265. [doi: 10.1016/j.maturitas.2015.07.019] [Medline: 26278873]
19. Hoorens S, Gallo F, Cave J, Grant J. Can assisted reproductive technologies help to offset population ageing? An assessment of the demographic and economic impact of ART in Denmark and UK. Hum Reprod 2007 Sep;22(9):2471-2475. [doi: 10.1093/humrep/dem181] [Medline: 17586831]
20. Habbema JDF, Eijkemans MJ, Nargund G, Beets G, Leridon H, Te Velde ER. The effect of in vitro fertilization on birth rates in western countries. Hum Reprod 2009 Jun;24(6):1414-1419. [doi: 10.1093/humrep/dep004] [Medline: 19233869]
21. Buyuk E, Seifer DB, Younger J, Grazi RV, Lieman H. Random anti-Müllerian hormone (AMH) is a predictor of ovarian response in women with elevated baseline early follicular follicle-stimulating hormone levels. Fertil Steril 2011 Jun;95(7):2369-2372. [doi: 10.1016/j.fertnstert.2011.03.071] [Medline: 21497340]
22. Nardo LG, Gelbaya TA, Wilkinson H, Roberts SA, Yates A, Pemberton P, et al. Circulating basal anti-Müllerian hormone levels as predictor of ovarian response in women undergoing ovarian stimulation for in vitro fertilization. Fertil Steril 2009 Nov;92(5):1586-1593. [doi: 10.1016/j.fertnstert.2008.08.127] [Medline: 18930213]
23. Durlinger AL, Kramer P, Karels B, de Jong FH, Uilenbroek JT, Grootegoed JA, et al. Control of primordial follicle recruitment by anti-Müllerian hormone in the mouse ovary. Endocrinology 1999 Dec;140(12):5789-5796. [doi: 10.1210/endo.140.12.7204] [Medline: 10579345]
24. Merhi Z, Zapantis A, Berger DS, Jindal SK. Determining an anti-Mullerian hormone cutoff level to predict clinical pregnancy following in vitro fertilization in women with severely diminished ovarian reserve. J Assist Reprod Genet 2013 Oct;30(10):1361-1365 [FREE Full text] [doi: 10.1007/s10815-013-0077-z] [Medline: 23963620]
25. Lukaszuk K, Liss J, Kunicki M, Jakiel G, Wasniewski T, Woclawek-Potocka I, et al. Anti-Müllerian hormone (AMH) is a strong predictor of live birth in women undergoing assisted reproductive technology. Reprod Biol 2014 Sep;14(3):176-181. [doi: 10.1016/j.repbio.2014.03.004] [Medline: 25152514]
26. Tal R, Seifer DB, Wantman E, Baker V, Tal O. Antimüllerian hormone as a predictor of live birth following assisted reproduction: an analysis of 85,062 fresh and thawed cycles from the Society for Assisted Reproductive Technology Clinic Outcome Reporting System database for 2012-2013. Fertil Steril 2018 Feb;109(2):258-265. [doi: 10.1016/j.fertnstert.2017.10.021] [Medline: 29331235]
27. Styer AK, Gaskins AJ, Brady PC, Sluss PM, Chavarro JE, Hauser RB, et al. Dynamic antimüllerian hormone levels during controlled ovarian hyperstimulation predict in vitro fertilization response and pregnancy outcomes. Fertil Steril 2015 Nov;104(5):1153-61.e1. [doi: 10.1016/j.fertnstert.2015.07.1161] [Medline: 26315051]
28. Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, et al. The International Glossary on Infertility and Fertility Care, 2017. Fertil Steril 2017 Sep;108(3):393-406 [FREE Full text] [doi: 10.1016/j.fertnstert.2017.06.005] [Medline: 28760517]

Abbreviations

AFA: Anti-Müllerian hormone level–Follicle-stimulating hormone level–Age
AFAA: Anti-Müllerian hormone level–Antral follicle count–Follicle-stimulating hormone level–Age
AFC: antral follicle count
AMH: anti-Müllerian hormone
AUC: area under the curve
FSH: follicle-stimulating hormone
GnRH: gonadotropin-releasing hormone

NPV: negative predicative value
PPV: positive predictive value

Edited by G Eysenbach; submitted 03.04.20; peer-reviewed by S Wu, N Pandiyan; comments to author 12.06.20; revised version received 25.06.20; accepted 25.06.20; published 21.09.20.

Please cite as:
Xu H, Shi L, Feng G, Xiao Z, Chen L, Li R, Qiao J
An Ovarian Reserve Assessment Model Based on Anti-Müllerian Hormone Levels, Follicle-Stimulating Hormone Levels, and Age: Retrospective Cohort Study
J Med Internet Res 2020;22(9):e19096
URL: https://www.jmir.org/2020/9/e19096
doi:10.2196/19096
PMID:32667898

Original Paper

Using Item Response Theory for Explainable Machine Learning in Predicting Mortality in the Intensive Care Unit: Case-Based Approach

Adrienne Kline[1,2,3], PhD; Theresa Kline[4], MSc, PhD; Zahra Shakeri Hossein Abad[3,5], MSc, PhD; Joon Lee[3,5,6], PhD

[1]Department of Biomedical Engineering, University of Calgary, Calgary, AB, Canada

[2]Undergraduate Medical Education, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[4]Department of Psychology, University of Calgary, Calgary, AB, Canada

[5]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[6]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Adrienne Kline, PhD
Department of Biomedical Engineering
University of Calgary
2500 University Drive NW
Calgary, AB
Canada
Phone: 1 5875831725
Email: askline1@gmail.com

Abstract

Background: Supervised machine learning (ML) is being featured in the health care literature with study results frequently reported using metrics such as accuracy, sensitivity, specificity, recall, or F1 score. Although each metric provides a different perspective on the performance, they remain to be overall measures for the whole sample, discounting the uniqueness of each case or patient. Intuitively, we know that all cases are not equal, but the present evaluative approaches do not take case difficulty into account.

Objective: A more case-based, comprehensive approach is warranted to assess supervised ML outcomes and forms the rationale for this study. This study aims to demonstrate how the item response theory (IRT) can be used to stratify the data based on how *difficult* each case is to classify, independent of the outcome measure of interest (eg, accuracy). This stratification allows the evaluation of ML classifiers to take the form of a distribution rather than a single scalar value.

Methods: Two large, public intensive care unit data sets, Medical Information Mart for Intensive Care III and electronic intensive care unit, were used to showcase this method in predicting mortality. For each data set, a balanced sample (n=8078 and n=21,940, respectively) and an imbalanced sample (n=12,117 and n=32,910, respectively) were drawn. A 2-parameter logistic model was used to provide scores for each case. Several ML algorithms were used in the demonstration to classify cases based on their health-related features: logistic regression, linear discriminant analysis, K-nearest neighbors, decision tree, naive Bayes, and a neural network. Generalized linear mixed model analyses were used to assess the effects of case difficulty strata, ML algorithm, and the interaction between them in predicting accuracy.

Results: The results showed significant effects (P<.001) for case difficulty strata, ML algorithm, and their interaction in predicting accuracy and illustrated that all classifiers performed better with easier-to-classify cases and that overall the neural network performed best. Significant interactions suggest that cases that fall in the most arduous strata should be handled by logistic regression, linear discriminant analysis, decision tree, or neural network but not by naive Bayes or K-nearest neighbors. Conventional metrics for ML classification have been reported for methodological comparison.

Conclusions: This demonstration shows that using the IRT is a viable method for understanding the data that are provided to ML algorithms, independent of outcome measures, and highlights how well classifiers differentiate cases of varying difficulty. This method explains which features are indicative of healthy states and why. It enables end users to tailor the classifier that is appropriate to the difficulty level of the patient for personalized medicine.

(J Med Internet Res 2020;22(9):e20268) doi:10.2196/20268

KEYWORDS

item response theory; machine learning; statistical model; mortality

Introduction

Background

This study aims to demonstrate an approach to assess the effectiveness of binary machine learning (ML) classification, which is an alternative to the more traditional single scalar measures in the literature. Our approach uses an item response theory (IRT) model to enhance the understanding of the data set on which ML protocols are run as well as the results of the classification outcomes. Aspects of IRT's utility have recently surfaced in the ML literature, including comparisons of collaborative filtering [1], evaluation of natural language processing systems [2], identification of initial computer adaptive learning items [3], assessment of the utility of ML classifiers [4], and ML feature selection [5]. However, using IRT in the manner proposed in this study has not yet been undertaken.

The varied and numerous contexts (eg, business, finances, medicine, home, government agencies) in which ML is being used is no less than staggering [6]. Since the advent of the development of ML classification protocols, there has been a commensurate interest in assessing and comparing their efficacy [7]. Evaluation techniques fall into several major categories [8-12]. Many techniques use estimates of single scalar values to summarize the quality of classification based on the frequencies in a confusion matrix (true-positive, false-positive, true-negative, and false-negative). The most common measures include accuracy, precision, negative predictive value, sensitivity, and specificity, although some combine sensitivity and specificity (eg, Youden index, likelihoods, and discriminant power) [13]. Further refinements of scalar estimates have been introduced including reducing the amount of bias and variance of the estimate that have enhanced their interpretability [14], presenting statistical comparisons (eg, *t* test, conservative z, McNemar test) between ML protocol scalar outcomes [15] and assessing the invariance of estimates with changes in the confusion matrix [16]. Graphical presentations of the confusion matrix data at various points along a continuum include gain and lift charts, receiver operating characteristic curves, and area under the curve (AUC). These provide a more comprehensive depiction of the various scalar measures [12] by contextualizing them.

Despite the advances in metric development, there is an interest in developing more extensive descriptions of ML classification outcomes. For example, it has been argued that "... any single scalar measure has significant limitations" and "that such measures ... oversimplify complex questions and combine things that should be kept separate" [17]. Another issue is that many different programs, search and optimize strategies, and evaluation approaches have populated the literature, sparking researchers to attempt to systematize the findings for more general consumption [18]. Some reviews have supported the contention that ML algorithms should outperform human experts [19], but others have found that overly complex approaches used in some ML models are no better than simpler, more intuitive models [20]. Studies should give readers an understanding of the reasons why algorithms perform differently, rather than simply providing results of differences between 2 scalar summary values [21,22].

These comments are consistent with general calls for a fuller explanation regarding the interpretability of ML studies [23], particularly in the biomedical and health fields that have been slower to exploit this technology. Overreliance on text-based information rather than contextual elements and the inherent uncertainty in medical decision making have been cited as problems in applying ML findings [24]. The associations within data sets that are theory-free offer little guidance with regard to improving clinical care [25]. Medical professionals are not well trained in how ML works and thus are not able to critically evaluate the utility of the reported findings [26].

The ML criticism of the lack of attention to the unique characteristics of the individual cases is the focus of this study. We propose to address this challenge using a more comprehensive, case-nuanced approach. Although there has been some work in this regard, such as the now accepted wisdom that standard classifiers do not work well with imbalanced data [27], a focus on the individual cases that fall into the *miss* or *false-positive* categories has only rarely been investigated as a point of interest [28].

This lack of attention is highlighted in the assessment study of various ML models; they often result in comparable outcomes as similar percentages of cases are misclassified regardless of the model used [29]. Some insight into this phenomenon was brought to light in a study on benchmarking data sets, where some sets had more *difficult*-to-classify cases and other sets contained largely *easy*-to-classify cases, providing similar results at the aggregate level regardless of the ML approach used [30]. It has been argued that cleaning up the data by eliminating some cases in the training data set is an appropriate tactic to improve classification accuracy [31]. We disagree with this approach and instead argue that hard-to-classify cases should be examined in a more systematic manner. This study shifts the focus from the level of utility of an ML only at the aggregate sample level toward pinpointing where that model falls short and, more importantly, why the model falls short. The process described identifies, *a priori*, which cases in the data set (balanced or imbalanced) will be more or less difficult to classify. The evaluation of ML algorithms across these cases will help to understand why these cases are difficult to classify. Examining this phenomenon in detail is as important as the classification accuracy index of the data set as a whole.

There are 2 fundamental building blocks to any ML system: the features of interest and the cases in the data set. To investigate the research question in this study, methods derived from IRT were employed as they simultaneously estimate the

characteristics of both features and cases. Understanding this phenomenon allows medical professionals to tailor the classifier to the patient.

Similarities Between ML and Test Taking

Following an examination, there are discussions by students about the test items; often they remark "that question was hard, what did you put?" or "that was as easy question." Such comments reflect the purposeful construction of the test items. Some items are designed to be relatively easy to pass whereas others are designed to be more difficult such that only a few can pass. Similarly, students talk about the test takers—"she always gets the highest score in the class" or "I think I have a 50-50 chance of passing." Test takers are quite cognizant of the fact that not all test items are created equal and that not all test takers have the same ability. These fundamental assumptions give rise to IRT, where the characteristics of the items and of the students are modeled together, providing a clearer picture about which items discriminate between which test takers.

A parallel can be drawn between a set of students passing or failing a test based on their performance on a set of items with a set of patients being classified into 1 of 2 categories (alive or not alive) based on their scores on a set of health-related features. Using the *test* item information, an ML classifier should be able to predict which students will pass or fail the test, and using the feature information, an ML classifier should be able to predict which patients will be alive or not alive. There is likely to be a base level at which it correctly classifies cases as belonging to 1 group or another by chance alone, and additional case information on each feature should enhance the prediction. Some cases can be easily partitioned into the *pass the test* (ie, they pass all the items) or *fail the test* (ie, they fail all the items). Cases with more moderate levels of mastery however would be expected to pass some and fail some items (features in ML terms). It is these *difficult* cases where classifiers would be less likely to successfully predict their probability of passing or failing the test. One option to enhance prediction is to add more features to help classify these more difficult cases, but doing so results in high dimensionality, overfitting models, difficult-to-interpret findings, and nongeneralizing results. This quandary is a classic optimization problem in the ML literature.

As not all test takers score 100% or 0% on an examination, some combination of right and wrong answers to questions provides an index of individual test-taker ability in completing the test. The term ability (symbolized by the term theta, θ) is used in the psychometric literature where IRT evolved and is used to describe any latent construct of interest being measured. In this study, within-range or out-of-range laboratory values and vital signs as well as demographic information comprise the features in our data sets. Thus, we can ascertain a case's placement with respect to the underlying distribution of *unhealthiness*. These individual case-based indices create a distribution of unhealthiness across all features (or in our case laboratory values, vital signs, and demographic information). Depending on where individual patients fall on the distribution, the ease with which ML classifiers correctly predict the outcome (mortality) is expected to be affected—those concentrated in the central area of the distribution will be more challenging to correctly classify relative to those cases lying more at the tails of the distribution. Thus, rather than labeling case scores as being on a healthy-unhealthy continuum, suggesting these scores might only be useful in a health context, we use the classification difficulty index (CDI) because of their ease or difficulty in being able to be correctly classified using supervised ML.

The process of generating CDIs on the unhealthiness continuum will be carried out without using the outcome variable of mortality itself, that is, IRT provides case-based scores (CDIs) that can be examined before the data as a collective is subjected to an ML protocol.

Specific Study Hypotheses and Research Questions

The IRT analysis provides case-based CDIs using a set of feature characteristics that do not use the information on the outcome classification variable. CDIs for the sample are generated along the normal distribution, with a mean 0.0 (SD 1.0). It is hypothesized that cases with more centrally located CDIs will be less likely to be classified correctly, whereas cases with more peripherally located CDIs will be more likely to be classified correctly. One research question is as follows: Will some ML classifiers be more accurate in classifying cases at all CDIs? Another research question is as follows: Will some ML classifiers be more accurate than others in classifying cases at different CDIs? Identifying these cases *a priori* provides an alternative manner to evaluate different ML protocols or classification methods and will advance our understanding of ML findings and the data they are being fed with.

Methods

Data Sets

Data were obtained through 2 large, freely available data sets. One was the MIMIC-III (Medical Information Mart for Intensive Care III) database housing health data of >40,000 critical care unit patients at the Beth Israel Deaconess Medical Center admitted between 2001 and 2012 [32,33]. The other was the electronic intensive care unit (eICU) Collaborative Research Database that houses data from critical care unit patients from across the continental United States admitted between 2014 and 2015 [34].

Case Inclusion

Databases were queried using the SQL plug-in for Python (Python Software Foundation). Case inclusion criteria were as follows: (1) age 16 years, (2) at least three-fourth of the features of interest were available for a select case (patient), leading to subsequent imputation, and (3) first hospital visit in the case of repeated patients. Features of predictive interest were selected based on 2 common severity of illness scores: Simplified Acute Physiology Score II and Acute Physiology and Chronic Health Evaluation IV for MIMIC-III and eICU, respectively. To test the hypothesis with both balanced and imbalanced data sets, the number of *death* cases in both data sets (coded *1*) was noted and the same number of cases of *no death* was then randomly selected and incorporated into the balanced data sets. Imbalanced data sets were created by randomly sampling twice as many *no death* cases compared with *death* cases. We used the 1/3:2/3

imbalance ratio to detect any change in results using a somewhat mildly imbalanced than an extremely imbalanced set.

For the MIMIC-III data set, there were 4039 cases that experienced *death in hospital*, resulting in a final and balanced sample size of 8078 and imbalanced sample size of 12,117. In the eICU data set, there were 10,970 *death in hospital* cases. Employing the same methodology resulted in a balanced sample size of 21,940 and an imbalanced sample size of 32,910.

Features

The features included demographic, procedural, pre-existing conditions, and laboratory values (Tables 1 and 2). Normal values were presented and were obtained from the Medical Council of Canada [35] unless otherwise noted. Laboratory values represent the worst values taken during the intensive care unit (ICU) stay in both data sets in the first 24 hours. In the IRT component of the analyses, variables were dichotomized into disease-promoting states (1) akin to *failing* the item on the test or disease-protective states (0; passing the item). Values that fell outside the normal laboratory ranges were coded as 1 (too low or too high). Pre-existing conditions were coded as 1 (present) or 0 (absent). Age was demarcated at 65 years, with those aged >65 years coded 1 and those aged ≤65 years coded 0 [36-39]. For the sex variable, men were assigned 1 and women were assigned 0 [40]. Imputation of missing data was performed using a multiple imputation chained equations technique using the *impyute* library in Python 3.7.7 to preserve the pre-existing distribution of features.

Table 1. Medical Information Mart for Intensive Care III variables based on Simplified Acute Physiology Score II.

Feature name	Description	Normal values, units
AIDS	Pre-existing diagnosis	Absent: 0, 0 or 1
Heme malignancy	Pre-existing diagnosis	Absent: 0, 0 or 1
Metastatic cancer	Pre-existing diagnosis	Absent: 0, 0 or 1
Minimum GCS[a]	Glasgow Coma Scale	15[b], 1-15
WBC[c] minimum	Lowest white blood cell	4-10, 10^9
WBC maximum	Highest white blood cell	4-10, 10^9
Na minimum	Sodium minimum	135-145, mmol/L
Na maximum	Sodium maximum	135-145, mmol/L
K minimum	Potassium minimum	3.5-5, mmol/L
K maximum	Potassium maximum	3.5-5, mmol/L
Bilirubin maximum	Bilirubin maximum	≤1.52, mg/dL
HCO_3 minimum	Bicarbonate minimum	24-30, mmol/L
HCO_3 maximum	Bicarbonate maximum	24-30, mmol/L
BUN[d] minimum	Blood urea nitrogen minimum	7-22, mg/dL
BUN maximum	Blood urea nitrogen maximum	7-22, mg/dL
PO_2	Partial pressure of oxygen	85-105, mm Hg
FiO_2	Fraction of inspired oxygen	21, %
Heart rate mean	Mean heart rate	60-100, bpm
BP mean	Mean systolic blood pressure	95-145, mm Hg
Max temp	Maximum temperature	36.5-37.5, ˚C
Urine output	Urine output	800-2000[e], mL/24h
Sex	Male or female	Male: 1, Female: 0, Male or female
Age	Age in years	≤65: 0, years
Admission type	Emergency or elective	Emergency: 1; else: 0, N/A[f]

[a]GCA: Glasgow Coma Scale.

[b]Teasdale and Jennett, 1974 [41]; Teasdale and Jennett, 1976 [42].

[c]WBC: white blood cell.

[d]BUN: blood urea nitrogen.

[e]Medical CMP, 2011 [43].

[f]N/A: not applicable.

Table 2. Electronic intensive care unit data set variables based on Acute Physiology and Chronic Health Evaluation IV.

Feature name	Description	Normal values, Units
GCS[a]	Glasgow Coma Scale	15[b], 1-15
Urine output	Urine output in 24 hours	800-2000[c], mL/24 hour
WBC[d]	White blood cell count	4-10, 10^9
Na	Serum sodium	135-145, mmol/L
Temperature	Temperature in Celsius	36.5-37.5[e], ˚C
Respiration rate	Highest white blood cell	12-20[f], breaths/min
Heart rate	Heart rate/min	60-100[f], bpm
Mean blood pressure	Mean arterial pressure	70-100[g], mm Hg
Creatinine	Serum creatinine	0.57-1.02 (F[h]); 0.79-1.36 (M[i]), mEq/L
pH	Arterial pH	7.35-7.45, N/A[j]
Hematocrit	Red blood cell volume	37-46 (F); 38-50 (M), %
Albumin	Serum albumin	3.5-5.0, g/dL
PO_2	Partial pressure of oxygen	85-105, mm Hg
PCO_2	Partial pressure carbon dioxide	35-45, mm Hg
BUN[k]	Blood urea nitrogen maximum	7-22, mg/dL
Glucose	Blood sugar level	68-200, mL/dL
Bili	Serum bilirubin	≤1.52, md/dL
FiO_2	Fraction of inspired oxygen	21[l], %
Sex	Male or female	Male: 1; female: 0, M or F
Age	Age in years	≤65: 0, years
Leukemia	Pre-existing diagnosis	Absent: 0, 0 or 1
Lymphoma	Pre-existing diagnosis	Absent: 0, 0 or 1
Cirrhosis	Pre-existing diagnosis	Absent: 0, 0 or 1
Hepatic failure	Pre-existing diagnosis	Absent: 0, 0 or 1
Metastatic cancer	Pre-existing diagnosis	Absent: 0, 0 or 1
AIDS	Pre-existing diagnosis	Absent: 0, 0 or 1
Thrombolytics	Medical intervention	Absent: 0, 0 or 1
Ventilator	Medical intervention	Absent: 0, 0 or 1
Dialysis	Medical intervention	Absent: 0, 0 or 1
Immunosuppressed	Medical intervention	Absent: 0, 0 or 1
Elective surgery	Medical intervention	Absent: 0, 0 or 1

[a]GCS: Glasgow Coma Scale.

[b]Teasdale and Jennett, 1974 [41]; Teasdale and Jennett, 1976 [42].

[c]Medical CMP, 2011 [43].

[d]WBC: white blood cell.

[e]Lapum et al. 2018 [44].

[f]MDCalc [45].

[g]Healthline [46].

[h]F: female.

[i]M: male.

[j]N/A: not applicable.

[k]BUN: blood urea nitrogen.

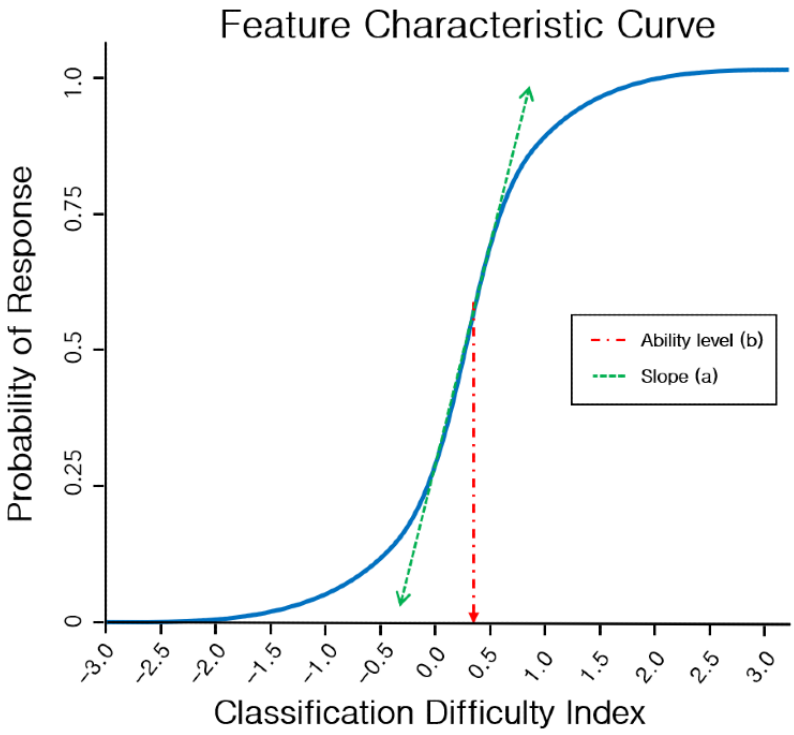
[l]eICU Collaborative Research Database [47].

IRT Analyses

Using the IRTPRO (Scientific Software International) program, a 2-parameter logistic model (2PL) was run on the dichotomous data. The program uses a marginal maximum likelihood estimation procedure to calculate feature and case parameters [48] and assumes that respondents belong to a population that can be characterized by their placement on a latent normal probability distribution—*unhealthiness* in this study with the left and right sides of the distribution indicating better and worse health, respectively [49]. Although higher scores on the latent distribution in IRT usually indicate better outcomes (eg, students have passed more items on a test), in this context, higher scores mean more of the patients' features were out of range and are thus associated with worse outcomes (ie, higher likelihood of death). The output generates logistic item characteristic curves that describe each feature's relationship to the underlying distribution. For each feature, 2 characteristics were estimated, slope and location.

Equation 1 shows a 2PL model in IRT; slope (a_i) captures the *discriminability* capacity of the feature. Feature functions with flat slopes indicate that they are not very discriminatory, whereas those with steep slopes are highly discriminatory, particularly at the inflection point. The location (b_i) denotes where along the function the inflection point occurs. As the functions are set along the standard normal distribution (mean 0.0, SD 1.0), this point indicates where along the unhealthiness continuum the feature is most likely to differentiate cases. An example is presented in Figure 1.

Figure 1. Characteristic curve using a 2-parameter logistic model.



CDI estimation in a 2PL model is calculated based on equation 2, where the probability of obtaining the correct answer is based on the scores on the items' u_i weighted by a_i.

Equation 3, where $u_i \in (0, 1)$ is the score on item i, is called the likelihood function. It is the probability of a response pattern given the CDIs and the item parameters across cases. There is 1 likelihood function for each response pattern, and the sum of all such functions equals 1 at any value of the distribution. On the basis of the pattern of each case's values on the features, the program uses a Bayesian estimation process that provides a CDI on the unhealthiness continuum for each case in the data set.

CDIs are reported on the standard normal distribution and typically range between −2.50 and +2.50. Each case's CDI has its own individual SE around it based on the individual's pattern of results across all features and their unique characteristics. Using the results from the 2PL model, it was possible to identify which of the cases were more centrally or more peripherally located on the distribution and thus would be less or more likely to be accurately classified into their respective categories (no death or death).

To allow for easy visualization and testing of effects, several strata bins were created into which continuous IRT CDIs could be assigned. These *bins* were separated at every 0.5 difficulty

change in the distribution of the data. The first bin was centered over the *0.0* mark to denote the most difficult cases and subsequent bins were demarcated at 0.5 levels toward each periphery. This process of bin allocation continued until all observed CDIs for the cases were accounted for.

ML Analyses

Multiple ML algorithms were tested using the original feature values for both MIMIC-III and eICU data sets. These included logistic regression, linear discriminant analysis, K-nearest neighbors, decision tree, naive Bayes, and neural network. Both the K-nearest neighbors and neural network had their hyperparameters optimized by a grid search. In the case of the K-nearest neighbors, the search grid included K from 1 to 40 and distance methods of Minkowski, Hamming, and Manhattan. The grid investigated for the neural network included activation functions such as softmax, softplus, softsign, relu, Tanh, sigmoid, and hard sigmoid; learning rates such as 0.001, 0.01, 0.1, 0.2, and 0.3; and hidden neurons in a single hidden layer of 1, 5, 10, 15, 20, 25, and 30. In each of these methods, a 10-fold cross-validation was performed, and the numerical prediction was extracted for each case and then reassociated with its subject ID number for graphical plotting. The evaluation methods, accuracy, precision, recall, F1, and AUC metrics were calculated. Accuracy was used to assess the hypotheses and research questions.

Comparison Analyses

To test the main effects of CDI and the repeated measure of the ML classifier as well as their interaction on each case's accuracy score (0,1), generalized linear mixed model (GLMM) [50] analyses were conducted using the GENLINMIX program of SPSS 23 [51]. GENLINMIX uses the penalized quasi-likelihood estimation method for fixed effects. Separate analyses for each of the balanced and imbalanced data sets were conducted. The standard form of the GLMM is shown in equations 4 and 5. y is a response vector, and b is the random effects vector. *Distr* is a conditional distribution of y given b. μ is the conditional mean, and is the dispersion parameter.

In equation 5, g(μ) is the logit link function that defines the relationship between the mean response μ and the linear combination of predictors. X represents the fixed effects matrix, and Z is a random effects matrix, where is simply an offset to the model.

The models specified that (1) all effects are fixed, (2) the dependent variable follows a binomial distribution, and thus the predictors and criterion are linked via a logit function, (3) the residual covariance matrix for the repeated measure (ML classifier) is diagonal, and (4) the reference category was set to 0. Follow-up paired-comparison tests on the estimated marginal and cell means used a *P* level of<.001 to protect against a type I error.

Results

IRT 2PL Model Results

Descriptive results of case CDIs are shown in Table 3, and frequency distributions are shown in Figures 2 and 3 (MIMIC-III) and Figures 4 and 5 (eICU).

It should be noted that the 2 data sets have different distributions, and this fingerprint is inherently unique to the data set processed.

Table 3. Item response theory case classification difficulty index results.

Data set	CDI[a] range	Overall, mean (SD)	Point-biserial correlations[b]		No death, mean (SD)	Death, mean (SD)	*Two-tailed t* value[c]	
			r value	*P* value			*t* test (df)	*P* value
MIMIC-III[d] balanced	−1.81 to +2.16	0.00 (0.85)	0.37	<.001	−0.32 (0.79)	0.32 (0.80)	35.76 (8077)	<.001
MIMIC-III imbalanced	−1.70 to +2.27	0.00 (0.85)	0.35	<.001	−0.21 (0.80)	0.42 (0.80)	40.88 (12116)	<.001
eICU[e] balanced	−2.63 to +2.83	0.00 (0.80)	0.50	<.001	−0.40 (0.73)	0.40 (0.64)	86.18 (21939)	<.001
eICU imbalanced	−2.55 to +2.93	0.00 (0.81)	0.51	<.001	−0.29 (0.73)	0.59 (0.61)	109.09 (32909)	<.001

[a]CDI: classification difficulty index.

[b]Between CDI and outcome (no death or death).

[c]Difference between no death and death means.

[d]MIMIC III: Medical Information Mart for Intensive Care.

[e]eICU: electronic intensive care unit.

Figure 2. Classification Difficulty Indexes in MIMIC-III (A) balanced and (B) imbalanced data. CDI: classification difficulty index; MIMIC: Medical Information Mart for Intensive Care.

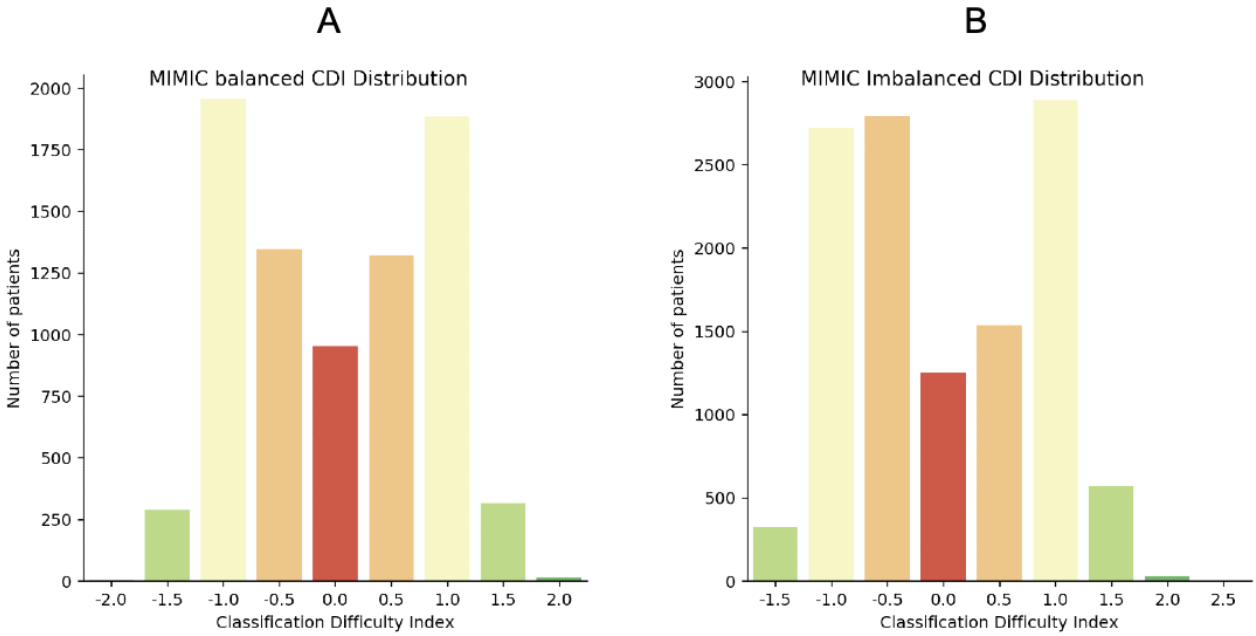
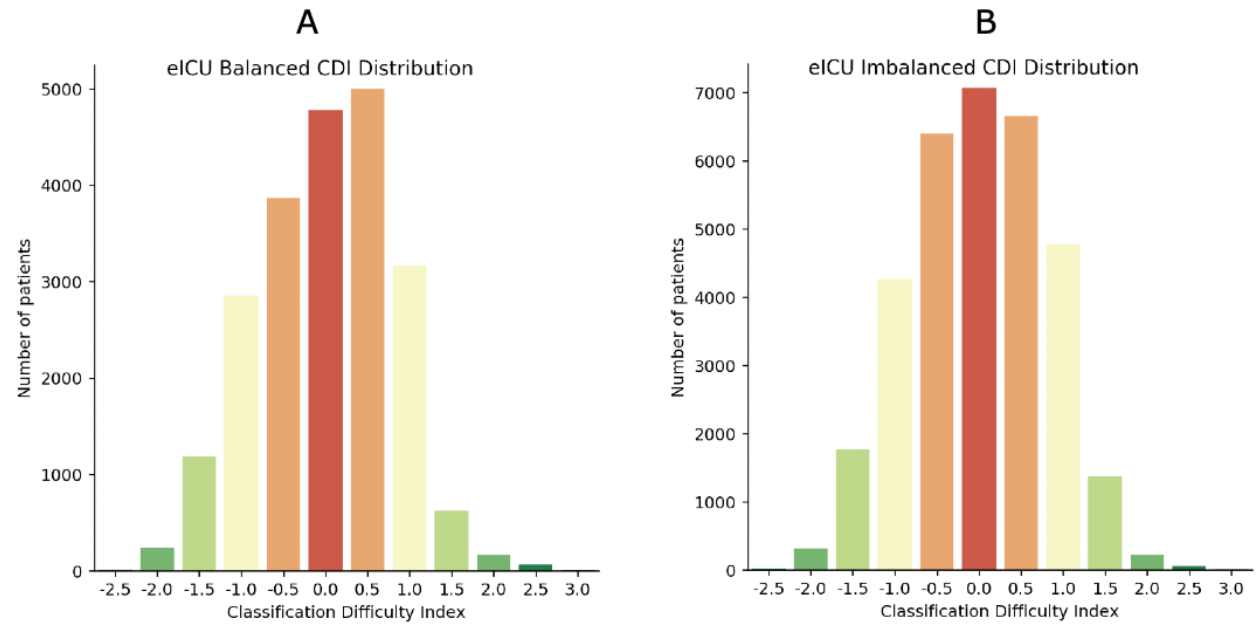


Figure 3. Classification Difficulty Indexes in eICU (A) balanced and (B) imbalanced data. eICU: electronic Intensive Care Unit; DT: decision tree; KNN: K-nearest neighbors; LDA: linear discriminant analysis; LR: logistic regression; NB: naive Bayes; NN: neural network.



Using the feature parameter estimates and case CDI, the unique differentiating capacity for each feature can be depicted by calculating the probability of each case falling into the 0 (no death) or 1 (death) categories. For example, the slope and location parameters for the blood urea nitrogen (BUN) minimum and urine output for the 2 MIMIC-III data sets are shown in Table 4. The higher slope of the BUN minimum feature is contrasted with the very low slope of the urine output feature. These differences highlight the importance of some features over others in terms of being useful in categorizing cases.

Table 4. Medical Information Mart for Intensive Care III feature parameters.

Feature parameters	Slope	Location
Balanced		
Blood urea nitrogen (minimum)	5.64	0.09
Urine output	0.15	−2.23
Imbalanced		
Blood urea nitrogen (minimum)	5.22	0.02
Urine output	0.09	−3.59

Similar to the MIMIC-III results, the IRT analyses of the eICU showed that BUN was a highly discriminating feature whereas urine output was not (Table 5). In fact, many of the features for both MIMIC-III and eICU were not discriminatory (slopes of <0.35 [52]).

Table 5. Electronic intensive care unit feature parameters.

Feature parameter	Slope	Location
Balanced		
Blood urea nitrogen (minimum)	1.55	−0.33
Urine output	0.04	−1.19
Imbalanced		
Blood urea nitrogen (minimum)	1.49	−0.1
Urine output	0.03	−1.39

ML Classification Results

Checking the K-nearest neighbors grid warranted using Manhattan distancing and 27 nearest neighbors for MIMIC-III and Manhattan distancing with 19 neighbors for eICU. The neural network grid search results returned an optimum learning rate of 0.001, activation function softmax, and a number of hidden nodes, 15 for MIMIC-III and 17 for eICU.

Traditional metrics of accuracy, precision, recall, F1, and AUC are presented for MIMIC-III in Table 6.

Table 6. Medical Information Mart for Intensive Care III classification performance in traditional metrics.

Metric	LR[a] (%)	LDA[b] (%)	KNN[c] (%)	DT[d] (%)	NB[e] (%)	NN[f] (%)
Balanced						
Accuracy	75.3	75.0	67.2	70.9	70.4	76.1
Precision	75.8	75.6	69.3	71.1	79.5	75.6
Recall	74.3	73.8	61.8	70.6	54.9	77.2
F1	75.0	74.7	65.3	70.8	64.9	76.4
AUC[g]	75.3	75.0	67.2	70.9	70.4	76.5
Imbalanced						
Accuracy	78.3	77.9	72.8	73.7	75.3	80.5
Precision	73.3	73.8	63.1	60.6	67.7	72.7
Recall	54.8	52.1	44.4	60.6	49.6	66.6
F1	62.7	61.1	52.2	60.6	57.3	69.5
AUC	72.4	71.4	65.7	70.9	68.9	76.9

[a]LR: logistic regression.

[b]LDA: linear discriminant analysis.

[c]KNN: K-nearest neighbor.

[d]DT: decision tree.

[e]NB: naive Bayes.

[f]NN: neural network.

[g]AUC: area under the curve.

In both the balanced and imbalanced MIMIC-III data sets, the neural network outperformed the other classifiers (balanced: accuracy was 76.1% and imbalanced: accuracy was 80.5%) using traditional metrics. It is worth highlighting the role an imbalanced data set has on an increased accuracy and a reduction in precision, recall, and F1.

Table 7 shows our proposed method of demonstrating accuracy as a function of CDI. The metric used in Table 7 is accuracy as F1, recall, and precision were undefined in the extreme negative (where features were predominantly 0), and no cases of *death* existed by which to divide. A parabolic relationship existed in the accuracy level and the strata values, where those more distant from the strata bin=0 were more likely to be classified correctly. ML researchers should be most interested in the *problematic* cases CDI bin 0.0 and where we observe that all classifiers struggle with prediction. These results suggest that even if a classifier outperforms its counterparts as shown in the traditional metrics of Table 6 (eg, neural network), it may be surpassed in the more fine-grained approach shown in Table 7 (eg, naive Bayes algorithm within the +1.5 CDI bin of the balanced data set).

In both the balanced and the imbalanced eICU data sets (Table 8), the neural network outperformed the other classifiers using traditional metrics. Similar to the MIMIC-III findings, the imbalanced data set resulted in increased accuracy and decreased precision, recall, and F1.

Table 9 shows our alternative method of demonstrating accuracy as a function of CDI. Cases that were more distant from the strata bin=0 were more likely to be classified correctly.

Table 7. Item response theory–based Medical Information Mart for Intensive Care III mortality prediction accuracy stratified by classification difficulty index.

Number of cases	CDI[a]	LR[b] (%)	LDA[c] (%)	KNN[d] (%)	DT[e] (%)	NB[f] (%)	NN[g] (%)
Balanced							
1	2.5	100.0	100.0	100.0	100.0	100.0	100.0
13	2.0	92.3	92.3	84.6	92.3	92.3	92.3
316	1.5	90.2	88.2	80.4	80.4	89.2	88.3
1884	1.0	75.6	74.9	68.2	68.8	68.4	77.0
1321	0.5	70.5	70.6	63.5	65.9	65.4	71.1
952	0.0	72.0	72.4	62.8	68.8	66.2	73.9
1346	−0.5	70.9	70.6	60.4	67.1	63.7	72.1
1955	−1.0	77.0	77.1	70.9	75.4	75.2	78.3
288	−1.5	94.8	94.8	83.3	91.0	95.5	94.5
3	−2.0	100.0	100.0	100.0	100.0	100.0	100.0
Imbalanced							
1	2.5	100.0	100.0	100.0	100.0	100.0	100.0
30	2.0	93.3	93.3	76.7	73.3	93.3	93.3
571	1.5	77.4	75.7	64.1	71.1	77.4	78.3
1886	1.0	70.6	70.3	63.9	65.0	64.6	73.3
1537	0.5	76.3	75.5	67.3	71.2	72.7	79.7
1251	0.0	78.7	78.0	75.6	74.5	76.8	80.3
2794	−0.5	75.0	74.5	71.0	72.1	72.3	78.4
2722	−1.0	88.3	88.3	85.0	83.3	87.1	89.1
325	−1.5	99.1	99.1	96.6	98.2	99.1	98.8

[a]CDI: classification difficulty index.

[b]LR: logistic regression.

[c]LDA: linear discriminant analysis.

[d]KNN: K-nearest neighbor.

[e]DT: decision tree.

[f]NB: naive Bayes.

[g]NN: neural network.

Table 8. Electronic intensive care unit classification performance in traditional metrics.

Metric	LR[a] (%)	LDA[b] (%)	KNN[c] (%)	DT[d] (%)	NB[e] (%)	NN[f] (%)
Balanced						
Accuracy	77.9	77.4	67.2	76.7	66.6	84.7
Precision	77.9	78.1	67.9	76.7	73.7	84.5
Recall	77.9	76.3	65.3	76.8	51.6	84.9
F1	77.8	77.2	66.6	76.7	60.7	84.7
AUC[g]	77.9	77.4	67.2	77.1	66.6	85.9
Imbalanced						
Accuracy	78.0	80.1	73.6	81.6	73.3	89.5
Precision	73.6	75.1	64.1	72.1	62.0	84.7
Recall	62.1	60.2	47.2	72.9	51.5	83.5
F1	67.4	66.8	54.4	72.5	56.3	84.1
AUC	75.5	75.1	67.0	79.3	67.9	87.8

[a]LR: logistic regression.

[b]LDA: linear discriminant analysis.

[c]KNN: K-nearest neighbor.

[d]DT: decision tree.

[e]NB: naive Bayes.

[f]NN: neural network.

[g]AUC: area under the curve.

Table 9. Item response theory–based electronic intensive care unit mortality prediction accuracy stratified by classification difficulty index.

Number of cases	CDI[a]	LR[b] (%)	LDA[c] (%)	KNN[d] (%)	DT[e] (%)	NB[f] (%)	NN[g] (%)
Balanced							
2	3.0	100.0	100.0	100.0	50.0	100.0	100.0
61	2.5	82.0	82.0	75.4	78.7	86.9	85.2
160	2.0	81.3	82.5	75.0	76.3	81.9	83.4
621	1.5	86.2	86.8	74.5	79.2	83.7	87.9
3167	1.0	83.7	82.9	72.1	78.3	66.3	85.4
4998	0.5	74.0	72.7	64.7	73.1	55.2	80.9
4776	0.0	70.9	70.1	58.5	71.5	57.3	80.0
3864	−0.5	73.8	74.5	63.3	74.4	67.4	84.3
2858	−1.0	85.4	85.5	74.4	84.8	83.1	91.8
1183	−1.5	92.5	92.6	84.3	91.7	91.9	96.4
240	−2.0	97.1	97.1	91.7	95.8	96.3	97.9
10	−2.5	100.0	100.0	100.0	100.0	100.0	100.0
Imbalanced							
6	3.0	66.7	83.3	83.3	66.6	66.6	83.3
58	2.5	82.8	81.0	69.0	75.9	87.9	84.5
215	2.0	79.1	78.6	67.0	72.6	76.3	82.3
1369	1.5	79.8	79.0	65.4	75.2	72.8	85.7
4776	1.0	72.2	72.4	61.6	74.8	58.4	83.9
6657	0.5	67.3	67.0	72.1	57.3	57.3	83.1
7068	0.0	76.4	76.9	70.0	78.8	70.3	88.5
6396	−0.5	87.1	87.3	83.2	87.3	83.4	93.7
4265	−1.0	94.8	95.0	92.0	94.3	92.7	97.7
1763	-1.5	98.0	98.0	97.1	97.9	97.3	99.4
317	-2.0	99.1	99.1	98.4	98.4	98.4	99.1
20	-2.5	100.0	100.0	100.0	100.0	100.0	100.0

[a]CDI: classification difficulty index.

[b]LR: logistic regression.

[c]LDA: linear discriminant analysis.

[d]KNN: K-nearest neighbor.

[e]DT: decision tree.

[f]NB: naive Bayes.

[g]NN: neural network.

Effect Testing

The CDI group sizes at the extreme ends were too small and were collapsed into the next level down for each data set. Tests of the effects of MIMIC-III are reported in Table 10 and Figure 4.

The MIMIC-III balanced data showed significantly better accuracies for the more peripheral than central CDI bins. K-nearest neighbors and decision tree were the poorest classifiers. Although there was a small significant interaction effect, by and large, the main effects were borne out.

Table 10. Tests of the effects of classification difficulty index, classifier, and their interaction for the Medical Information Mart for Intensive Care III data set.

Effect	Significance		Significant paired comparisons (P<.001; higher accuracies listed first)
	F test (df)	P value	
Balanced			
CDI[a]	123 (6,48456)	<.001	• −1.5 vs −1.0, −0.5, 0.0 • −1.0 vs −.05, 0.0 • +1.0 vs +0.5, 0.0 • +1.5 vs +1.0, +0.5, 0.0
ML[b] classifier	52 (5,48456)	<.001	• LR[c], LDA[d], NB[e], NN[f] vs KNN[g], DT[h] • DT vs KNN
CDI×ML classifier	2 (30,48456)	<.001	• −1.5: LR, LDA, NB, NN, DT vs KNN • −1.0: LR, LDA, NB, NN, DT vs KNN • −0.5: LR, LDA, DT, NN vs NB, KNN • 0.0: LR, LDA, DT, NN vs NB, KNN • +0.5: LR, LDA, NN vs NB, KNN, DT • +1.0: LR, LDA, NN vs NB, KNN, DT • +1.5: LR, LDA, NB, NN vs KNN DT
Imbalanced			
CDI	314 (6,72660)	<.001	• −1.5 vs −1.0, −0.5, 0.0 • −1.0 vs −.05, 0.0 • 0.0 vs −0.5, +0.5, +1.0 • +0.5 vs +1.0 • +1.5 vs +1.0
ML classifier	12 (5,72660)	<.001	• LR, LDA, NB, NN vs KNN, DT
CDI×ML classifier	2 (30,72660)	.004	• −1.5: no differences • −1.0: LR, LDA, NB, NN vs KNN, DT • −0.5: LR, LDA, NN vs NB, KNN, DT • 0.0: NN vs DT

[a]CDI: classification difficulty index.

[b]ML: machine learning.

[c]LR: logistic regression.

[d]LDA: linear discriminant analysis.

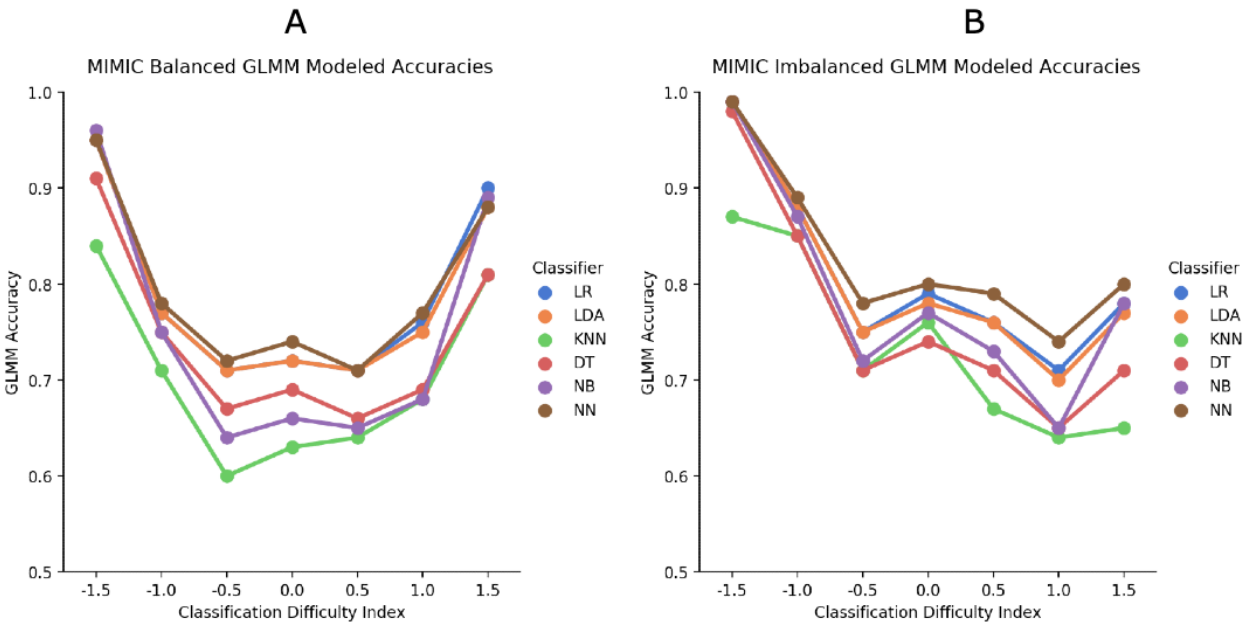
[e]NB: naive Bayes.

[f]NN: neural network.

[g]KNN: K-nearest neighbor.

[h]DT: decision tree.

Figure 4. Medical Information Mart for Intensive Care (MIMIC) III generalized linear mixed model (GLMM) accuracy results; machine learning classifier against CDI for (A) balanced and (B) imbalanced data. DT: decision tree; KNN: K-nearest neighbors; LDA: linear discriminant analysis; LR: logistic regression; NB: naive Bayes; NN: neural network.



The MIMIC-III imbalanced data set showed that at the healthier end of the CDI continuum, more peripheral cases were accurately classified. This was not the case at the central and unhealthier end of the continuum. Like the balanced data set, K-nearest neighbors and decision tree were the poorest classifiers. Although the interaction was significant, most of the paired comparisons supported the main effect findings.

Tests of the effects from eICU are reported in Table 11 and Figure 5.

Table 11. Tests of the effects of classification, classifier, and their interaction for the electronic intensive care unit data set.

Effect	Significance		Significant paired comparisons (P<.001; higher accuracies listed first)
	F test (df)	P value	
Balanced			
CDI[a]	382 (8,131586)	<.001	• −2.0 vs −1.5, −1.0, −0.5, 0.0 • −1.5 vs −1.0, −0.5, 0.0 • −1.0 vs −.05, 0.0 • +1.0 vs +0.5, 0.0 • +1.5 vs +1.0, +0.5, 0.0 • +2.0 vs +0.5, 0.0
ML[b] classifier	58 (5,131586)	<.001	• NN[c] vs LR[d], LDA[e], DT[f] vs NB[g] vs KNN[h]
CDI×ML classifier	9 (40,131586)	<.001	• −2.0: NN vs KNN • −1.5: NN vs LR, LDA, NB, DT vs KNN • −1.0: NN vs LR, LDA, NB, DT vs KNN • −0.5: NN vs LR, LDA, DT vs NB vs KNN • 0.0: NN vs LR, LDA, DT vs NB vs KNN • +0.5: NN vs LR, LDA, DT vs KNN vs NB • +1.0: NN vs LR, LDA vs DT vs KNN vs NB • +1.5: NN, LR, LDA vs NB vs DT vs KNN • −2.0: NN vs KNN
Imbalanced			
Difficulty CDI	1138 (8,197406)	<.001	• −2.0 vs −1.0, −0.5, 0.0 • −1.5 vs −1.0, −0.5, 0.0 • −1.0 vs −.05, 0.0 • −0.5 vs 0.0 • 0.0 vs +0.5, +1.0 • +1.0 vs +0.5 • +1.5 vs +0.5, +1.0 • +2.0 vs +1.0, +0.5
ML classifier	28 (5,197406)	<.001	• NN vs LR, LDA vs DT vs NB, KNN
CDI×ML classifier	4 (40,197406)	<.001	• −2.0: no differences • −1.5: NN vs LR, LDA, NB, KNN, DT • −1.0: NN vs LR, LDA, DT vs KNN, NB • −0.5: NN vs LR, LDA, DT vs KNN, NB • 0.0: NN vs LR, LDA vs DT vs KNN, NB • +0.5: NN vs LR, LDA vs DT vs KNN, NB • +1.0: NN vs LR, LDA, DT vs KNN, NB • +1.5: NN, LR vs LDA vs DT, NB vs KNN • +2.0: NN, LR vs KNN

[a]CDI: classification difficulty index.

[b]ML: machine learning.

[c]NN: neural network.

[d]LR: logistic regression.

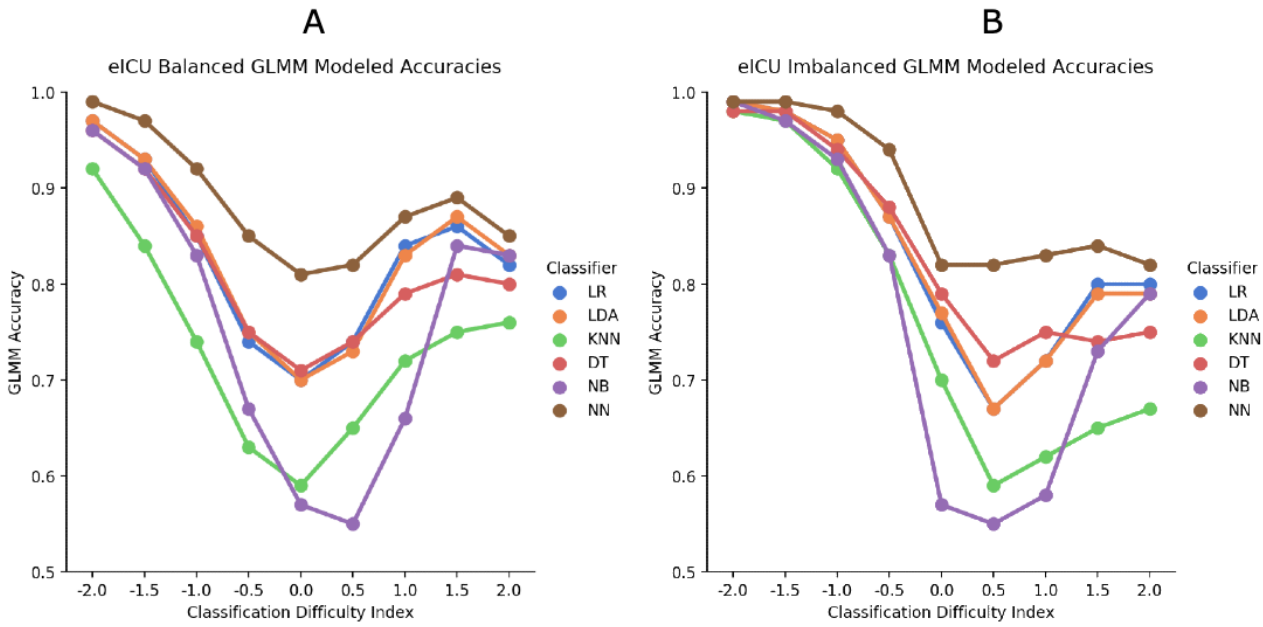
[e]LDA: linear discriminant analysis.

[f]DT: decision tree.

[g]NB: naive Bayes.

[h]KNN: K-nearest neighbor.

Figure 5. Electronic intensive care unit (eICU) generalized linear mixed model (GLMM) accuracy results; machine learning classifier against CDI for (A) balanced and (B) imbalanced data. DT: decision tree; KNN: K-nearest neighbors; LDA: linear discriminant analysis; LR: logistic regression; NB: naive Bayes; NN: neural network.



For the eICU balanced data set, moving away from the central bin showed significantly better accuracy, except at the +2.0 level, which was similar to the +1.5 ML classifier estimated means showed that the neural network had significantly better accuracy than all other classifiers. The overall interaction effect was significant, but the paired comparisons were similar to the main effects.

For the eICU imbalanced data set, more peripheral cases were accurately classified at the healthier end of the distribution, whereas there was only a slight improvement at the unhealthier end. Similar to the other analyses, the neural network showed the best classification accuracy. Although the overall interaction was significant, the neural network continued to be the best classifier.

Discussion

Principal Findings

The results generally supported the hypothesis that cases with more extreme IRT-based CDI values are more likely to be correctly classified than cases with more central CDI values. This provides a unique manner to evaluate the utility of ML classifiers in a health context. We were able to demonstrate that ML classifiers performed similarly for the extreme cases, whereas for the centrally located cases, there were more differences between classifiers. Thus, ML classifiers can be evaluated based on their relative performance with cases of varying difficulty.

Although these were the general results, there were several specific findings that are worth noting. First, the neural network classifier was the best across all situations. The logistic regression and linear discriminant analysis classifiers were close to the second-best classifiers, whereas K-nearest neighbors almost always performed the worst. It is possible, as found in this study, that classifiers may turn out to be consistent over all levels of difficulty. However, owing to the unique characteristics of both data sets and classifiers selected, some algorithms may yield better results at various levels of case difficulty in other samples.

It was also clear that the *peripheral-central* trend of correct classification was most closely adhered to for cases with negative CDI values (ie, at the *healthier* end of the CDI distribution), and this trend was particularly pronounced with the imbalanced (2/3 *nondeath*) data sets. We adopted this modest imbalance in this research to detect trends such as these. This finding is pertinent to ML training protocols in that it is best to train them on balanced data sets before running them on imbalanced ones. There is a clear training effect toward negative CDI or the majority class in our case.

On the basis of the IRT analysis results, easier- and harder-to-classify cases were identified. This has implications for research and clinical practice. Once the cases have been identified, other information gathered from their patient-specific data may provide clues about why they are easier or harder to classify, diagnose, or treat. The features themselves that have varying weighted importance in the indexing process can be examined to assess for any differences in a patient's CDI, that is, not just how many they got *wrong* but which they got *wrong* or *correct* to justify their position in eluding an ML classifier.

As an example of how one could examine more closely the *problematic* patients, we selected the neural network accuracies for each case in the *0* CDI bin in the MIMIC-III balanced data set. This provided 952 cases, 704 (73.9%) were correctly classified and 248 (26.1%) were not. A series of chi-square analyses were conducted using the *in and out-of-range* coding for each of the features crossed with accuracy. Not surprisingly, these cases did not differ on most of the features; the only ones with differences were WBC max (χ^2_1=5.6; *P*=.02), where those

who were more accurately classified had out-of-range scores; bilirubin max (χ^2_1=4.2; *P*=.04), where those who were more accurately classified had normal scores; and mean heart rate (χ^2_1=7.1; *P*=.008, where those who were more accurately classified had normal scores. Using an approach like this can assist in determining which features in more problematic cases may be differentiated.

Relationship With Previous Work

An IRT analysis can assist in providing a better understanding of why the classification process works well or falls short on the set of features and cases under investigation. This moves the field closer to having interpretable and explainable results [53,54]. Recent research with another ICU data set also argues about the importance of explainable processes as well as results [55]. Early research into ML focused on knowledge as an outcome and adopted an informal approach to evaluation. As the field has progressed, the focus shifted to large data sets, mathematical formulae, single evaluation metrics, and statistics, which has impoverished the discipline [22]. "Choosing performance metrics and confidence estimation methods blindly and applying them without any regard for their meaning and the conditions governing them, is not a particularly interesting endeavor and can result in dangerously misleading conclusions" [9].

Limitations and Future Research

Limitations of this research include the fact that classifiers showcased here were not exhaustive, only ICU data sets were used, and converting an out-of-range laboratory value as either *in range=0* or *out of range=1* is reductive. Although this is true, the purpose of this study is to demonstrate a new evaluation metric using a basic 2PL model with binary data.

There are several ways to extend this work. Future research calls for (1) applying this method to other data sets to generalize its use, (2) using polytomous IRT models (eg, 0=in range, 1=somewhat out of range, and 2=very out of range) for more fine-grained case CDI scoring, (3) using multidimensional IRT models to obtain CDIs on >1 underlying dimension, and (4) using this approach to compare human versus machine classification accuracy across case difficulty. We can extend the intersection of ML with clinical medicine if we liken a physician to an ML classifier using feature data. It would be particularly interesting to compare case accuracies based on traditional ML versus clinical classifiers for cases of varying difficulty using an approach similar to that demonstrated in this study. Identifying which cases clinical classifiers are better suited to address, and which cases should be offloaded to an automated system allows for the optimal use of scarce resources. As clinical expertise is developed over time, the use of ML algorithms to assist any single individual would be a moving target and would also serve as a source of future research.

Another way to improve the veracity of the findings would be to address the issue of extraneous features. Several of the features in MIMIC-III and eICU had very low (<0.35) discrimination (slope) parameters, suggesting that there was a lot of *noise* in the cases' CDIs as well as in the ML classifications. It would be a useful exercise to *a priori* determine the most useful features [5] and then run the analyses outlined in this study using a more refined feature set.

Conclusions

As more ML methods are investigated in the health care sphere, concerns have risen because of a lack of understanding regarding why they are successful, especially when compared with physician counterparts. This study has suggested an IRT-based methodology as one way to address this issue by examining the case difficulty in a data set that allows for follow-up into possible reasons why cases are or are not classified correctly.

Using the methods described in this study would signal a change in the way we evaluate supervised ML. Adopting them would move the field toward more of an evaluation system that characterizes the entire data set on which the classifiers are being trained and tested. Doing so circumvents the pitfalls associated with 1 classifier being cited as more accurate or more precise and generates a more tailored approach to ML classifier comparisons. In addition, this methodology lends itself well to *post hoc* inspections of the data as to what makes difficult cases challenging.

The method here presents an intersection of personalized medicine and ML that maintains its explainability and transparency in both feature selection and modeled accuracy, both of which are pivotal to their uptake in the health sphere.

Authors' Contributions

AK contributed to idea generation, study and method design, literature search, data acquisition (MIMIC-III data set), figures, tables, data analysis, and writing. TK contributed to data analysis, writing, and proofing the manuscript. ZA contributed to data acquisition of eICU data set. JL contributed to proofing and journal selection.

Conflicts of Interest

None declared.

References

1. Bergner Y, Dröschler S, Kortemeyer G, Rayyan S, Seaton D, Pritchard D. Model-based Collaborative Filtering Analysis of Student Response Data: Machine-Learning Item Response Theory. In: 5th International Conference on Educational Data Mining. 2012 Presented at: EDM'12; June 19-21, 2012; Chania, Greece. [doi: 10.1007/978-1-4899-7687-1_945]

2. Lalor J, Wu H, Yu H. Building an Evaluation Scale Using Item Response Theory. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016 Presented at: EMNLP'16; November 1-5, 2016; Austin, Texas, USA. [doi: 10.18653/v1/d16-1062]
3. Pliakos K, Joo S, Park J, Cornillie F, Vens C, Van den Noortgate W. Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. Comput Educ 2019 Aug;137:91-103 [FREE Full text] [doi: 10.1016/j.compedu.2019.04.009]
4. Martínez-Plumed F, Prudêncio R, Martínez-Usó A, Hernández-Orallo J. Item response theory in AI: analysing machine learning classifiers at the instance level. Artif Intell 2019 Jun;271:18-42 [FREE Full text] [doi: 10.1016/j.artint.2018.09.004]
5. Kline A, Kline T, Hossein Abad ZS, Lee J. Novel Feature Selection for Artificial Intelligence Using Item Response Theory for Mortality Prediction. In: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2020 Presented at: EMBC'20; July 20-24, 2020; Montreal, QC, Canada p. 5729-5732. [doi: 10.1109/EMBC44109.2020.9175403]
6. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
7. Greiner R, Silver B, Becker S, Groninger M. A review of machine learning at AAAI-87. Mach Learn 1988 Aug;3(1):79-92. [doi: 10.1007/bf00115011]
8. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 2015 Mar 31;5(2):1-11. [doi: 10.5121/ijdkp.2015.5201]
9. Japkowicz N. Why Question Machine Learning Evaluation Methods? (an Illustrative Review of the Shortcomings of Current Methods). In: 34th AAAI Conference on Artificial Intelligence. 2006 Presented at: AAAI'06; February 7-12, 2006; New York, USA p. 11.
10. Kotthoff L, Gent I, Miguel I. A Preliminary Evaluation of Machine Learning in Algorithm Selection for Search Problems. In: Proceedings of the Combustion Institute. 2011 Presented at: SoCS'11; March 8-11, 2011; Washington, DC. [doi: 10.3233/aic-2012-0533]
11. Singh K. Performance evaluation of five machine learning algorithms and three feature selection algorithms for IP traffic classification. Int J Comput Appl 2011;1:32 [FREE Full text]
12. Srivastava T. 11 Important Model Evaluation Metrics for Machine Learning Everyone Should Know. Commonly Used Machine Learning Algorithms: Data Science. 2019. URL: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/ [accessed 2020-08-25]
13. Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-Score and ROC: Performance evaluation of five machine learning algorithms and three feature selection algorithms for IP traffic classification. In: Advances in Artificial Intelligence Systems. Berlin, Heidelberg: Springer; 2006.
14. Santafe G, Inza I, Lozano JA. Dealing with the evaluation of supervised classification algorithms. Artif Intell Rev 2015 Jun 30;44(4):467-508. [doi: 10.1007/s10462-015-9433-y]
15. Stapor K. Evaluating and Comparing Classifiers: Review, Some Recommendations and Limitations. In: International Conference on Computer Recognition Systems. 2017 Presented at: CORES'17; May 22-24, 2017; Polanica Zdroj, Poland. [doi: 10.1007/978-3-319-59162-9_2]
16. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009 Jul;45(4):427-437. [doi: 10.1016/j.ipm.2009.03.002]
17. Drummond C, Japkowicz N. Warning: statistical benchmarking is addictive. Kicking the habit in machine learning. J Exp Theor Artif In 2010 Mar;22(1):67-80. [doi: 10.1080/09528130903010295]
18. Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl 2017;09(01):1-16. [doi: 10.4236/jilsa.2017.91001]
19. Senders J, Staples P, Karhade AV, Zaki M, Gormley W, Broekman M, et al. Machine learning and neurosurgical outcome prediction: a systematic review. World Neurosurg 2018 Jan;109:476-86.e1. [doi: 10.1016/j.wneu.2017.09.149] [Medline: 28986230]
20. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019 Jun;110:12-22. [doi: 10.1016/j.jclinepi.2019.02.004] [Medline: 30763612]
21. Drummond C. Machine Learning as an Experimental Science. AAAI. 2006. URL: https://www.aaai.org/Papers/Workshops/2006/WS-06-06/WS06-06-002.pdf [accessed 2020-09-08]
22. Langley P. The changing science of machine learning. Mach Learn 2011 Feb 18;82(3):275-279. [doi: 10.1007/s10994-011-5242-y]
23. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May 13;1(5):206-215. [doi: 10.1038/s42256-019-0048-x]
24. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. J Am Med Assoc 2017 Aug 8;318(6):517-518. [doi: 10.1001/jama.2017.7797] [Medline: 28727867]
25. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. N Engl J Med 2017 Jun 29;376(26):2507-2509 [FREE Full text] [doi: 10.1056/NEJMp1702071] [Medline: 28657867]

26. Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. AJR Am J Roentgenol 2019 Jan;212(1):38-43. [doi: 10.2214/AJR.18.20224] [Medline: 30332290]
27. Stefanowski J. Dealing with data difficulty factors while learning form imbalanced data. In: Matwin S, Mielniczuk J, editors. Challenges in Computational Statistics and Data Mining. Switzerland: Springer; 2016:333-363.
28. Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting through the safety data haystack: using machine learning to identify individual case safety reports in social-digital media. Drug Saf 2018 Jun;41(6):579-590 [FREE Full text] [doi: 10.1007/s40264-018-0641-7] [Medline: 29446035]
29. Goetz J, Brenning A, Petschko H, Leopold P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 2015 Aug;81:1-11. [doi: 10.1016/j.cageo.2015.04.007]
30. Olson RS, la Cava W, Orzechowski P, Urbanowicz RJ, Moore JH. PMLB: a large benchmark suite for machine learning evaluation and comparison. BioData Min 2017;10:36 [FREE Full text] [doi: 10.1186/s13040-017-0154-4] [Medline: 29238404]
31. Jeatrakul P, Wong KW, Fung CC. Data cleaning for classification using misclassification analysis. J Adv Comput Intell Intell Inform 2010 Apr 20;14(3):297-302. [doi: 10.20965/jaciii.2010.p0297]
32. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M. Data Descriptor: MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:1-9. [doi: 10.1038/sdata.2016.35]
33. Johnson AEW, Mark RG. Real-time mortality prediction in the Intensive Care Unit. 2017 Presented at: AMIA Annu Symp proceedings; 2018; San Francisco p. 994-1003.
34. Pollard TJ, Johnson AE, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. Sci Data 2018 Sep 11;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]
35. Clinical Laboratory Tests - Adult Normal Values. Medical Council of Canada. URL: https://mcc.ca/objectives/normal-values/ [accessed 2020-08-25]
36. Fuchs L, Novack V, McLennan S, Celi LA, Baumfeld Y, Park S, et al. Trends in severity of illness on ICU admission and mortality among the elderly. PLoS One 2014;9(4):e93234 [FREE Full text] [doi: 10.1371/journal.pone.0093234] [Medline: 24699251]
37. Lindemark F, Haaland OA, Kvåle R, Flaatten H, Johansson KA. Age, risk, and life expectancy in Norwegian intensive care: a registry-based population modelling study. PLoS One 2015;10(5):e0125907 [FREE Full text] [doi: 10.1371/journal.pone.0125907] [Medline: 26011281]
38. Alam R, Haque M, Haque M. An appraisal of mortality in intensive care unit of a level III military hospital of Bangladesh. Indian J Crit Care Med 2017 Sep;21(9):594-598 [FREE Full text] [doi: 10.4103/ijccm.IJCCM_250_17] [Medline: 28970660]
39. Atramont A, Lindecker-Cournil V, Rudant J, Tajahmady A, Drewniak N, Fouard A, et al. Association of age with short-term and long-term mortality among patients discharged from intensive care units in France. JAMA Netw Open 2019 May 3;2(5):e193215 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.3215] [Medline: 31074809]
40. Mahmood K, Eldeirawi K, Wahidi MM. Association of gender with outcomes in critically ill patients. Crit Care 2012 May 22;16(3):R92 [FREE Full text] [doi: 10.1186/cc11355] [Medline: 22617003]
41. Teasdale G, Jennett D. Assessment of coma and impaired consciousness. Lancet 1974;304(7872):81-84.
42. Teasdale G, Jennett B. Assessment and prognosis of coma after head injury. Acta Neurochir (Wien) 1976;34(1-4):45-55. [doi: 10.1007/BF01405862] [Medline: 961490]
43. Urine 24-Hour Volume. Penn State Hershey Health Information Library. 2011. URL: http://pennstatehershey.adam.com/content.aspx?productId=117pid=1gid=003425 [accessed 2020-08-25]
44. Lapum J, Verkuyl M, Garcia W, St-Amant O, Tan A. Vital Sign Measurement Across the Lifespan – 1st Canadian Edition. Ryerson University Pressbooks. 2018. URL: https://pressbooks.library.ryerson.ca/vitalsign/ [accessed 2020-09-08]
45. APACHE II Score. MDCalc. URL: https://www.mdcalc.com/apache-ii-score [accessed 2020-08-25]
46. Understanding Mean Arterial Pressure. Health Line. 2019. URL: https://www.healthline.com/health/mean-arterial-pressure [accessed 2020-08-25]
47. apacheApsVar. eICU Collaborative Research Database. URL: https://eicu-crd.mit.edu/eicutables/apacheapsvar/ [accessed 2020-08-25]
48. Bock R, Aitkin M. EM Solution of the Marginal Likelihood Equations. Psychometrika 1981;46(4):443-459.
49. Scientific Software Inc. 2018. URL: https://ssicentral.com/ [accessed 2020-09-08]
50. West BT, Galecki AT. An overview of current software procedures for fitting linear mixed models. American Statistician 2015;135(2):612-615 [FREE Full text] [doi: 10.1198/tas.2011.11077] [Medline: 23606752]
51. Overview (GENLINMIXED Command). IBM - United States. 2020. URL: https://www.ibm.com/support/knowledgecenter/SSLVMB_24.0.0/spss/advanced/syn_genlinmixed_overview.html [accessed 2020-02-10]
52. Baker F. Item Response Theory the Basics of Item Response Theory. New York, USA: ERIC Clearinghouse on Assessment and Evaluation; 2001.
53. Bennett C, Miller M. How reliable are the results from functional magnetic resonance imaging? Ann NY Acad Sci 2010 Mar;1191:133-155. [doi: 10.1111/j.1749-6632.2010.05446.x] [Medline: 20392279]

54. Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell 2019 Feb;267:1-38. [doi: 10.1016/j.artint.2018.07.007]
55. Thorsen-Meyer H, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit Heal 2020 Apr;2(4):e179-e191. [doi: 10.1016/s2589-7500(20)30018-2]

Abbreviations

2PL: 2-parameter logistic
AUC: area under the curve
BUN: blood urea nitrogen
CDI: classification difficulty index
eICU: electronic intensive care unit
GLMM: generalized linear mixed model
ICU: intensive care unit
IRT: item response theory
MIMIC: Medical Information Mart for Intensive Care
ML: machine learning

Edited by G Eysenbach; submitted 14.05.20; peer-reviewed by J Lalor, M Sokolova; comments to author 22.06.20; revised version received 02.07.20; accepted 08.08.20; published 25.09.20.

Please cite as:
Kline A, Kline T, Shakeri Hossein Abad Z, Lee J
Using Item Response Theory for Explainable Machine Learning in Predicting Mortality in the Intensive Care Unit: Case-Based Approach
J Med Internet Res 2020;22(9):e20268
URL: http://www.jmir.org/2020/9/e20268/
doi:10.2196/20268
PMID:32975523

Original Paper

Giving Children With Osteogenesis Imperfecta a Voice: Participatory Approach for the Development of the Interactive Assessment and Communication Tool Sisom OI

Maia Siedlikowski[1], RN, MSc; Frank Rauch[2], MD; Argerie Tsimicalis[1,2], RN, PhD

[1]Ingram School of Nursing, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada

[2]Shriners Hospitals for Children - Canada, Montreal, QC, Canada

Corresponding Author:
Argerie Tsimicalis, RN, PhD
Ingram School of Nursing
Faculty of Medicine and Health Sciences
McGill University
Suite 1800 Room 1835
680 Sherbrooke Street West
Montreal, QC,
Canada
Phone: 1 5147706069
Email: argerie.tsimicalis@mcgill.ca

Abstract

Background: Children with osteogenesis imperfecta (OI) experience acute and chronic symptoms that expose them to physical, mental, and social challenges. Empowering these children by involving them in their care can help them to cope with OI. Sisom is an interactive assessment and communication tool designed to help children aged 6-12 years with chronic illnesses express their symptoms. This tool has not yet been adapted to the unique needs of OI.

Objective: The aim of this study was to develop a Sisom OI paper prototype by seeking the perspectives of end users.

Methods: A participatory approach was adopted to develop the prototype overseen by an expert panel of 9 clinicians at a university-affiliated pediatric hospital. Purposive sampling was used to recruit 12 children with OI who were aged 6-12 years. The study was carried out over the course of 3 feedback cycles. Data were deductively interpreted using content analysis techniques.

Results: Overall, 64% (57/89) of the Sisom symptoms were deemed relevant for inclusion in Sisom OI, with 42% (37/89) directly incorporated and 22% (20/89) incorporated with changes. In total, 114 symptoms were used to create the prototype, of which 57 were newly generated. The relevant symptoms addressed children's thoughts and feelings about hospitalization and their wishes for participation in their own care. The new symptoms addressed fractures, body image, and social isolation related to difficulties with accessibility and intimidation.

Conclusions: Once developed, Sisom OI will offer clinicians an innovative and child-centered approach to capture children's perspectives on their condition.

(J Med Internet Res 2020;22(9):e17947) doi:10.2196/17947

KEYWORDS

child health; symptom assessment, communication, mobile apps, software

Introduction

In health care, children need to be enabled to make their views known on issues that affect them [1,2]. In particular, those living with chronic illnesses have a thorough understanding of their condition [3,4]. This positions them at the center of their own care. Yet, clinicians are confronted with numerous challenges in eliciting these children's views [5]. There is a risk that their symptoms will be underdiagnosed and undertreated. Children's perceptions of their symptoms may be inadequately evoked, their rights for participation poorly applied, and thus their needs often neglected [6-10]. A growing body of evidence suggests that interactive software offers an innovative way to enable children to express themselves [11-14].

Sisom (Norwegian acronym derived from “Si det som det er” meaning “Tell it as it is”) is an award-winning, rigorously tested, interactive, computerized tool that helps children aged 6-12 years with chronic illnesses express their symptoms [14]. It is also considered as a creative system to help clinicians better understand children’s perspectives [12]. It utilizes spoken text, sound, and animations to depict symptoms that are each represented by an animated scene within one of the 5 symptom islands. One first creates an “avatar” and is then prompted by Sisom to indicate the presence and severity of the symptoms displayed by using a 5-point Likert-type scale. Upon completion, Sisom generates a child-friendly Symptom Report that can be shared with family and clinicians. The working of Sisom can be viewed on the demo clip available on the internet. Sisom, which was originally designed for children with cancer, adapted for children with congenital heart disease, and in the process of being adapted for children with diabetes, also has the potential to engage children with other chronic conditions to participate in their own care [15].

Osteogenesis imperfecta (OI) is a chronic condition, which has not yet been studied with respect to children’s perceptions of their symptoms. Moreover, to our knowledge, no interactive computerized tools have yet been designed to attend to the unique needs of this population. OI is the most common of the inherited bone disorders and is usually caused by mutations in collagen type I encoding genes [16,17]. The principal clinical feature of OI is bone fragility that leads to frequent fractures. Since there is no cure for OI, health services focus on rehabilitation as well as pharmacological and surgical interventions to prevent or treat fractures and to maximize mobility [18]. The therapeutic goal is an increase in function and a decrease in fractures. However, pain, fatigue, and varying degrees of physical limitations may hinder participation in daily activities, acceptance by peers, and lead to feelings of fear, otherness, and isolation [13,17,19-21]. Overall, relatively little attention has been directed toward understanding the day-to-day symptoms of children with OI from their own perspective.

Sisom addresses the challenges associated with capturing the child’s perspective on their own health status. In studies in Norway and the United States, children with cancer who used Sisom felt better prepared and expressed twice as many symptoms than their peers during “conventional” consults [12,15,22-24]. When oncologists and registered nurses used the Sisom Symptom Report, they asked a large number of clarifying questions, gave more detailed explanations, and communicated with greater empathy, all within the same period allocated for “conventional” consults [12,22-24]. In studies in Canada and the United States, children have expressed an overwhelming interest in using Sisom in a variety of settings such as at home, school, and clinical environments. These children have also remarked the many benefits of Sisom in helping them express themselves [14,22,25].

Adapting Sisom for children with OI has the potential to generate useful and meaningful data that will serve to establish a more comprehensive and “child-friendly” model of care for this population [13]. The purpose of this study was to develop the Sisom OI paper prototype. A participatory approach was used to seek the perspectives of end users, particularly children and clinicians, to inform the development of Sisom OI.

Methods

Design

Following institutional review board approval (A06-B29-17B), this descriptive study was conducted at a university-affiliated, nonprofit, pediatric, orthopedic hospital in Montreal, Quebec, specialized in OI care.

Participants and Recruitment

Recruitment took place between August 2017 and December 2017. Purposive sampling was used to allow for maximum variation in age, self-identified gender, and type of OI for children. A sample size of 10-15 children and 5-10 clinicians was proposed [26,27]. Previous Sisom studies with similar designs as this study have included between 5-12 participants and have successfully established the validity and usability of the tool [14,15,22]. The clinicians were approached by an email sent by a nonauthoritative colleague, not affiliated with the study, and invited to participate. The children with OI were recruited by reaching out to clinicians, who assisted by identifying, screening, and approaching families to determine if they were interested in hearing more about the study. One member of the study team was responsible for providing a verbal and written explanation of the study to those interested in obtaining written informed consent or assent.

Data Collection and Procedure

This study was carried out over the course of 3 feedback cycles with 2-6 semistructured, face-to-face, individual child interviews per cycle. The parent(s) or legal guardian(s) were given a choice to be present during the interview. The lead author conducted the audio-recorded interviews. The child was invited to use Sisom, which was installed on a laptop, with the lead author. Throughout the interview, children were prompted to answer questions related to the content in Sisom and to consider whether it reflected their own symptoms. Potential drawings that emerged were collected, described, and will be shared in the future with Sisom OI developers. The length of the interview depended on the interest of the child and varied from 20 to 60 minutes in length. Field notes were recorded during and immediately after each interview.

Following each child feedback cycle, the lead author summarized the children’s input on Sisom symptoms, vignettes, and avatars, and hosted an expert panel meeting wherein the synthesized data generated from each child feedback cycle were relayed back to the clinicians for input. Moreover, any discrepancies, similarities, or ambiguities in the children’s responses were discussed. These meetings were carried out with all clinicians in person as one group at the study site and were facilitated by the lead author and another member of the study team. Clinicians were also given the opportunity to view Sisom themselves and were prompted to consider changes that they thought were relevant to OI. Field notes were recorded during and immediately after each meeting. Multiple data sources were collected and they included the self-reported sociodemographic questionnaires for clinicians and children, the Sisom checklists

for clinicians and children (Multimedia Appendix 1), transcribed audio recordings from child interviews, written summaries of the audio recordings from expert panel meetings, and field notes from child interviews and expert panel meetings, which included detailed descriptions of nonverbal data, other observations, impressions, and any drawings generated.

Data Analysis

In the following analyses, a symptom is defined as any question asked by Sisom. Self-report sociodemographic questionnaire data were descriptively analyzed to characterize our samples. Data analysis occurred concurrently with data collection [28,29].

Expert Panel Meetings

During the first meeting, clinicians viewed Sisom and used their clinical expertise to categorize symptoms according to the following predetermined mutually exclusive categories: (1) Relevant, (2) Irrelevant, (3) To modify, (4) To add, and (5) Unsure. Any symptoms initially coded as "Unsure" were subsequently coded into one of the other categories by following the procedures of data triangulation. During the second expert panel meeting, data collected from the second cycle of the child interviews were shared and critically examined. This contributed to establishing the focus of the third cycle of child interviews. During the third expert panel meeting, data collected from the third cycle of the child interviews were shared and a comprehensive list of Sisom OI symptoms was established. The study team then reviewed this list for feasibility and the final Sisom OI paper prototype was created. During the fourth expert panel meeting, the final Sisom OI paper prototype was shared with the clinicians.

Child Interviews

These data were analyzed according to island, deductively, by using content analysis techniques. Guided by the following framework, data pertaining to symptoms, vignettes, or avatars were coded into 5 categories for each island: (1) Relevant, (2) Irrelevant, (3) To modify, (4) To add, and (5) Unsure. The children's rationales for coding the content were highlighted by matching their quotes to the corresponding symptoms, vignettes, or avatars. Drawings were matched to corresponding symptoms and used to showcase children's suggestions of vignettes for future development. Following the conclusion of each cycle, the child interviews for that cycle were compiled and synthesized in preparation for the next expert panel meeting.

Integration of Expert Panel Meetings and Child Interviews

Integration consisted of an iterative process of data reduction of transcripts, summaries, and field notes; data display in the form of lists, tables, and figures; conclusion drawing from recurrent patterns; and verification by drawing contrasts [30]. Constant comparative methods were applied within each cycle. This involved a comparison of elements present in one data source with those in another to determine similarities [28]. In this fashion, commonalities were identified among data sources. The content that clinicians, children, and the study team all agreed on was the symptoms that were considered significant and ranked the highest in terms of priority for software development. The content was then integrated and tabulated to create the Sisom OI paper prototype (Multimedia Appendix 1). An audit trail composed primarily of methodological and analytical documentation was kept, which permitted the reproducibility and the transferability of the process [31].

Results

Sample Characteristics

In total, 12 children participated in this study (Table 1). No child withdrew from the study. Altogether, 9 clinicians participated in this study (Table 2). The clinician participation rate was 100% (9/9). No clinician withdrew from the study. The clinicians recruited included 4 nurses, 1 physiotherapist, 1 occupational therapist, 1 social worker, 1 special education teacher, and 1 child life specialist.

Table 1. Demographic characteristics of the children (n=12).

Characteristics	Values
Age (years), range 6-12, mean (SD)	9 (2)
Number of family members living at home, range 2-6, mean (SD)	4 (1)
Gender, n (%)	
Male	7 (58)
Female	5 (42)
Nationality, n (%)	
Provincial (Quebec)	5 (42)
National (Canada)	3 (25)
International	4 (33)
Language(s) spoken at home, n (%)	
English	4 (33)
French	4 (33)
Bilingual (English and French)	1 (8)
Bilingual (English and other)	3 (25)
Current fracture, n (%)	
Yes	5 (42)
No	7 (58)
Use of mobility device(s), n (%)	
Wheelchair	2 (17)
Wheelchair and walker	4 (33)
None	6 (50)
Use of computer or tablet at school, n (%)	
Yes	9 (75)
No	3 (25)
Use of computer or tablet at home, n (%)	
Yes	12 (100)
No	0 (0)
Use of own mobile phone, n (%)	
Yes	5 (42)
No	7 (58)

Table 2. Demographic characteristics of the clinicians (n=9).

Characteristics	Values
Full-time employment, n (%)	9 (100)
Experience	
Number of years in profession, range 7-35, mean (SD)	20 (11)
Cumulative number of years in profession	176
Number of years in OI[a] care for children, range 3-30, mean (SD)	15 (11)
Cumulative number of years in OI care for children	136

[a]OI: osteogenesis imperfecta.

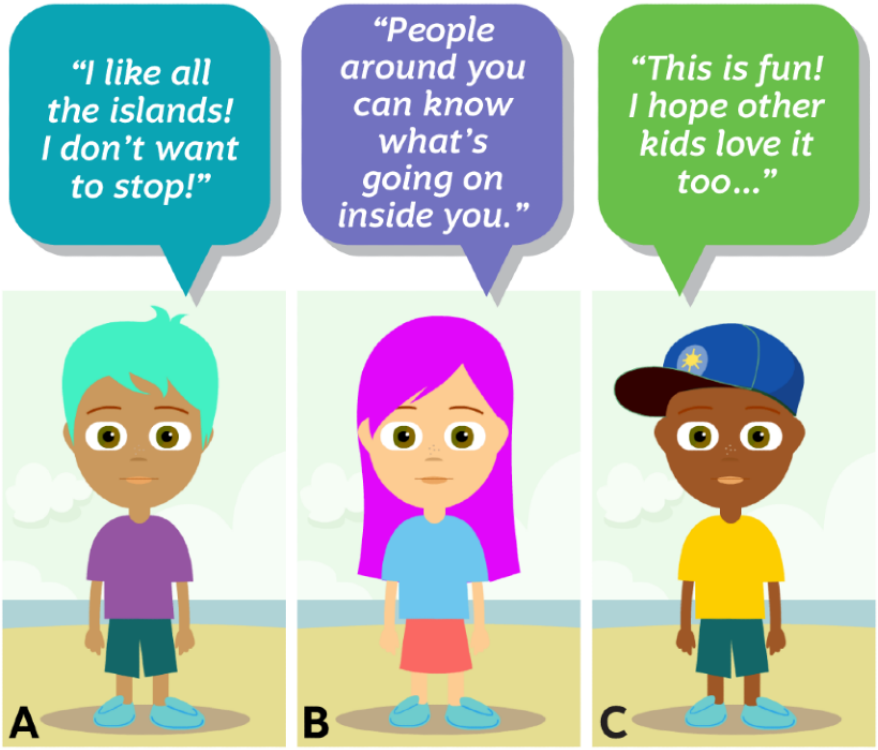
Expert Panel Meeting Findings

In the first meeting, clinicians revealed that they were keen to implement Sisom OI into their practice. The clinicians also assessed the 89 Sisom symptoms for relevance. The first cycle of child feedback was shared with clinicians who highlighted the importance of having Sisom OI adopt a more positive lens. Syntax changes were suggested. Then, a new list of OI-specific symptoms was created. Based on their clinical expertise, clinicians identified several islands that would benefit from further child input: "My Body," "At The Hospital," and "About Managing Things." During the second meeting, clinicians agreed on the need to capture the diverse experiences of pain in OI within the "My Body" island. The clinicians suggested that the "At the Hospital" subisland be divided into several core areas based on the health care trajectory of a child who arrives with a fracture. In the third meeting, minor changes were made to the "The Bathroom" subisland's symptoms to reflect the extra help required for mobilization. The clinicians suggested the creation of a new subisland in "About Managing Things" called "Getting Around." This was in order to capture the importance of accessibility for children with OI and the challenges associated with engaging in leisure activities in public spaces. During the fourth meeting, the clinicians reviewed all Sisom OI content. This meeting concluded with a rich discussion about implications for practice. It was acknowledged that factors such as context of completion, level of parent involvement, and level of clinician involvement would all depend on the child.

Child Interview Findings

These findings are summarized according to Sisom islands. Half of the children (n=6) were accompanied by one of their parents during their interviews. There were 4 mothers and 2 fathers. All children enjoyed using Sisom and completed it with excitement (Figure 1). One-third of the children (n=4) expressed themselves through songs during their Sisom journey. One-third of the children (n=4) chose to check their Sisom Symptom Reports. Finally, some children discussed implementation wishes (Figure 1).

Figure 1. General impressions of Sisom. A. Quote by Participant #6, 6 years old. B. Quote by Participant #4, 8 years old. C. Quote by Participant #5, 10 years old.



Avatar

All children expressed a strong desire for the avatar to reflect their identity, which they defined by their physical appearance, inside and out, and by their preferences. Several children explained what helped support their bodies from the inside (eg, surgical rods) and from the outside (eg, mobility devices). To create space for the gender spectrum, "Choose Boy or Girl" was removed. There were 3 main additions: "Choose your current mood," "Choose what helps you get around," and "Choose what helps support your body" to celebrate their uniqueness.

> *Makes you feel like you're not like everybody else. You can see yourself through a computer technological program!* [Participant #4, 8 years old]

About Me

Several children spoke of the importance of creating a space to help others get to know them.

> *I think there should be one more island about your imagination and what you expect from others. There should also be an area where you can express your feelings about who you are.* [Participant #5, 10 years old]

In light of this, 3 main additions were made: "Here you can tell about your imagination, your wishes, and your dreams," "Here you can tell about your family," and "Here you can tell about your friends" (Table 3). Table 3 illustrates the symptoms from Sisom that were deemed "Relevant" and directly incorporated into Sisom OI, the symptoms that were deemed "To Modify" and incorporated into Sisom OI with changes to syntax, answer options, location, or vignette, and the symptoms that were "To Add," that is, entirely newly generated symptoms from the child feedback cycles overseen by the expert panel. Two children spontaneously made the same suggestion of creating a space to journal their answers to allow for flexibility through drawn, written, or spoken entries.

Table 3. Composition of the Sisom OI islands.

Island	Total number of symptoms	Relevant symptoms, n (%)	Modified symptoms, n (%)	New symptoms, n (%)
Avatar				
Sisom	4	1 (25)	2 (50)	N/A[a]
OI[b] version	8	N/A	N/A	5 (63)
About Me				
Sisom	0	0 (0)	0 (0)	N/A
OI version	3	N/A	N/A	3 (100)
At the Hospital				
Sisom	18	7 (39)	2 (11)	N/A
OI version	19	N/A	N/A	10 (53)
My Body				
Sisom	28	12 (43)	4 (14)	N/A
OI version	28	N/A	N/A	12 (43)
About Managing Things				
Sisom	21	6 (29)	7 (33)	N/A
OI version	27	N/A	N/A	14 (52)
Thoughts and Feelings				
Sisom	10	7 (70)	3 (30)	N/A
OI version	18	N/A	N/A	8 (44)
Things One Might Be Afraid of				
Sisom	8	4 (50)	2 (25)	N/A
OI version	11	N/A	N/A	5 (45)
Total				
Sisom	89	37 (42)	20 (22)	N/A
OI version	114	N/A	N/A	57 (50)

[a]Not applicable.

[b]OI: osteogenesis imperfecta.

At the Hospital

Overall, 50% (9/18) of the "At the Hospital" symptoms were retained. However, this whole island was reorganized to fit the experiences of children with OI.

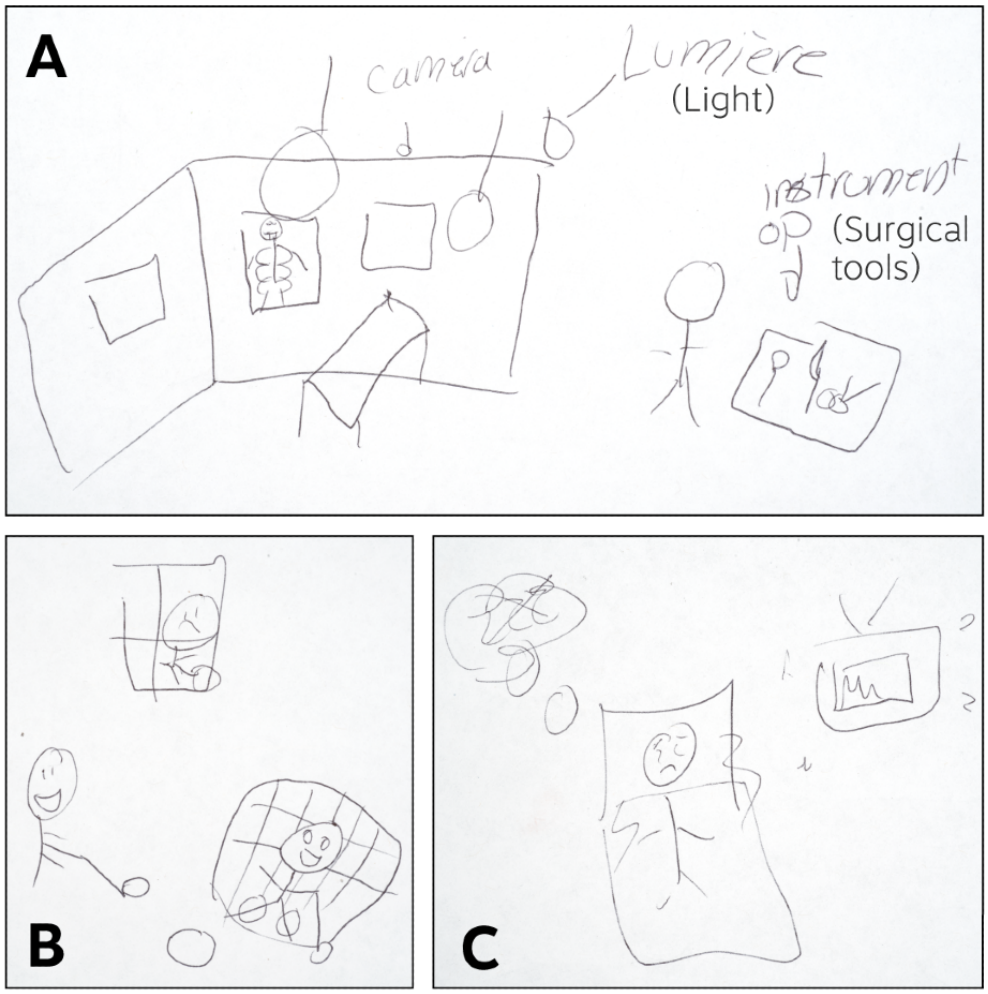
> *Sometimes I say the hospital and I are friends cause OI kids go to the hospital a lot for check-ups and stuff.* [Participant #5, 10 years old]

The new subislands included "The Clinic and The Unit," "In The Operating Room," "The Cast Room," and "The Rehabilitation Room." On this island, children were able to say with certainty about the symptoms that they had never experienced, for instance, "How is it for you to get a feeding tube?" These symptoms that the children had never experienced were removed (Table 4). This table illustrates the symptoms from Sisom that were retained in Sisom OI, that is, deemed "Relevant" or "To Modify" as well as the symptoms that were deemed "Irrelevant" and eliminated by the child feedback cycles overseen by the expert panel. The subisland "About Making Your Own Decisions" was preserved. One child provided vignette suggestions for "How does it feel like when you enter the operating room?", "How do you feel after you wake up after a surgery?", and "How do you feel with a cast?" (Figure 2).

Table 4. Transferability of Sisom symptoms to the Sisom OI paper prototype.

Island	Total number of symptoms	Irrelevant symptoms, n (%)	Retained symptoms, n (%)
Avatar	4	1 (25)	3 (75)
At the Hospital	18	9 (50)	9 (50)
My Body	28	12 (43)	16 (57)
About Managing Things	21	8 (38)	13 (62)
Thoughts and Feelings	10	0 (0)	10 (100)
Things One Might Be Afraid of	8	2 (25)	6 (75)
Total	89	32 (36)	57 (64)

Figure 2. Suggested new symptoms and the corresponding vignettes. A. How is it for you to enter the operating room? Participant #11, 11 years old: "In the room, there are lots of lights. There is a bed and sometimes, you can see the staff looking at their sharp silver instruments and that's stressful for a child. There are also cameras and radiographic images! Those can be really scary." B. How do you feel when you are not able to do certain activities? Participant #11, 11 years old: "Show someone in a cast looking out of the window sadly watching their friends play hockey. It affects you to see your friends happy while you have to stay inside doing nothing. That's it…" C. How is it for you to wake up after a surgery? Participant #11, 11 years old: "Show someone that is not well. Show the machines that beep. See? He's moving. He's in pain. Show someone who wants to sleep but can't.".



My Body

Overall, 57% (16/28) of the "My Body" symptoms were retained. Most revisions and additions reflected the need to capture the children's use of mobility devices, their accessibility issues, and their constipation issues. Some changes also reflected the perception the children shared about the importance of their bones.

> *There should be an 'x-ray' button!* [Participant #7, 9 years old]
>
> *You could show where you often break so that you could explain that's where you are the most fragile.* [Participant #11, 11 years old]

The vignette for the "Pain and Discomfort" subisland was thus modified to allow children to view a skeleton to help them identify the exact bones that caused them discomfort.

About Managing Things

Overall, 62% (13/21) of the "About Managing Things" symptoms were retained. Here, children attributed the greatest importance to the "School Yard" subisland where the symptoms "Do you ever feel left out?" and "Do others ever bully you?" garnered the most attention.

> *Ok. There should definitely be an area for how to handle bullies.* [Participant #5, 10 years old]

One child provided vignette modifications for these 2 symptoms (Figure 3). One child stated:

> *I can tell that people often speak about me and laugh at me (…) No one wants to be with me because of my disease.* [Participant #8, 7 years old]

Another reflected aloud:

> *I get teased a lot. I am starting to realize this is not a world where nice people are (…). There won't always be someone to support you.*

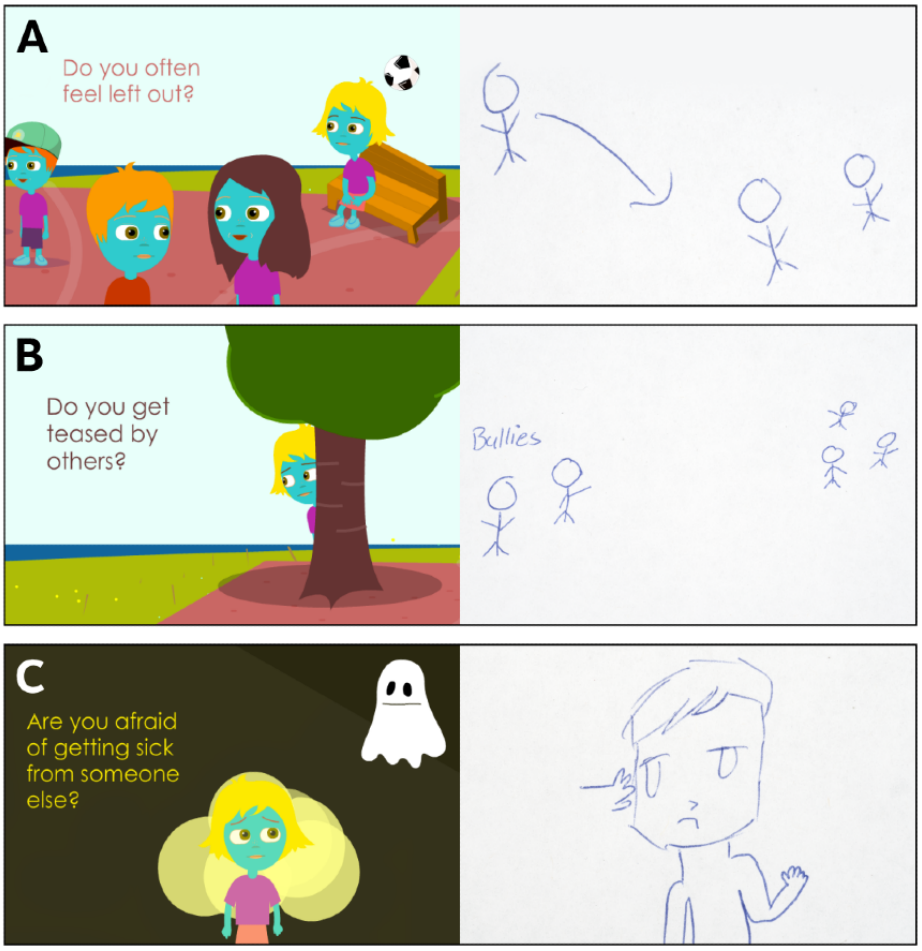
This narrative was common among all participants. Children also shared challenges related to their independence, their social lives, as well as the gazes of others:

> *People give me strange looks at the store and I find that very unpleasant...When I go to the restaurant, people often stare at me. I don't like that.* [Participant #11, 11 years old]

Further additions were made "At Home" and "At School" to reflect accessibility issues:

> *We need a lot of help because of our condition. These questions are very good questions.* [Participant #5, 10 years old]

Figure 3. Suggested modifications to symptoms and the corresponding vignettes. A. "Do you ever feel left out?" changed from "Do you often feel left out?" Participant #5, 10 years old: "Instead of showing someone sitting on a bench, show a person walking up to a group of friends and when that person tries to talk to them, they just go somewhere else. That makes you feel left out..." B. "Do others ever bully you?" changed from "Do you get teased by others?" Participant #5, 10 years old: "Instead of someone hiding behind a tree, there should be bullies teasing and pointing." C. "Are you afraid of getting hurt by someone?" changed from "Are you afraid of getting sick from someone else?" Participant #5, 10 years old: "Show a person sitting down who is being punched".



Thoughts and Feelings

Overall, 100% (10/10) of the "Thoughts And Feelings" symptoms were retained. Feelings of otherness expressed by the children were what drove the changes made to this island. Their perceptions of uniqueness were viewed both positively and negatively. In response to "Do you feel different from the other children?", one child pointed out,

> *Ya but I think it's a good thing. I think that because I'm in a wheelchair I can do things other kids can't do… But being in a wheelchair also has its disadvantages.* [Participant #5, 10 years old]

Further additions reflected children's experiences of having to cope with others' invasive questions about their bodies.

Things One Might Be Afraid Of

Altogether, 75% (6/8) of the "Things One Might Be Afraid Of" symptoms were retained. The main revisions and additions on this island reflected the unique nature of OI as a health condition. One child provided a symptom and the corresponding vignette modification suggestion for "Are you afraid of getting sick from someone else?" to "Are you afraid of getting hurt by someone else?" (Figure 3). Others explained as follows:

> *Overall I'm most scared of fractures.* [Participant #2, 9 years old]

> *When I get a fracture, sometimes it's really bad and I can hear that snap! It makes me a bit shivery just thinking about it…* [Participant #7, 9 years old]

Overall, this was the island that most children visited first and the only island that several children visited more than once.

Discussion

Brief Summary of the Findings

Overall, 64% (57/89) of the Sisom symptoms were deemed relevant for inclusion in Sisom OI, with 42% (37/89) directly incorporated and 22% (20/89) incorporated with changes. The relevant symptoms addressed children's thoughts and feelings about hospitalization and wishes for participation in their own care. In total, 114 symptoms were used to create the prototype among which 57 were newly generated. These new symptoms addressed fractures, body image, and social isolation related to difficulties with accessibility and intimidation. This indicates the need to create Sisom OI to attend to all the specific needs of this population. These new symptoms may also represent the experiences of other children that have similar negative experiences because of societal structures and attitudes premised upon ableism [32].

Children With Chronic Conditions Share Symptoms

The overwhelming number of symptoms that were transferable from Sisom to Sisom OI show that children, regardless of their specific health condition, want to be involved in decisions that affect them, have similar fears within the health care setting, and experience challenges with integration among peers [8,33]. These findings are supported by the literature in which children with other chronic conditions, life-long illnesses, and learning difficulties share similar illness experiences [3,4,34,35]. Some common themes that emerged from qualitative studies with these populations include an aspiration for "normalcy," a life of ups and downs, and changes for the whole family [3,34,36,37]. Other common narratives include a desire to be included and informed, develop assertiveness, gain responsibility, live daily life in stride, and participate in social activities [3,4,34,35,38].

Symptoms as Positive and Negative Experiences

Currently, Sisom assumes that if a symptom is present, it is a source of distress in either a mild, moderate, or severe way. Yet, children explained that the presence of a symptom was not necessarily lived as a disturbance. This finding was also reported in previous Sisom studies in which children emphasized the need for Sisom to use neutral language [14,15,22,25].

In the literature, symptoms experienced by children have been described as "feeling states" lived as a function of context as opposed to isolated measurable sensations assumed to be sources of suffering [37,38]. The children in this study wanted Sisom OI to embrace an outlook in which they could describe their sources of happiness, pride, and support. This discourse is one that is shared by the OI community. By adopting a positive outlook on life, individuals with OI can live their lives to the fullest, despite the many difficult symptoms they experience [39].

Children as Partners in Research and Design

Our participatory approach builds upon how Sisom was originally created [12,40,41]. The principles we applied reflect those described in the Agile Manifesto under which requirements and solutions evolve through the collaborative effort of the end users [42-50]. Clinicians were touched by the children's personal stories about themselves and their relationships, insights about their OI, and deep understanding of their strengths and challenges [1,8,51]. Together, end users agreed that children's experiences of stigma and resulting feelings of otherness must be addressed as early as possible [13]. With the addition of the "About Me" island and the revisions to the "About Managing Things" island, Sisom OI will have the potential to screen for these particular experiences and offer clinicians an opening to address what is of concern to the child.

Limitations

One important factor in any study is the quality of the target group representation during investigations. In this study, we have involved few participants. One criterion that was used to judge that enough data had been collected to conduct an analysis was data saturation, which was reached at the point where a sense of closure was attained because new data from our multitude of sources yielded redundant information [26-28]. Multiple data sources provided an opportunity to evaluate the extent to which a consistent and coherent picture of the content to include in the Sisom OI paper prototype emerged [28]. In the future, this prototype is to be subjected to further verification, validation, and evaluation. In this study, the views of the primary caregivers of these children were not elicited [34,52]. To what extent disruptions of interviews by primary caregivers had an impact on the children's participation remains unknown. Our anecdotal reports suggest that those caregivers who were present expressed interest in using Sisom OI to enhance their children's communication of symptoms. This was a similar finding to those from other Sisom studies [12,14,15,22,24]. Future research will incorporate their views.

Implications for Clinical Practice

This work demonstrates that it is feasible to involve children and clinicians in the creation of software designed for them. Once fully developed, there is potential that the data generated by Sisom OI be collected centrally in order to track trends in the symptoms of an individual, community, or population. Sisom

OI may improve communication between children and their clinicians as encountered in previous Sisom studies and become fully integrated into the practice setting such as in Norway [12,14,15,22,24]. It may also enhance children's participation in their own care by promoting discussions about what these children deem most important. Empowering children by actively involving them in their care may help them to cope with the difficult physical, mental, and social challenges they face. This, in turn, may ease the transfer of self-management skills, ultimately resulting in a greater quality of life.

Conclusion

Interactive software such as Sisom offer a solution to assessing the complex symptoms of children with OI and eliciting their perspectives on their health. The successful Sisom tool that addresses the child directly has the potential to change the communication patterns between children and clinicians and could strengthen children's empowerment [53,54]. Future directions for this work include an inductive analysis of what the children shared when consulted as partners in the development of the Sisom OI paper prototype. It will also be necessary to secure the funds required to create Sisom OI, subject it to further testing, determine incorporation into practice, and evaluate outcomes.

Acknowledgments

We would like to convey our deepest thanks to all the study participants as well as Dr. Cornelia Ruland for her continuous support. We would also like to express our utmost gratitude to Claudette Bilodeau and Khadidja Chougui for their technical assistance in this study. We would like to specially acknowledge Guylaine Bédard for her help with the design of all graphics. This work was generously supported by the Tunis Shriners, Newton Foundation, and Scotiabank®.

Authors' Contributions

Principal investigator AT conceived and designed the study, was responsible for ethics and budget, oversaw recruitment, data collection, and analysis, participated in data analysis and interpretation, and reviewed the manuscript. Clinician FR reviewed the findings, provided clinical expertise, and edited and reviewed the manuscript. Research Trainee MS contributed to the study design, was responsible for recruitment, data collection, analysis, and interpretation, and wrote, edited, and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sisom OI paper prototype.
[DOCX File , 30 KB - jmir_v22i9e17947_app1.docx]

References

1. United Nations. Convention on the Rights of the Child. 1990. URL: http://www.ohchr.org/EN/ProfessionalInterest/Pages/CRC.aspx [accessed 2018-04-18]
2. United Nations. The State of the World's Children. 2003. URL: https://www.unicef.org/sowc/archive/ENGLISH/The%20State%20of%20the%20World%27s%20Children%202003.pdf [accessed 2018-04-18]
3. Alderson P, Sutcliffe K, Curtis K. Children as partners with adults in their medical care. Arch Dis Child 2006 Apr;91(4):300-303 [FREE Full text] [doi: 10.1136/adc.2005.079442] [Medline: 16399782]
4. Sartain S, Clarke C, Heyman R. Hearing the voices of children with chronic illness. J Adv Nurs 2000 Oct;32(4):913-921. [Medline: 11095230]
5. Van Dulmen AM. Children's contributions to pediatric outpatient encounters. Pediatrics 1998 Sep 01;102(3 Pt 1):563-568. [doi: 10.1542/peds.102.3.563] [Medline: 9738177]
6. Söderbäck M, Coyne I, Harder M. The importance of including both a child perspective and the child's perspective within health care settings to provide truly child-centred care. J Child Health Care 2011 Jun 10;15(2):99-106. [doi: 10.1177/1367493510397624] [Medline: 21685225]
7. Kelsey J, Abelson-Mitchell N. Adolescent communication: perceptions and beliefs. Journal of Children's and Young People's Nursing 2007 May;1(1):42-49. [doi: 10.12968/jcyn.2007.1.1.23307]
8. Carnevale FA, Campbell A, Collin-Vézina D, Macdonald ME. Interdisciplinary Studies of Childhood Ethics: Developing a New Field of Inquiry. Child Soc 2013 Oct 04;29(6):511-523. [doi: 10.1111/chso.12063]
9. Peña ALN, Rojas JG. Ethical aspects of children's perceptions of information-giving in care. Nurs Ethics 2014 Mar 23;21(2):245-256. [doi: 10.1177/0969733013484483] [Medline: 23702897]
10. Virkki M, Tolonen T, Koskimaa T. Children as decision-makers in health care: An integrative review. Clinical Nursing Studies 2015;3(1):47 [FREE Full text] [doi: 10.5430/CNS.V3N1P47]

11. Driessnack M. Children's drawings as facilitators of communication: A meta-analysis. J Pediatr Nurs 2005 Dec;20(6):415-423. [doi: 10.1016/j.pedn.2005.03.011] [Medline: 16298282]
12. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. J Biomed Inform 2008 Aug;41(4):624-635 [FREE Full text] [doi: 10.1016/j.jbi.2007.10.004] [Medline: 18082468]
13. Tsimicalis A, Denis-Larocque G, Michalovic A, Lepage C, Williams K, Yao T, et al. The psychosocial experience of individuals living with osteogenesis imperfecta: A mixed-methods systematic review. Qual Life Res 2016 Aug 19;25(8):1877-1896. [doi: 10.1007/s11136-016-1247-0] [Medline: 26894269]
14. Tsimicalis A, Stone PW, Bakken S, Yoon S, Sands S, Porter R, et al. Usability testing of a computerized communication tool in a diverse urban pediatric population. Cancer Nurs 2014;37(6):E25-E34 [FREE Full text] [doi: 10.1097/NCC.0000000000000115] [Medline: 24457227]
15. Arvidsson S, Gilljam B, Nygren J, Ruland CM, Nordby-Bøe T, Svedberg P. Redesign and validation of Sisom, an interactive assessment and communication tool for children with cancer. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e76 [FREE Full text] [doi: 10.2196/mhealth.5715] [Medline: 27343004]
16. Bardai G, Moffatt P, Glorieux FH, Rauch F. DNA sequence analysis in 598 individuals with a clinical diagnosis of osteogenesis imperfecta: Diagnostic yield and mutation spectrum. Osteoporos Int 2016 Dec;27(12):3607-3613. [doi: 10.1007/s00198-016-3709-1] [Medline: 27509835]
17. Dahan-Oliel N, Oliel S, Tsimicalis A, Montpetit K, Rauch F, Dogba MJ. Quality of life in osteogenesis imperfecta: A mixed-methods systematic review. Am J Med Genet A 2016 Jan 14;170A(1):62-76. [doi: 10.1002/ajmg.a.37377] [Medline: 26365089]
18. Trejo P, Rauch F. Osteogenesis imperfecta in children and adolescents: New developments in diagnosis and treatment. Osteoporos Int 2016 Dec 5;27(12):3427-3437. [doi: 10.1007/s00198-016-3723-3] [Medline: 27492436]
19. Shapiro JE, Germain-Lee EL. Osteogenesis imperfecta: Effecting the transition from adolescent to adult medical care. J Musculoskelet Neuronal Interact 2012 Mar;12(1):24-27 [FREE Full text] [Medline: 22373948]
20. Fegran L, Hall EO, Uhrenfeldt L, Aagaard H, Ludvigsen MS. Adolescents' and young adults' transition experiences when transferring from paediatric to adult care: A qualitative metasynthesis. Int J Nurs Stud 2014 Jan;51(1):123-135. [doi: 10.1016/j.ijnurstu.2013.02.001] [Medline: 23490470]
21. Nghiem T, Louli J, Treherne SC, Anderson CE, Tsimicalis A, Lalloo C, et al. Pain experiences of children and adolescents with osteogenesis imperfecta: An integrative review. Clin J Pain 2017 Mar;33(3):271-280. [doi: 10.1097/AJP.0000000000000395] [Medline: 27258996]
22. Tsimicalis A, Le May S, Stinson J, Rennick J, Vachon M, Louli J, et al. Linguistic validation of an interactive communication tool to help French-speaking children express their cancer symptoms. J Pediatr Oncol Nurs 2017;34(2):98-105. [doi: 10.1177/1043454216646532] [Medline: 27226380]
23. Tsimicalis A, Stone PW, Bakken S, Yoon S, Sands S, Porter R, et al. Usability testing of a computerized communication tool in a diverse urban pediatric population. Cancer Nurs 2014;37(6):E25-E34 [FREE Full text] [doi: 10.1097/NCC.0000000000000115] [Medline: 24457227]
24. Baggott C, Baird J, Hinds P, Ruland CM, Miaskowski C. Evaluation of Sisom: A computer-based animated tool to elicit symptoms and psychosocial concerns from children with cancer. Eur J Oncol Nurs 2015 Aug;19(4):359-369. [doi: 10.1016/j.ejon.2015.01.006] [Medline: 25691300]
25. Tsimicalis A, Rennick J, Le May S, Stinson J, Sarkis B, Séguin K, et al. “Tell it as it is”: How Sisom prompts children and parents to discuss their cancer experience. Cancer Reports 2019 Apr 23;2(3):1-15. [doi: 10.1002/cnr2.1173]
26. Jakob Nielsen. Nielsen Norman Group World Leaders in Research-Based User Experience. Why You Only Need to Test with 5 Users. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2018-04-18]
27. Nielsen JT, Landauer TK. A Mathematical Model of the Finding of Usability Problems. 1993 Presented at: Proceedings of the INTERACT 93 and CHI 93 Conference on Human Factors in Computing Systems; 1993; Amsterdam p. 24-29. [doi: 10.1145/169059.169166]
28. Polit DC, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice. In: Nursing Research: Generating and Assessing Evidence for Nursing Practice. Hong Kong: Wolters Kluwer Health - Lippincott Williams and Wilkins; 2017.
29. Thorne S. Data analysis in qualitative research. Evid Based Nurs 2000;3(3):68-70. [doi: 10.1136/ebn.3.3.68]
30. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. In: Qualitative Data Analysis: An Expanded Sourcebook. London: Sage Publications Incorporated; 1994.
31. Rodgers BL, Cowles KV. The qualitative research audit trail: A complex collection of documentation. Res Nurs Health 1993 Jun;16(3):219-226. [doi: 10.1002/nur.4770160309] [Medline: 8497674]
32. Policy on Ableism and Discrimination Based on Disability. Ontario Human Rights Commission. URL: http://www.ohrc.on.ca/en/policy-ableism-and-discrimination-based-disability [accessed 2018-04-18]
33. King G, Batorowicz B, Rigby P, Pinto M, Thompson L, Goh F. The leisure activity settings and experiences of youth with severe disabilities. Dev Neurorehabil 2014 Aug 19;17(4):259-269. [doi: 10.3109/17518423.2013.799244] [Medline: 23869878]

34. Garth B, Aroni R. 'I Value What You have to Say'. Seeking the Perspective of Children with a Disability, Not Just their Parents. Disability & Society 2003 Aug;18(5):561-576. [doi: 10.1080/0968759032000097825]
35. Livesley J, Long T. Children's experiences as hospital in-patients: Voice, competence and work. Messages for nursing from a critical ethnographic study. Int J Nurs Stud 2013 Oct;50(10):1292-1303. [doi: 10.1016/j.ijnurstu.2012.12.005] [Medline: 23332686]
36. Dogba MJ, Bedos C, Durigova M, Montpetit K, Wong T, Glorieux FH, et al. The impact of severe osteogenesis imperfecta on the lives of young patients and their parents: A qualitative analysis. BMC Pediatr 2013 Sep 30;13(1):153 [FREE Full text] [doi: 10.1186/1471-2431-13-153] [Medline: 24074180]
37. Woodgate RL, Degner LF, Yanofsky R. A different perspective to approaching cancer symptoms in children. Journal of Pain and Symptom Management 2003 Sep;26(3):800-817. [doi: 10.1016/s0885-3924(03)00285-9]
38. Wyness M. Adults' involvement in children's participation: Juggling children's places and spaces. Children and Society 2009;23(6):395-406. [doi: 10.1111/j.1099-0860.2008.00181.x]
39. Care 4 Brittle Bones Ambassadors. Care 4 Brittle Bones. URL: http://www.care4brittlebones.org/en/care4brittlebones/who-is-who/ambassadors/ [accessed 2018-04-18]
40. Ruland C. Handheld technology to improve patient care: Evaluating a support system for preference-based care planning at the bedside. J Am Med Inform Assoc 2002;9(2):192-201 [FREE Full text] [doi: 10.1197/jamia.m0891] [Medline: 11861634]
41. Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of a computerized system to support shared decision making in symptom management of cancer patients: Preliminary results. J Am Med Inform Assoc 2003;10(6):573-579 [FREE Full text] [doi: 10.1197/jamia.M1365] [Medline: 12925545]
42. Beck K, Beedle M, Van BA. 2001. Manifesto for Agile Software Development. URL: https://agilemanifesto.org/ [accessed 2020-04-02]
43. Steele GC, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: Adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
44. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]
45. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: Application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
46. The CHAOS Report. The Standish Group. 1994. URL: https://www.standishgroup.com/sample_research_files/chaos_report_1994.pdf [accessed 2020-04-02]
47. Bate P, Robert G. Experience-based design: From redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
48. Boyd H, McKernon S, Mullin B, Old A. Improving healthcare through the use of co-design. N Z Med J 2012 Jun 29;125(1357):76-87. [Medline: 22854362]
49. Christensen T. The evolution of patient-centered care and the meaning of co-design. Institute for Healthcare Improvement. 2017. URL: http://www.ihi.org/communities/blogs/evolution-of-patient-centered-care-and-the-meaning-of-co-design [accessed 2018-04-18]
50. Chokshi SK, Mann DM. Innovating from within: A process model for user-centered digital development in academic medical centers. JMIR Hum Factors 2018 Dec 19;5(4):e11048 [FREE Full text] [doi: 10.2196/11048] [Medline: 30567688]
51. Davis JM. Understanding the meanings of children: a reflexive process. Children & Society 1998 Nov;12(5):325-335. [doi: 10.1111/j.1099-0860.1998.tb00089.x]
52. Gannoni AF, Shute RH. Parental and child perspectives on adaptation to childhood chronic illness: A qualitative study. Clin Child Psychol Psychiatry 2010 Jan;15(1):39-53. [doi: 10.1177/1359104509338432] [Medline: 19914942]
53. McNaughton D, Light J. The iPad and mobile technology revolution: Benefits and challenges for individuals who require augmentative and alternative communication. Augment Altern Commun 2013 Jun;29(2):107-116. [doi: 10.3109/07434618.2013.784930] [Medline: 23705813]
54. Raaff C, Glazebrook C, Wharrad H. A systematic review of interactive multimedia interventions to promote children's communication with health professionals: Implications for communicating with overweight children. BMC Med Inform Decis Mak 2014 Jan 22;14:8 [FREE Full text] [doi: 10.1186/1472-6947-14-8] [Medline: 24447844]

Abbreviations

OI: osteogenesis imperfecta

Edited by G Eysenbach; submitted 27.01.20; peer-reviewed by M Grant, C Reis; comments to author 23.03.20; revised version received 05.04.20; accepted 10.04.20; published 22.09.20.

Please cite as:
Siedlikowski M, Rauch F, Tsimicalis A
Giving Children With Osteogenesis Imperfecta a Voice: Participatory Approach for the Development of the Interactive Assessment and Communication Tool Sisom OI
J Med Internet Res 2020;22(9):e17947
URL: http://www.jmir.org/2020/9/e17947/
doi:10.2196/17947
PMID:32960176

Original Paper

Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies: Qualitative Mixed-Method Study of Military and Civilian Cases

Laurie Lovett Novak[1], MHSA, PhD; Christopher L Simpson[1], MA; Joseph Coco[1], MS; Candace D McNaughton[2,3], MD, MPH, PhD; Jesse M Ehrenfeld[4,5], MD, MPH; Sean M Bloos[2], BS; Daniel Fabbri[1], PhD

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Geriatric Research Education and Clinical Center, Tennessee Valley Healthcare System VA Medical Center, Nashville, TN, United States

[4]Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, United States

[5]Department of Anesthesiology, Medical College of Wisconsin, Milwaukee, WI, United States

Corresponding Author:
Laurie Lovett Novak, MHSA, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Avenue, Suite 1475
Nashville, TN, 37203
United States
Phone: 1 615 936 6497
Email: laurie.l.novak@vumc.org

Abstract

Background: Current methods of communication between the point of injury and receiving medical facilities rely on verbal communication, supported by brief notes and the memory of the field medic. This communication can be made more complete and reliable with technologies that automatically document the actions of field medics. However, designing state-of-the-art technology for military field personnel and civilian first responders is challenging due to the barriers researchers face in accessing the environment and understanding situated actions and cognitive models employed in the field.

Objective: To identify design insights for an automated sensing clinical documentation (ASCD) system, we sought to understand what information is transferred in trauma cases between prehospital and hospital personnel, and what contextual factors influence the collection, management, and handover of information in trauma cases, in both military and civilian cases.

Methods: Using a multi-method approach including video review and focus groups, we developed an understanding of the information needs of trauma handoffs and the context of field documentation to inform the design of an automated sensing documentation system that uses wearables, cameras, and environmental sensors to passively infer clinical activity and automatically produce documentation.

Results: Comparing military and civilian trauma documentation and handoff, we found similarities in the types of data collected and the prioritization of information. We found that military environments involved many more contextual factors that have implications for design, such as the physical environment (eg, heat, lack of lighting, lack of power) and the potential for active combat and triage, creating additional complexity.

Conclusions: An ineffectiveness of communication is evident in both the civilian and military worlds. We used multiple methods of inquiry to study the information needs of trauma care and handoff, and the context of medical work in the field. Our findings informed the design and evaluation of an automated documentation tool. The data illustrated the need for more accurate recordkeeping, specifically temporal aspects, during transportation, and characterized the environment in which field testing of the developed tool will take place. The employment of a systems perspective in this project produced design insights that our team would not have identified otherwise. These insights created exciting and interesting challenges for the technical team to resolve.

(J Med Internet Res 2020;22(9):e17978) doi:10.2196/17978

KEYWORDS

trauma handoffs; military field medicine; documentation; trauma; health records; hospital; emergency

Introduction

Optimizing Communication Processes During Trauma Handoffs

When military personnel or civilians are injured, field medics are the first to respond. Their objectives include stabilizing and transporting the patient to a trauma facility. Optimizing patient outcomes depends on accurate information sharing between field personnel and receiving physicians, including the context of the injury and clinical interventions [1]. These patient and information transfers in combat settings are highly variable and can range from minimal communication (eg, pointing to a limb with a tourniquet when a helicopter picks up a patient from a scene in hostile territory) to verbal handoff when the patient is transported directly to the next level of care. When written documentation is generated, the documentation process may distract vital cognitive efforts away from patient care. Moreover, both written and verbal communication methods are vulnerable to rapid changes in clinical status, human cognitive biases, and mistakes in data collection, processing, and sharing. As a result, the information may be incomplete, inaccurate, or lost in communication [1-3]. Multiple handoffs further complicate the process and likely increase the risk of errors and miscommunication during transport.

Timely and accurate clinical documentation occurs when a sociotechnical system is designed and optimized around the relevant people, tasks, technologies, and physical and social environments [4]. Challenges include time pressure, the unique stress of providing care in combat situations, the use of personal protective equipment, limited visibility, and constrained working spaces. In addition, even when documentation is generated, it is rarely transmitted in a way that is timely, clear, or effective [5]. Given the challenges of using traditional technologies to document clinical care at the point of injury and during transport, new systems are needed that can ensure better, more consistent, and clearer communication among care teams.

Our main research effort is to develop novel technologies to automatically generate a clinical care record without requiring the active participation of personnel in the field [6]. This automated sensing clinical documentation (ASCD) technology observes the tasks the medic performs using a combination of sensors [7]. During its observation, the system outputs the list of clinical procedures that are being performed, ideally with high accuracy.

Designing ASCD involves understanding the other elements of the sociotechnical system into which the ASCD must fit [8]. These elements include information the system must capture and the social and physical contexts in which it will be deployed. Direct assessment of the current state of military trauma handoffs is impractical due to safety and logistical concerns [9]. Relying on civilian ambulance observations produces data from a limited number of trauma cases, typically in an environment that is unlike a military field operation. Therefore, through a multi-modal analysis that includes focus groups and trauma-bay video review, this paper analyzes current trauma handoff practices to categorize information needs and contextual factors involved in trauma handoffs.

Background

The overall objective of our project is to develop an ASCD system that can be used on the battlefield by military personnel or by civilian medics in the field. The technology will involve a combination of off-the-shelf sensors, accelerometers, and cameras aligned with a software system that automatically detects the motion signatures associated with key clinical tasks and generates an abbreviated care record, which can be transmitted upstream in real time. The system will passively collect data from a combination of accelerometers and cameras. Machine learning, activity detection, and summarization algorithms will analyze the collected data to construct an abbreviated care record. This care record will provide patient clinical status, interventions, and anticipated resources needed upon arrival, without requiring active input from personnel in the field. Open research challenges to building such documentation systems include the accuracy of predicting clinical events, usability, and deployment robustness.

In the US conflicts in Iraq and Afghanistan, the nation has suffered a total death toll of 4432 and 2351 deaths, respectively, as of December 4, 2019 [10]. Since many fatalities occur between the point of injury and the medical treatment facility (MTF), the military has incorporated the use of Tactical Combat Casualty Care (TCCC) cards to document the mechanism of injury, injury locations, vital signs and symptoms, and treatments [11-13]. This allows the first responders to triage the most critical patients in the prehospital (eg, battlefield, vehicle) environment [12,13]. The military's documentation of the treatment during this period “is critical to ensuring continuity of care” [14]. After completing the card, the first responder attaches the TCCC card to the patient in a visible location as the record of provided treatment. Medical personnel in the receiving MTF are instructed to include the TCCC card with the paper medical record and to enter the TCCC data into the patient's electronic health record and appropriate trauma registry. Despite some evidence of a lack of compliance with the policy, the Defense Health Agency states, “The military will continue to use the TCCC card until it fields a prehospital documentation platform that supports an electronic version” [14].

The transfer of a patient from a field medic to an MTF is a handover, defined as “the transfer of professional responsibility and accountability for some or all aspects of care for a patient, or group of patients, to another person or professional group on a temporary or permanent basis” [15]. In this paper, we use the term “handoff” with the same definition. Handoffs in health care have received significant attention in recent years as a period of high risk for the patient's safety. In comparison with

handovers in combat conditions, civilian handoffs are characterized by safe working conditions for the involved clinicians and an indoor Emergency Room physical setting with controlled lighting, climate, and (typically) noise. Additionally, trauma patients in civilian settings often undergo only one or two handoffs to reach the final hospital destination, whereas a patient in a combat situation may experience handoffs in several stages between the battlefield and the hospital. Despite these conditions, a review found that in civilian handoffs between medics and hospital-based emergency departments, the key issues were a lack of common understanding, lack of active listening, variable quality and quantity of information exchanged, lack of clear leadership, lack of teamwork skills, busy and complex environment, and handoff repetition [16]. Organizations have tried to resolve issues with handoffs through interventions to standardize communications, with mixed results [17,18]. Our project uses a systems perspective to examine an understudied topic that is especially challenging in military medicine: the capture of clinical documentation in the field, especially in battle conditions.

Findings in this paper are organized with a health care systems engineering model that has been extensively used in the study of both handoffs [19] and information technologies [20,21]. The Systems Engineering Initiative for Patient Safety (SEIPS) is a systems approach for understanding human activity in its context [22]. The fields of human factors and industrial engineering spurred the development of the framework to help frame research and innovation as technology was introduced into all areas of health care. The model was subsequently extended as SEIPS 2.0 to incorporate patient engagement, patient work, and work practice adaptations [23].

Methods

Research Questions

The research questions guiding this work were the following: (1) What information is transferred in trauma cases between prehospital and hospital personnel? (2) What contextual factors influence the collection, management, and handover of information in trauma cases?

Research Site

Vanderbilt University Hospital provides trauma care for 65,000 square miles. The Division of Trauma at Vanderbilt University Hospital handles close to 5800 acute traumas, admitting 4300 of those annually. Its facilities are essential for the quality of trauma care provided by Vanderbilt University Hospital. These include a 20-bed burn unit and a 31-bed integrated acute and subacute care unit, which contains a 14-bed intensive care unit, a 7-bed acute admission area, and a 10-bed subacute unit, as well as LifeFlight, which is an active air medical transport program. The trauma units' unique geography allows close integration and management of patient progress from admission to discharge. LifeFlight provides rapid access to the tertiary care facilities of the Trauma Center for all patients within a 140-mile radius of Nashville. In addition to LifeFlight, Vanderbilt receives patient transport from local and rural emergency medical services (EMS).

Research Approaches

Data Sources

Our methods included (1) a structured review of routinely captured videos of trauma handoffs in the Vanderbilt University Medical Center (VUMC) Emergency Department (ED), and (2) focus groups with ED providers, prehospital personnel (such as emergency medical technicians and paramedics), and military field medics. The research was conducted at VUMC and the Army's Rascon School of Combat Medicine on Fort Campbell, Kentucky.

The study protocols were reviewed and approved by the Vanderbilt University Institutional Review Board. Given the infeasibility of observational research of the activities of front-line military medical personnel, we used triangulation of data [24] from 2 different methods to gather information about the work of field medics (also referred to as prehospital personnel) and the handoffs between prehospital and hospital personnel.

Trauma Video Reviews

VUMC level 1 trauma cases are recorded for quality improvement purposes and reviewed weekly. These videos capture the pre-brief (in which EMS personnel and trauma team members from the ED and trauma team review facts about the arriving case and discuss a plan of action) and treatment while in the ED trauma bay. We reviewed 50 randomly selected videos to identify information transmitted via conversations during the handoff from EMS to hospital personnel. Videos are stored with no identifying linkages to patients and are deleted after a specified period. The videos were a way for us to observe handoffs without any patient-identifying information being collected.

A structured form facilitated the collection of relevant data from the videos. In order to refine a preliminary data collection form for the reviews, 5 videos were reviewed and documented by 3 reviewers. After the videos were reviewed, discussion of the results and any discrepancies in documentation were moderated by an independent arbiter. The reviewers came to a consensus on the types of information transferred from prehospital to hospital personnel and developed a data collection form to be used by 1 expert observer. The observer, a registered nurse, has extensive experience in trauma nursing and experience with the review of handoff videos. This observer viewed 50 trauma handoff videos and recorded observations on the forms. After completion of the reviews, the data from the observation forms were entered into a REDCap [25] database (Vanderbilt University) for analysis and tabulation.

Focus Groups

We conducted four 60-90-minute focus groups, each led by a team of 2 anthropologists (LN and CS) experienced in qualitative research. The focus group leaders had no prior relationship with the participants, although some participants had interacted with other research team members previously. The leaders had only general prior knowledge of the activities and contexts discussed in the focus groups and no personal experiences that would introduce bias.

Of the 4 focus groups that were conducted, 2 included civilian prehospital personnel (ambulance-based medics and aircraft-based flight medics), 1 included hospital personnel (physicians), and 1 was conducted with military combat medics. For the civilian focus groups, participants were recruited through email-based snowball sampling; for the military group, participants were recruited through a convenience sample of personnel available on the date of the study. Other members of the research team were present during the focus groups. There was no subsequent contact between the research team and the participants.

The goals of the focus groups were to gather information from providers and medics with trauma experience to (1) identify information transmitted in handoffs, (2) identify gaps in current handoff procedures, and (3) understand the social and physical context into which the technology will be deployed. These goals were communicated to the participants of each group. The sessions explored participant experience with the transportation of patients to the hospital, including the elicitation of actual experiences in a combat environment when possible. Questions posed during the focus groups included the following: (1) What information is normally shared during handoffs? (2) What information is most useful to determine the next steps of care management? (3) Why and how is this information shared? (4) What information is not useful to determine care management?

Based on the information shared in the session, we added probes to better understand the physical actions involved in transporting patients from the field or scene to the hospital, including the implications of incorporating wearable technologies, cameras, and other devices into the process.

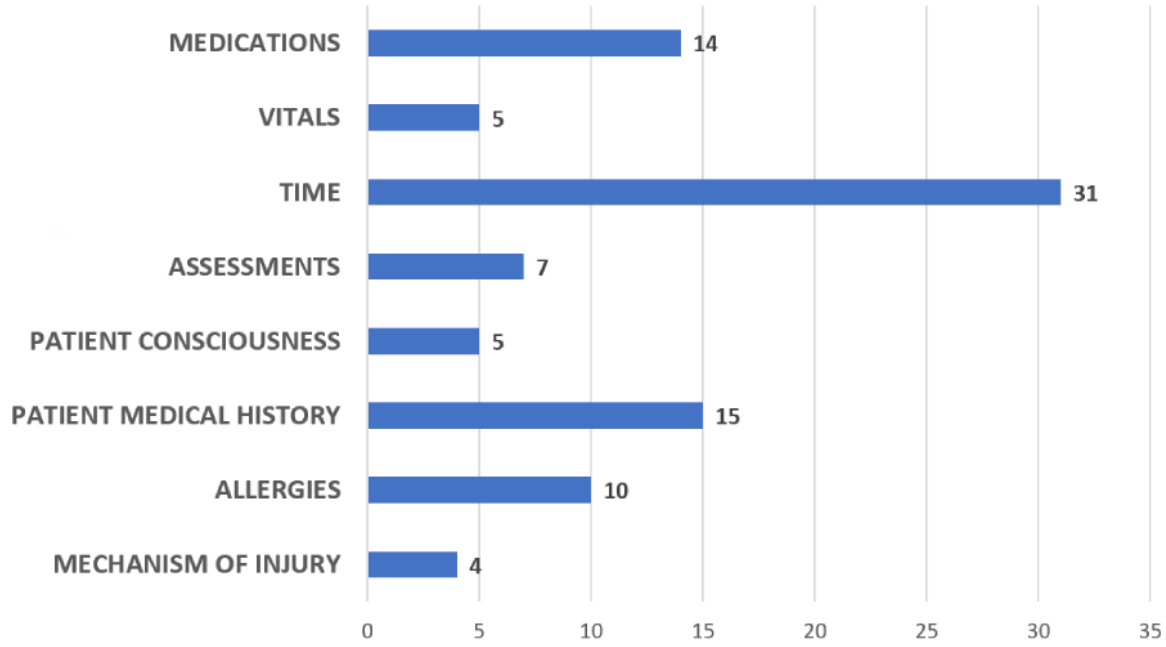
The sessions were audio-recorded and transcribed for analysis. The transcripts were analyzed using a qualitative data analysis tool, Dedoose (SocioCultural Research Consultants), which facilitates the selection of text excerpts and labeling with one or more codes by a human coder and displays a variety of summaries of the coded data. Given the variety of information shared by participants on information needs and context, we used an open coding procedure, identifying all themes that arose in the data. There were 3 researchers who coded the data, supported by discussions in frequent team meetings about findings and the organization of the data. Then the data was organized using the SEIPS 2.0 model for presentation and consideration by the team's technology designers.

Results

Trauma Video Reviews

The handoff videos revealed information that is routinely relayed to the hospital team from the prehospital team. Figure 1 describes the content of each category of information in the 50 videos. Categories of information included clarifying questions asked by the receiving medical team, procedures performed, mechanism of injury, medications and fluids given during transport, time of intervention and injury, blood pressure, heart rate, respiratory rate, oxygen saturation, and episodes of hypotension changes in clinical status.

Figure 1. Content of clarifying questions in the trauma videos.



Upon analysis of the data, it became apparent that clarifying questions were an important part of the prehospital-to-trauma team handoff. Clarifying questions are defined as questions from the hospital team directed to the prehospital team during handoff that are intended to obtain additional information that was not provided in the initial handoff. Of the 50 videos reviewed, 40 (80%) contained clarifying questions.

The clarifying questions that we observed in the videos consisted of questions about medication (eg, dosages, timing), personal medical history (if known), Glasgow Coma Scale or other (mostly neurological) exam results, time and mechanism of injury, allergies, whether or not restraints were used in accidents in vehicles, length of time tourniquets have been in place, and

fluctuations in vitals or neurological signs (blood pressure, heart rate, respiratory rate, oxygen saturation, etc).

The results for the other categories of information captured during observations are shown in Figures 2-5.

Figure 2. Frequency of procedures performed during transport in the trauma videos.

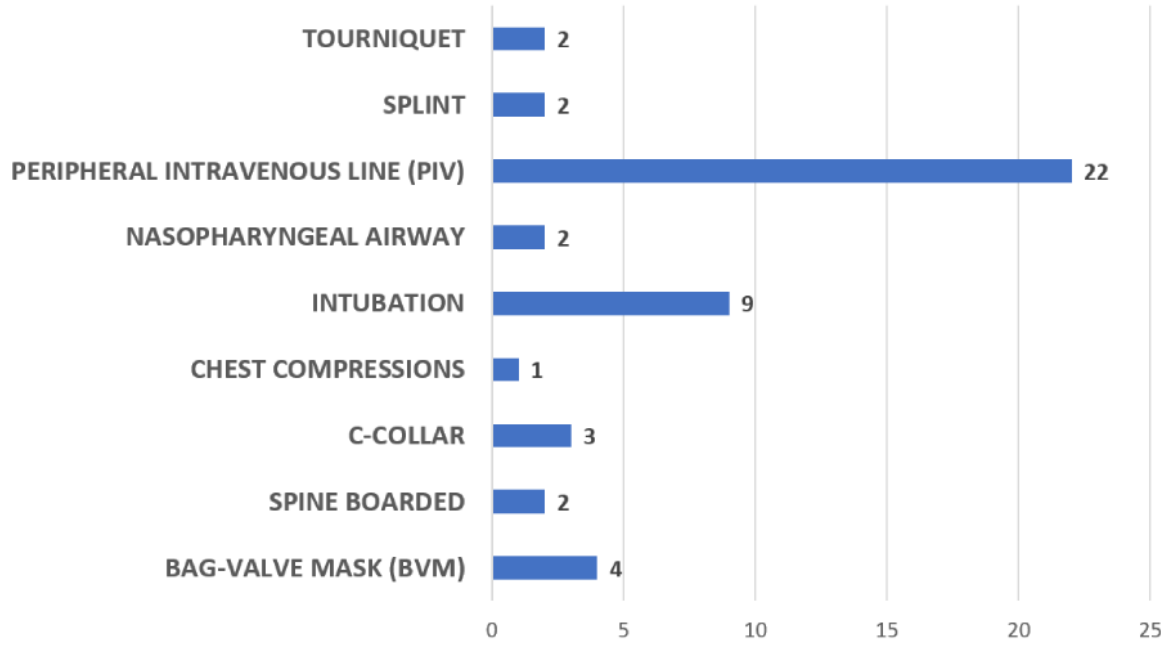


Figure 3. Frequency of each mechanism of injury in the trauma videos.

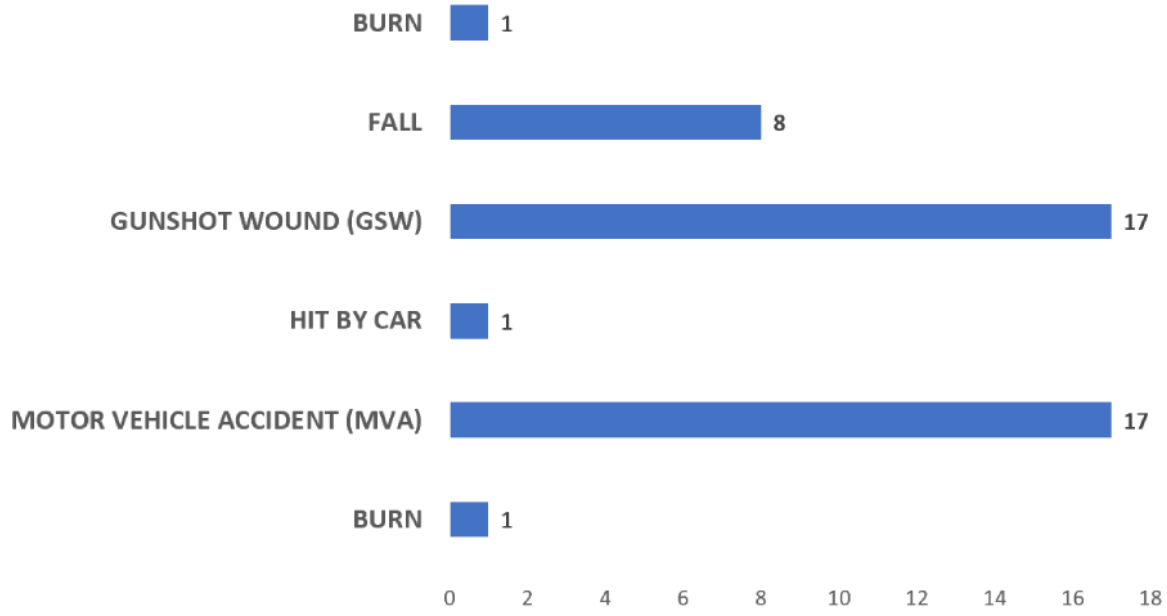


Figure 4. Medications and fluids administered during transport in the trauma videos.

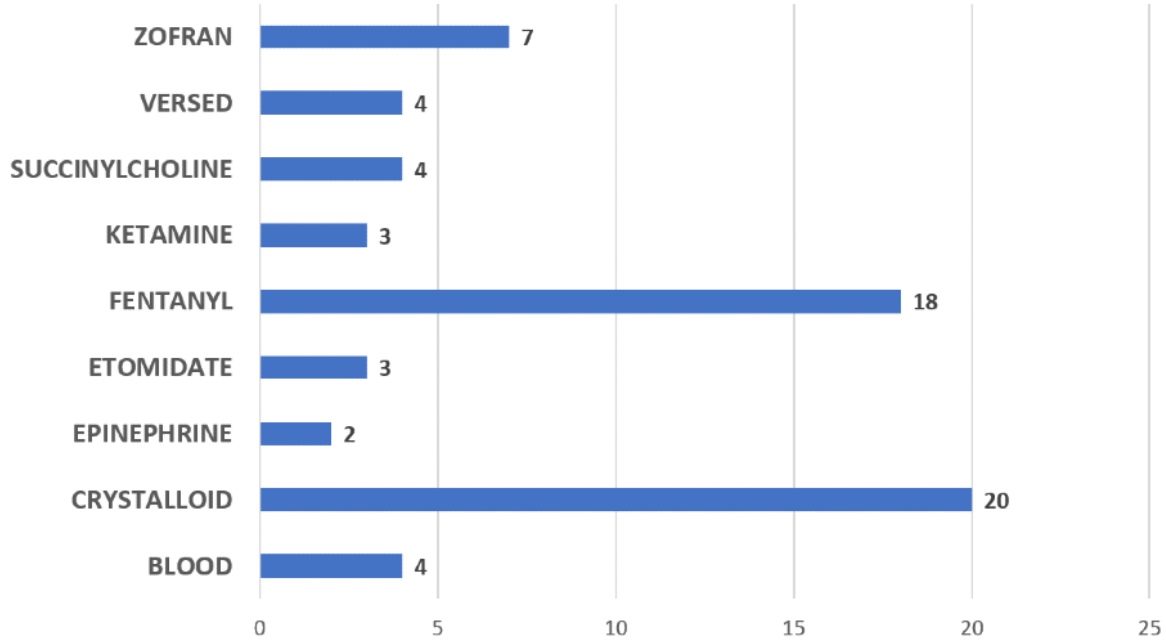
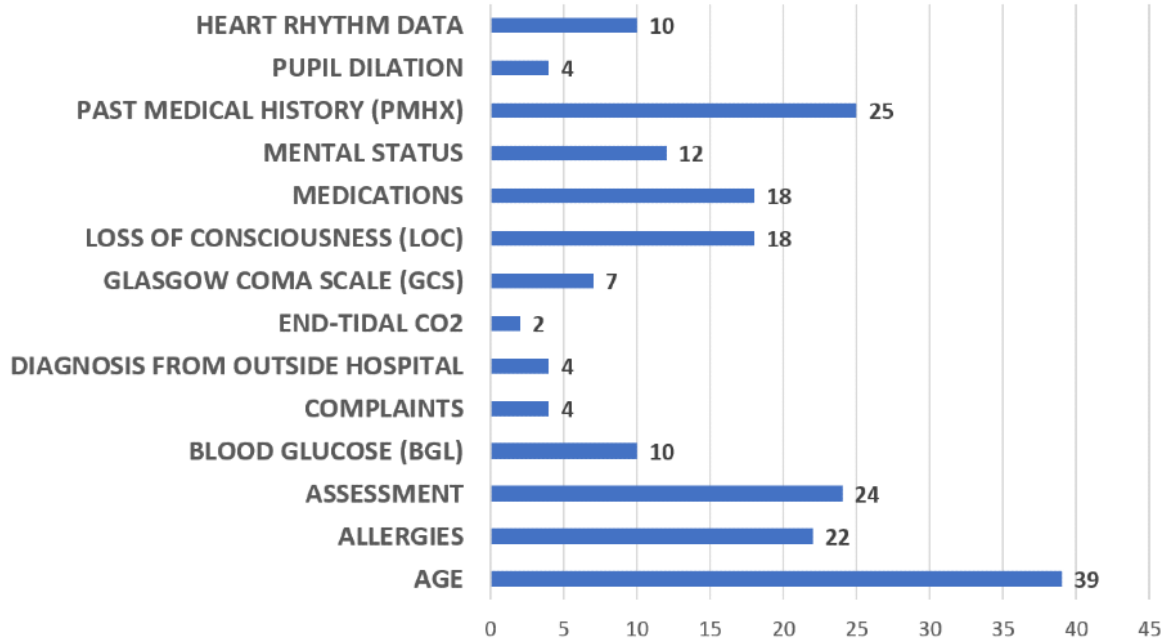


Figure 5. Other handoff information reported in the trauma videos.



Focus Groups

We conducted 4 focus groups comprising 19 participants, with no participants dropping out of the study. Participants included prehospital personnel (ambulance-based medics and aircraft-based flight medics), hospital personnel (physicians), and military medical personnel. Findings, comprising a comparison of the military and civilian experiences, are summarized using the SEIPS framework in Table 1.

Table 1. Work system analysis for documentation in the field.

Feature	Civilian prehospital system	Military prehospital system	Insights for the development of hardware and software tools
Technology and tools			
	Written or electronic documentation of prehospital care	TCCC[a] (universal documentation card), sometimes partially completed by service member prior to mission	*Ad hoc use of tools, determined by the setting and characteristics of the patient
	Gloves, paper, tape (used for recording written information)	Communication headsets	*Information transmitted in advance could help hospital allocate resources (physician group)
	Vital signs monitoring technology	Medics carry medical gear and combat gear	*In civilian setting, a simple statistic representing level of medic activity could potentially give early indication of severity of patient injury (physician group)
			*Mounting a camera in the vehicle is a challenge due to privacy issues in civilian setting
			*Object detection algorithms could potentially detect specific medication packages; however, packaging not currently standardized in military or civilian setting
Tasks			
	Documentation: vital signs, demographics, medications, allergies, time of events, procedures, pain level, mechanism of injury	Documentation: vital signs, procedures, mechanism of injury	Priority information for handoff:
			*Timing and sequence of procedures can suggest cause and effect
			*Worst and most recent vital signs are most useful
	Procedures	Procedures	*In military setting, trends in data were more useful
	Radio communication	Radio communication	
	N/A[b]	Triage	
	N/A	Active combat activities	
	N/A	Maintaining tactical awareness	
Organization			
	Information systems in the EMS[c] vehicle did not communicate with the hospital emergency department.	Large-scale, contracted military technology implementations sometimes lack coordination in technology updates, resulting in lost communication between system components.	Transmitting information to hospital can reduce miscommunication, but also result in information overload.
Physical environment			
	Extreme heat is common	Extreme heat is common, exacerbated by excessive gear	*Need lightweight, small sensors; armbands will be hot and uncomfortable
	N/A	Dusty	*Voice technology not feasible because of noise
	N/A	High noise level in all settings	*Sensors should conserve power when not in use
	N/A	Note-taking is difficult	*Wearable devices must withstand a substantial amount of sweat from the wearer
	N/A	Rough terrain/unstable vehicle	*Devices must be physically durable
	N/A	Low light in combat settings	

[a]TCCC: Tactical Combat Casualty Care.

[b]N/A: not applicable.

[c]EMS: emergency medical services.

Discussion

Principal Findings

Findings from the videos illustrated that the most medically important information is not always effectively conveyed during the handoff from prehospital to hospital personnel. Of note were the clarifying questions observed during the review of the videos of the handoffs. Clarifying questions were observed in 80% (40/50) of recorded handoffs and most commonly involved temporal aspects of the case. Temporal questions included queries about the time the injury occurred, when a procedure was performed, and when a medication was given. Questions related to timing (eg, when medications were administered) were present in 27 of the 40 videos in which clarifying questions were asked of the prehospital staff. The next most commonly

asked clarifying question involved either medications (drugs given, doses, timing, etc) or the patient's past medical history. Both types of questions were present in 13 of the 40 videos in which clarifying questions were asked during the handoff.

Data from the observations support the findings from the 3 focus groups that more accurate information is needed at the time of handoff, specifically regarding time and sequences of procedures and medications. The hospital focus group emphasized that the most important information needed by the trauma team involved the timing of events, especially regarding the sequence of procedures performed during transport. The trauma videos revealed mechanisms of injury that would be less common in military environments (eg, falls and being hit by a motor vehicle). However, we note that it is difficult to speculate on what types of trauma injuries may be seen in future combat situations, and it is likely short-sighted to design only for wounds produced by gunshots or explosions.

The prehospital and hospital teams have different priorities and capabilities in the performance of their roles in their respective environments. Prehospital teams need to get the patient into the vehicle and perform needed procedures during transport to get the patient to superiorly resourced care teams, often geared toward surgical intervention. Meanwhile, the receiving trauma team wants to be able to appropriately allocate resources based on procedures performed and patient trajectory during transport. These differences result in an inadvertent conflict about the priority of recording specific times of medication administration and the performance and sequence of procedures during transport.

The findings from the video review and focus groups produced insights that informed device choices, software development, and evaluation strategy. Some surveillance technologies, such as microphones that could potentially be useful to support documentation, are not practical for noisy and insecure military field settings. While no tool will be able to capture every aspect of prehospital care, documentation through automated sensing can potentially enable medics to offer a more complete handoff to the receiving hospital.

Implications for Design

Various activities are detectable through sensors. We identified numerous opportunities to capture activity (such as medical procedures or administration of medications) through motion detection and the relationship of motion signatures to locations on the patient's body, as well as the use of physical artifacts such as medication packaging. However, there is heterogeneity in how procedures are performed and noise in the data. A robust system of data collection and analysis will be needed to deal with the forces of real-world deployments. Challenges such as vehicle motion and sensor failure due to the environment (eg, a wearable sensor exposed to extensive sweat) may be universal. Challenges specific to military environments include the lack of lighting, a high possibility of network failure, and the possibility of active battle conditions while treatment is being carried out.

Conclusion

An ineffectiveness of communication is evident in both the civilian and military worlds. We used multiple methods of inquiry to study the information needs of trauma care and handoff, and the context of medical work in the field. Our findings informed the design and evaluation of an automated documentation tool. The data illustrated the need for more accurate recordkeeping, specifically temporal aspects, during transportation, and characterized the environment in which field testing of the developed tool will take place. Solutions will need to address the environmental constraints of low lighting, heat, dust, noise, and vehicle instability. In addition, sensor power conservation is critical in field combat settings. The employment of a systems perspective in this project produced design insights that our team would not have identified otherwise. These insights created exciting and interesting challenges for the technical team to resolve.

Acknowledgments

This work was supported by the Department of Defense Contract Number W81XWH-17-C-0252 from the CDMRP Defense Medical Research and Development Program. We would also like to thank the participants in our study for the generous contribution of their time and insight. Finally, we are grateful to the Vanderbilt University Medical Center Department of Emergency Medicine, and specifically Kevin High and Sally Dye, for their assistance with the video analysis project.

Conflicts of Interest

None declared.

References

1. Carter AJE, Davis KA, Evans LV, Cone DC. Information Loss in Emergency Medical Services Handover of Trauma Patients. Prehospital Emergency Care 2009 Aug 13;13(3):280-285. [doi: 10.1080/10903120802706260]
2. Walderhaug S, Meland P, Mikalsen M, Sagen T, Brevik J. Evacuation support system for improved medical documentation and information flow in the field. International Journal of Medical Informatics 2008 Feb;77(2):137-151. [doi: 10.1016/j.ijmedinf.2007.01.006]
3. Scott LA, Brice JH, Baker CC, Shen P. An Analysis of Paramedic Verbal Reports to Physicians in the Emergency Department Trauma Room. Prehosp Emerg Care – 2003;7(2):247-251. [doi: 10.1080/10903120390936888]

4. Berg M. Patient care information systems and health care work: a sociotechnical approach. International Journal of Medical Informatics 1999 Aug;55(2):87-101. [doi: 10.1016/s1386-5056(99)00011-8]
5. Therien SP, Nesbitt ME, Duran-Stanton AM, Gerhardt RT. Prehospital Medical Documentation in the Joint Theater Trauma Registry: A Retrospective Study. The Journal of Trauma: Injury, Infection, and Critical Care 2011;71(supplement):S103-S108. [doi: 10.1097/ta.0b013e3182218fd7]
6. Heard J, Paris R, Scully D, McNaughton C, Ehrenfeld J, Coco J, et al. Automatic Clinical Procedure Detection for Emergency Services. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society; July 23, 2019; Berlin, Germany p. 337-340. [doi: 10.1109/embc.2019.8856281]
7. Paris R, Sullivan P, Heard J, Scully D, McNaughton C, Ehrenfeld J, et al. Heatmap generation for emergency medical procedure identification. 2019 Presented at: SPIE 10951, Medical Imaging 2019: ImageGuided Procedures, Robotic Interventions, and Modeling; March 8, 2019; San Diego, CA p. 1095130. [doi: 10.1117/12.2513122]
8. Bloos S, McNaughton C, Coco J, Novak L, Adams J, Bodenheimer R, et al. Feasibility Assessment of a Prehospital Automated Sensing Clinical Documentation System. Proceedings of the American Medical Informatics Association Annual Symposium 2019:24 [FREE Full text]
9. King DR. Initial Care of the Severely Injured Patient. N Engl J Med 2019 Feb 21;380(8):763-770. [doi: 10.1056/nejmra1609326]
10. URL: https://dod.defense.gov/News/Casualty-Status [accessed 2019-03-11]
11. Headquarters DOTA. Casualty Care, HQDA ATP (Army Techniques Publication) 4-02. Washington, DC: Department of Defense; 2013:1-38.
12. U.S. ACAC. Tactical Combat Casualty Care Handbook, Version 5. Ft. Leavenworth, KS: United States Army; 2017.
13. Joint COS. Health Service Support, Joint Publication 4-02. Washington, DC: Department of Defense; 2017:I-3-I-4.
14. Defense HA. Implementation Guidance for the Utilization of DD Form 1380, Tactical Combat Casualty Care (TCCC) Card, Defense Health Agency Procedural Instruction #6040. In: Implementation Guidance for the Utilization of DD Form 1380, Tactical Combat Casualty Care (TCCC) Card, Defense Health Agency Procedural Instruction #6040. Washington, DC: Department of Defense; 2017:01.
15. Merten H, van Galen LS, Wagner C. Safe handover. BMJ 2017 Oct 09:j4328. [doi: 10.1136/bmj.j4328]
16. Bost N, Crilly J, Patterson E, Chaboyer W. Clinical handover of patients arriving by ambulance to a hospital emergency department: A qualitative study. International Emergency Nursing 2012 Jul;20(3):133-141. [doi: 10.1016/j.ienj.2011.10.002]
17. Cohen MD, Hilligoss PB. The published literature on handoffs in hospitals: deficiencies identified in an extensive review. 2010 Apr 08;19(6):493-497. [doi: 10.1136/qshc.2009.033480]
18. Talbot R, Bleetman A. Retention of information by emergency department staff at ambulance handover: do standardised approaches work? Emergency Medicine Journal 2007 Aug 01;24(8):539-542. [doi: 10.1136/emj.2006.045906]
19. Agha RA. Handover in Trauma and Orthopaedic Surgery – A Human Factors Assessment. Annals of Medicine and Surgery 2012;1:25-29. [doi: 10.1016/s2049-0801(12)70009-2]
20. Carayon P, Cartmill R, Blosky MA, Brown R, Hackenberg M, Hoonakker P, et al. ICU nurses' acceptance of electronic health records. J Am Med Inform Assoc 2011 Nov 01;18(6):812-819. [doi: 10.1136/amiajnl-2010-000018]
21. Martinez VI, Marquard JL, Saver B, Garber L, Preusse P. Consumer Health Informatics Interventions Must Support User Workflows, Be Easy-To-Use, and Improve Cognition: Applying the SEIPS 2.0 Model to Evaluate Patients' and Clinicians' Experiences with the CONDUIT-HID Intervention. International Journal of Human–Computer Interaction 2017 Jan 11;33(4):333-343. [doi: 10.1080/10447318.2016.1278340]
22. Carayon P, Schoofs Hundt A, Karsh B, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec 01;15(suppl 1):i50-i58. [doi: 10.1136/qshc.2005.015842]
23. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013 Nov;56(11):1669-1686. [doi: 10.1080/00140139.2013.838643]
24. Denzin N, Lincoln Y, editors. Collecting and Interpreting Qualitative Materials. Thousand Oaks: Sage Publications; 1998.
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010]

Abbreviations

ASCD: automated sensing clinical documentation
ED: Emergency Department
EMS: emergency medical services
MTF: medical treatment facility
SEIPS: Systems Engineering Initiative for Patient Safety
TCCC: Tactical Combat Casualty Care
VUMC: Vanderbilt University Medical Center

Edited by G Eysenbach; submitted 26.01.20; peer-reviewed by J Mihanovic, D Pförringer; comments to author 12.06.20; revised version received 07.07.20; accepted 26.07.20; published 25.09.20.

Please cite as:
Novak LL, Simpson CL, Coco J, McNaughton CD, Ehrenfeld JM, Bloos SM, Fabbri D
Understanding the Information Needs and Context of Trauma Handoffs to Design Automated Sensing Clinical Documentation Technologies: Qualitative Mixed-Method Study of Military and Civilian Cases
J Med Internet Res 2020;22(9):e17978
URL: http://www.jmir.org/2020/9/e17978/
doi:10.2196/17978
PMID:32975522

Original Paper

Evaluating Two Common Strategies for Research Participant Recruitment Into Autism Studies: Observational Study

Kelli L Ahmed[1,2], PhD; Andrea R Simon[1,2], BA; Jack R Dempsey[3,4], PhD; Rodney C Samaco[1], PhD; Robin P Goin-Kochel[1,2], PhD

[1]Baylor College of Medicine, Houston, TX, United States

[2]Autism Center, Texas Children's Hospital, Houston, TX, United States

[3]Children's Hospital Colorado, Aurora, CO, United States

[4]University of Colorado School of Medicine, Aurora, CO, United States

Corresponding Author:
Robin P Goin-Kochel, PhD
Autism Center
Texas Children's Hospital
8080 N Stadium Drive, Suite 100
Houston, TX, 77054
United States
Phone: 1 8328243390
Email: kochel@bcm.edu

Abstract

Background: Ongoing research is necessary to better understand the causes of autism spectrum disorder (ASD), the developmental outcomes for individuals diagnosed with ASD, and the efficacy of the interventions. However, it is often difficult to recruit sufficient numbers of participants for studies, and despite the prevalence of ASD (currently estimated to affect 1 in 54 children), little research has focused on how to efficiently recruit participants with ASD.

Objective: The aim of this study was to determine the efficacy of two different paid advertisements—social media and radio advertising—in recruiting participants for a study enrolling people with ASD and their family members by examining the number of participants enrolled, the cost per participant, and the geographic reach of each type of advertising.

Methods: We examined participant enrollment in a study following nonoverlapping paid advertisements on a popular FM radio station (aired in three cities across two states) and Facebook (six advertisements that ran in five cities across two states). The total paid investment in the radio campaign was $12,030 and that in the Facebook campaign was $2950. Following the advertising campaigns, 1391 participants in the study who were affiliated with the Houston, Texas, site received email invitations to participate in a brief survey about the ways in which they learned about the study (eg, social media, medical provider, website) and which of these were most influential in their decisions to participate; 374 (26.8%) of the participants completed this survey.

Results: Social media advertising outperformed radio in all three parameters examined by enrolling more participants (338 vs 149), with a lower average cost per participant ($8.73 vs $80.74) and a wider geographic reach, based on a comparison of the number of zip codes within and outside of Texas for questionnaire respondents who rated social media as the most influential method of contact (n=367, χ^2_1=5.85, *P*=.02). Of the 374 survey participants, 139 (37.2%) reported that they had seen the study on social media prior to enrollment, while only 9 (2.4%) said they heard about it via radio.

Conclusions: Our findings suggest that advertising on social media can efficiently reach a large pool of potential participants with ASD, increasing the likelihood of meeting study enrollment goals. Researchers should consider allocating at least some portion of recruitment dollars to social media platforms as a means of quickly and inexpensively reaching out to their target populations, including for studies with in-person procedures.

(J Med Internet Res 2020;22(9):e16752) doi:10.2196/16752

KEYWORDS

autism spectrum disorder; participant recruitment; social media; Facebook; radio; genetic studies

Introduction

Background

A genetic contribution to the development of autism spectrum disorder (ASD) has been well established; however, for the majority of individuals with ASD, the genetic contributions remain unknown [1,2]. Continued study of ASD is critical for identifying additional causal mechanisms as well as for developing and improving personalized treatment approaches that facilitate optimal outcomes. However, successful recruitment for such studies can be challenging and can require substantial time and financial investments [3]. Using the National Library of Medicine clinical trials registry, Carlisle et al [4] showed that of 2579 phase 2 and 3 intervention clinical trials that closed in 2011, 19% terminated early because of low enrollment or enrolled less than 85% of their target enrollment numbers. A number of factors may contribute to recruitment challenges, including the demands of the study, the appeal of the advertisement, and importantly, the population being recruited. Understanding these factors and how they affect study enrollment is key to reducing the likelihood of underenrollment. Underenrolled studies consume resources that could be dedicated to other projects. Further, they may lack the statistical power necessary to uncover meaningful results and support the conclusions of the study [5]. Thus, understanding how to recruit participants *effectively* and *efficiently* is of the utmost importance to ensure successful completion of studies and to advance scientific knowledge [6].

Challenges in reaching clinical recruitment goals are ubiquitous; however, this is not due to lack of effort. Research teams have used a variety of recruitment methods to seek participants, with one such strategy being radio advertising. Recruiting for research through radio advertisements has been a popular strategy for decades [7], and its appeal lies in the ability to reach a wide audience with limited effort on behalf of the study team. According to a recent Nielsen report, nearly 92% of Americans are weekly radio listeners, with comparable listening rates across a variety of demographics (eg, sex, age cohort, and racial and ethnic groups [8]). Moreover, more than 93% of radio listeners continue to listen during commercials [9]. However, reports using radio advertising for research recruitment show that it can be costly, ranging from $80 to $827 per enrolled participant [10-14]. In four of these studies, radio advertising was the most expensive recruitment strategy used [11-14]. Additionally, the success of radio advertising in helping to achieve recruitment goals varies; in some published reports, radio advertising helped recruit the most participants of all strategies used [10,11,15], while in another study, only 1 participant (0.2%) was recruited in this way [16].

A modern recruitment strategy that has gained popularity in recent years is advertising on social media platforms, such as Facebook or Instagram. In the United States, 88% of people aged 18 to 29 years and 78% of people aged 30 to 49 years are connected to one or more social media platforms [17]. The most popular of these is Facebook, which is used by 68% of adults in the United States, 74% of whom visit the site daily [17]. Indeed, in a review of 27 studies, advertising on Facebook for research recruitment tended to be more time- and cost-effective than traditional research advertising strategies, with costs ranging from $1.36 to $110 per completed participant [3]. Moreover, Facebook ads can be tailored to target a specific audience. Focusing on families managing ASD, research has shown that these parents often rely heavily on *other* parents of children with ASD for information and support, including via Facebook groups [18-20]. Collectively, this suggests that social media advertising may be a particularly appealing way to reach select audiences and broadly share information about research opportunities, including to parents of children with ASD.

Purpose

Despite the high prevalence of ASD (1 of every 54 children in the United States) [21], there is a dearth of research that directly addresses how to best recruit participants with ASD and their families into research studies. In fact, to our knowledge, there is only one such report, which focused specifically on recruitment of Hispanic participants with ASD [22]. Thus, there is a great need for additional research that identifies the most effective recruitment strategies for participants with ASD. To this end, we examined two different mechanisms of paid advertising (radio and Facebook) in terms of number of participants recruited, cost, and geographic reach within the context of the SPARK (Simons Foundation Powering Autism Research for Knowledge) project. For the purposes of the current study, we focused on radio and Facebook recruitment strategies at the Houston SPARK site. Additionally, we solicited feedback from Houston-affiliated participants about their recruitment and enrollment experiences to compare against enrollment numbers and reflect participant perceptions about the most influential recruitment strategies.

SPARK: Project Overview

SPARK is a national, multi-site effort to enroll 50,000 individuals with ASD and their biological family members into a web-based genetic and phenotypic repository [23]. Briefly, more than 20 clinical sites across the United States form a clinical network to recruit potential participants and assist them throughout the enrollment process. SPARK participation is open to any individual living in the United States with a professional diagnosis of ASD or a dependent with a professional diagnosis of ASD. ASD diagnosis is ascertained through self-reporting, which has previously been shown to have >90% reliability [24]. Interested participants visit the study website, create a web-based profile, and complete the web-based enrollment forms to create a “primary” account. Because this “primary” account holder is required to be an independent adult, they can be either a parent or guardian of an individual with ASD or an independent adult with ASD. Once enrolled, each participating member of the family receives an at-home saliva collection kit with instructions on sample collection. Participants subsequently mail completed kits back to the laboratory for DNA analysis. Families can follow this process on their own or can contact a research coordinator for support.

At the Houston SPARK site, multiple strategies were used to recruit participants for SPARK; however, because existing literature suggested that the potential reach, popularity, and success of paid radio and Facebook campaigns may be similar,

these two approaches were singled out at the Houston site for further examination. Both campaigns occurred during nonoverlapping months within a 6-month timeframe, during which enrollment numbers were tracked daily through the web-based coordinator portal.

Methods

Recruitment Strategies

For the radio outreach, three FM radio campaigns were scheduled across two states (three cities) at a total investment of $12,030. All three campaigns ran during a 4-week period between November 28 and December 26, 2016, on a popular station that played holiday music. The aired advertisement was 30 seconds long and ran a total of 342 times across markets. Interested radio listeners were invited to text the word "SPARK" to a short-code telephone number (555-888) to receive information about a web-based US autism research study. These "subscribers" received three immediate text messages with 1) the study's institutional review entity–approved call to action, 2) the hyperlinked URL to register on the internet, and 3) the research coordinator's contact information. Two weeks after texting the short-code, subscribers received an automated reminder message with the enrollment link to the study.

For the social media campaign, a series of six paid Facebook ads were placed across two states (five cities) between March 23 and May 22, 2017, at a total investment of $2950. Five of these campaigns consisted of institutional review board–approved recruitment language, with a 2:07 minute video of the principal investigator at the Houston site explaining the study, while one campaign used the approved call-to-action language and an image to invite participants to enroll in person with the assistance of the study team. The geographic radius, campaign length, and dollar investment for each Facebook campaign varied slightly. However, all six campaigns were targeted using identical audience criteria for adults aged 22 to 55 years with interests in autism awareness organizations or special education. The Facebook ads targeted individuals who previously endorsed interest in the National Autism Association, World Autism Day, National Autistic Society, Special Education, Autism Spectrum Awareness, Stand Up for Autism, Asperger syndrome awareness, Autism Community Network, Autism Awareness, Autism Society of America, or Autism Speaks.

During both paid advertising campaigns, a minimal number of other traditional recruitment methods were used simultaneously. For example, flyers were posted in the clinic before, during, and between the two campaigns. Other "background" recruitment methods included physician referral, word of mouth, and information about the project on the clinic's website. These efforts were not consistently tracked; however, they remained consistent throughout both the radio and social media campaigns.

Questionnaire

Study data were collected and managed using Research Electronic Data Capture (REDCap) tools hosted at Texas Children's Hospital [25]. The questionnaire ascertained information about the participants' recruitment and enrollment experiences with SPARK, including all the ways they heard about the project before creating their web-based profile and which way was most influential in their decision to enroll. An invitation with a link to the questionnaire was emailed to the primary account holders, followed by three additional email reminders that were each sent one week apart.

Participants

The sampling pool for the questionnaire consisted of 1391 primary account holders affiliated with the Houston SPARK site who enrolled and consented to providing a saliva sample between April 21, 2016, and February 29, 2018. These participants received an invitation via email to complete a questionnaire about their recruitment and enrollment experiences with SPARK. A total of 374/1391 (26.8%) participants responded (see Table 1). The mean age of the 1391 respondents was 39 years (SD 8.7), 91% (1266/1391) identified as female, and 96% (1335/1391) were the parent or grandparent of a person with ASD, while 11 (3%) reported that they themselves had ASD (mean 30 years, SD 10.1 years). The mean age of the individuals with ASD (n=404, including the 11 independent adults) at the time of survey completion was 9.9 years (SD 6.6 years). On average, the 374 survey respondents reported a total of four people living in their household, with between 2 and 3 people participating in the SPARK study; 38 people (10.2%) reported that more than one person with ASD lived in their home.

Demographic Information

Race and ethnicity data were gathered for SPARK participants through two sources. For participants who are also affiliated with our clinical site, we obtained race and ethnicity data through our Epic medical record system. Additionally, SPARK distributed a survey in July 2017 to allow families to voluntarily provide race and ethnicity data. The demographic data for the state of Texas were obtained from the US census website [26]. Approved researchers can obtain the SPARK population dataset described in this study by applying at the Simons Institute Autism Research Initiative portal [27].

Table 1. Demographics of the survey recipients (N=1391), n (%).

Characteristic	Survey responders (n=374)	Survey nonresponders (n=1017)
Race		
White	243 (65.0)	536 (52.7)
Black or African American	36 (9.6)	92 (9.0)
Asian or Pacific Islander	15 (4.0)	35 (3.4)
Native American or American Indian	2 (0.5)	1 (<0.1)
More than one race	25 (6.7)	39 (3.8)
Other	13 (3.5)	17 (1.7)
Data unavailable	40 (10.7)	297 (29.2)
Ethnicity		
Hispanic	108 (35.2)	252 (35.0)
Highest level of education		
High school diploma, GED[a], or less	31 (8.3)	N/A[b]
Some college, no degree	88 (23.5)	N/A
Associate's degree	54 (14.4)	N/A
Bachelor's degree	101 (27.0)	N/A
Graduate degree	100 (27.0)	N/A

[a]GED: General Education Development.

[b]N/A: not applicable.

Data Analysis

To determine the number of participants from the radio campaign, we included any enrollments during the campaign, which lasted one month, as well as during the month following the campaign. This allowed all subscribers to receive a two-week reminder about enrollment and gave them an additional two weeks to enroll. More simply, any participant who enrolled in SPARK from November 28, 2016, through January 31, 2017, was counted for the radio campaign. To keep the amount of time consistent between the radio and Facebook campaign groups, we only counted participants who enrolled during the Facebook campaign itself, which ran for two months, from March 23 to May 22, 2017. Participants from both campaigns were given a deadline of October 30, 2017, to return the saliva kit for the individual with ASD.

To determine the effectiveness of each recruitment campaign, we examined three factors: number of participants, cost, and geographic reach. Two different numbers of participants were examined: the number of individuals with ASD who were enrolled and returned the saliva kit to the lab during the course of each campaign, and the number of primary account holders who reported on the questionnaire that they had heard about the study on radio or social media before registering. The cost per participant was determined by dividing the total cost of the campaign by the number of individuals with ASD who enrolled on the internet during that campaign. Geographic reach was examined using the Mapping toolbox in MATLAB (R2018b, MathWorks) to plot the zip codes provided by participants upon registration by latitude and longitude [23]. Using the zip codes provided by the survey participants, the numbers of zip codes in Texas versus outside of Texas were compared using chi-square analysis for the participants who responded that social media was their most influential method of contact.

Results

Number of Participants Recruited

Radio and Facebook Campaigns

Across the six-month time period encompassing both the radio and Facebook advertising campaigns, 568 individuals with ASD enrolled in the SPARK project. Of those 568 individuals, 520 (91.5%) also consented to providing a saliva sample, and 295 saliva kits were returned for an individual with ASD (as of October 30, 2017). During the radio advertising campaign, 378 people texted the short code, and 149 individuals with ASD were enrolled in the study; 140 of these individuals (94.0%) consented to providing a saliva sample, and 83 (55.7%) ultimately returned the saliva kit. During the Facebook campaign, 338 people with ASD were enrolled, of whom 312 (92.3%) consented to providing a saliva sample and 167 (49.4%) returned the saliva kit (see Tables 2 and 3).

Table 2. Number of participants enrolled during and between the media campaigns.

Participants	Radio campaign (11/28/2016-01/31/2017)	Time between campaigns (02/01/2017-03/22/2017)	Social media campaign (03/23/2017-05/23/2017)	Total sample (03/01/2016-12/31/2018)
Enrolled, n	149	81	338	2051
Consented to DNA test, n (%)	140 (94.0)	68 (84.0)	312 (92.3)	1925 (93.9)
Returned saliva kit, n (%)	83 (55.7)	45 (55.6)	167 (49.4)	1126 (54.9)

Table 3. Race and ethnicity of participants from media campaigns compared to the population of Texas.

Characteristic	Radio campaign (11/28/2016-01/31/2017; n=149), n (%)	Time between campaigns (02/01/2017-03/22/2017; n=81), n (%)	Social media campaign (03/23/2017-05/23/2017; n=338), n (%)	Total sample (03/01/2016-12/31/2018; n=2051), n (%)	Texas, % [26]
Race					
White	84 (75.0)	48 (75.0)	191 (74.9)	1155 (73.4)	78.8
Black or African American	12 (10.7)	5 (7.8)	33 (12.9)	198 (12.6)	12.8
Asian or Pacific Islander	8 (7.1)	4 (6.3)	6 (2.4)	73 (4.6)	5.3
Native American or American Indian	0 (0.0)	0 (0.0)	0 (0.0)	4 (0.3)	1.0
More than one race	5 (4.5)	4 (6.3)	18 (7.1)	92 (5.8)	2.0
Other	3 (2.7)	3 (4.7)	7 (2.7)	52 (3.3)	N/A
Data unavailable	37 (24.8)	17 (21.0)	83 (24.6)	477 (23.3)	N/A
Ethnicity					
Hispanic	41 (35.0)	25 (39.7)	73 (28.5)	546 (34.3)	39.4
Non-Hispanic	76 (65.0)	38 (60.3)	183 (71.5)	1044 (65.7)	N/A
Data unavailable	32 (21.5)	18 (22.2)	82 (24.3)	461 (22.5)	N/A

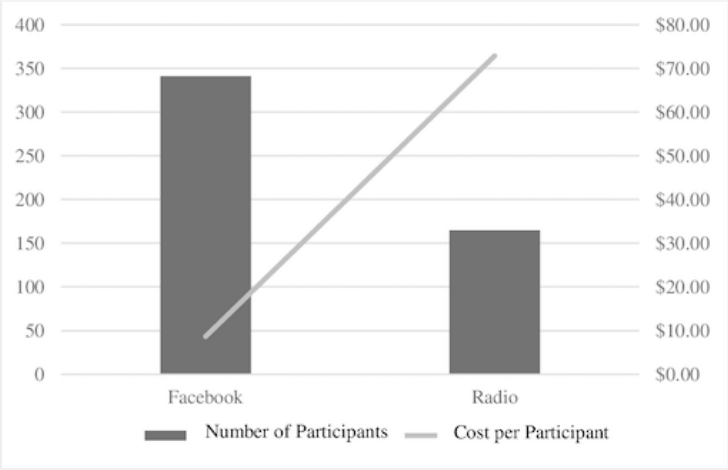
Questionnaire Respondents

Among the 374 participants who responded to the REDCap questionnaire, 139 (37.2%) reported seeing the study advertised via social media prior to registering. Among this group, 75/139 (54.0%) rated this form of contact as the most influential in their decision to enroll. In contrast, only 9/374 survey respondents (2.4%) reported hearing the advertisement for the study over the radio, and among this group, 6 (66.7%) said that this form of contact was the most influential.

Cost

As depicted in Figure 1, at an investment of $12,030 for radio, the cost per enrolled participant with ASD was $80.74 ($12,030/149). The cost of the Facebook campaign was $2950, for a per-participant cost of $8.73 ($2950/338).

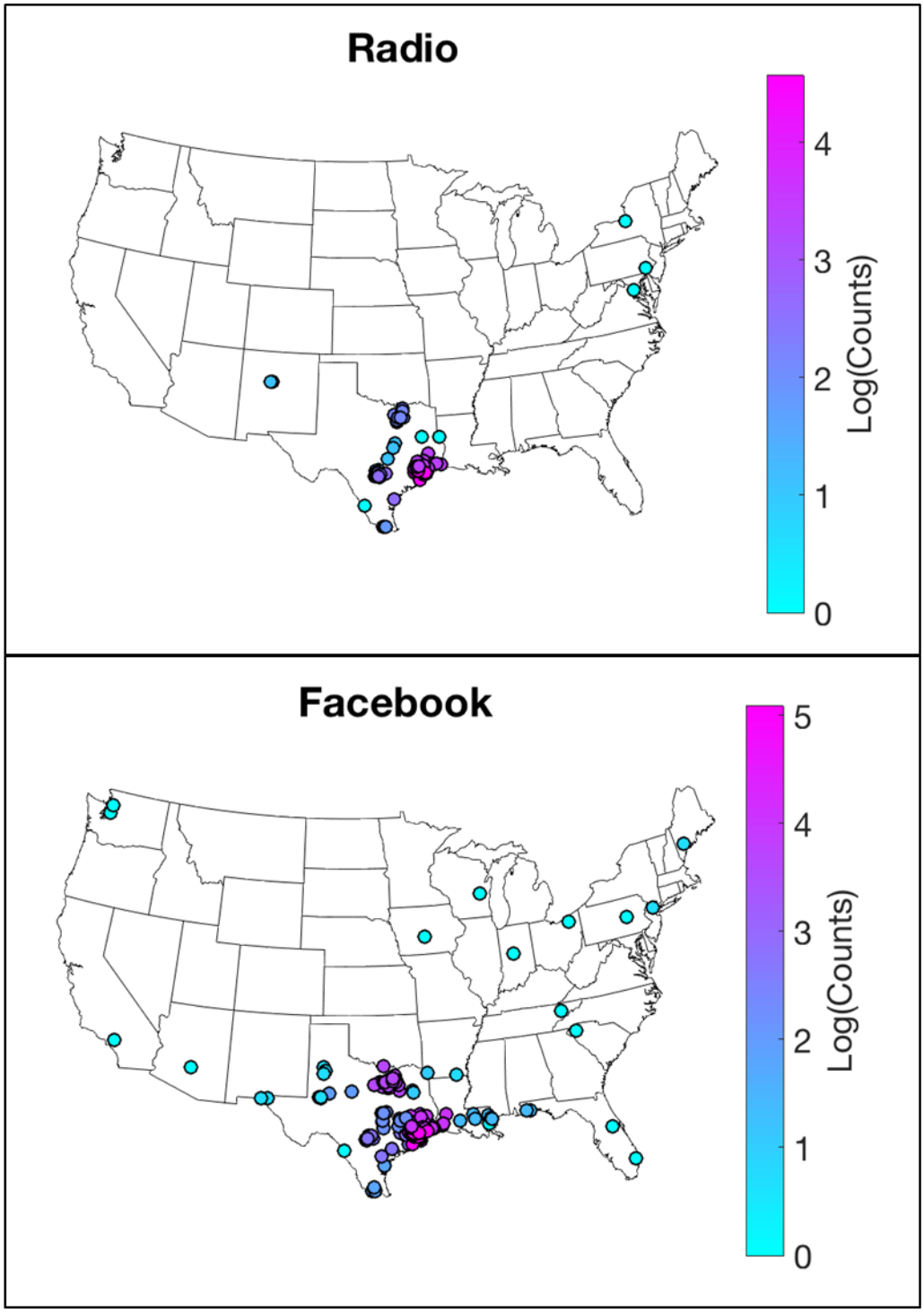
Figure 1. Comparison of the cost-efficiency of the Facebook and radio advertising campaigns.



Geographic Reach

Participants who enrolled in the SPARK project during the radio campaign were associated with 113 zip codes (Figure 2). These included the two states in which the radio campaign took place, plus four additional states. However, during the Facebook campaign, enrollment profiles were created from 208 zip codes across 16 states: the 2 states where the campaigns occurred, plus 14 additional states. Among participants who completed the REDCap questionnaire, those who identified social media as the most influential form of contact exhibited significant geographic differences compared to the rest of the sample. Of the individuals most influenced by social media, 13% indicated residence in a state outside of Texas, compared to only 5% of those most influenced by other strategies (n=367, χ^2_1=5.85, *P*=.02). There were no significant differences between these groups in terms of the likelihood to return saliva samples, racial or ethnic composition, or educational level. Only six participants endorsed radio advertisements as the most influential method of contact; this small subsample precluded comparable analyses for this recruitment method.

Figure 2. Comparison of the geographic reach of the radio and Facebook advertising campaigns.



Discussion

Principal Results

In this study, we sought to determine the effectiveness of two paid advertising campaigns for recruitment in a web-based autism study. The results clearly demonstrated that paid advertising on social media substantially outperformed paid advertising on radio for the SPARK project. This was shown across the three examined parameters of numbers of enrolled participants, cost, and geographic reach. More than twice as many individuals with ASD were enrolled in the project during the social media campaign, and over half of the individuals who learned about SPARK via social media indicated that this form of contact was the most influential in their decision to participate. The Facebook campaign also proved to be more cost-effective; advertising on radio cost almost 10 times as much per enrolled participant as advertising on social media. Finally, the zip codes provided by the participants when they enrolled on the internet showed that social media had a much broader geographic reach than radio; participants enrolled via Facebook from 14 states *outside* the 2 states where the advertisements originated, compared to only 4 outside states for the radio campaign.

Several reasons could explain the success of Facebook advertising compared to radio advertising for recruiting SPARK participants. First, advertisements on social media can be tailored specifically to people who meet certain criteria. The advertisements used here were targeted to participants within a specified age range who had already endorsed having an interest in ASD or special needs. This allowed our study team to avoid wasting "paid impressions" on people who were not likely to be interested in participating, which was not possible with the radio advertising campaign.

Another likely reason that Facebook outperformed radio is its ability to expand the reach of an advertisement through a single click. Paid Facebook advertisements can be easily shared to additional profiles, pages, and groups, yielding many additional "organic impressions" for paid content at no extra cost. One Facebook share can reach hundreds or even thousands of other people depending on that person's social circle and, if it is shared again, a virtual snowball phenomenon can emerge. This organic sharing process is also likely the reason we received so many registrations from states in which we did not advertise. Indeed, this particular advertisement received hundreds of shares and comments, and often the person sharing would "tag" people they thought would be interested in the advertisement, causing a personalized notification to be sent directly to that individual. In contrast, radio advertising is not easily shared and is certainly not shareable in real time. Furthermore, in radio advertising, a specific number of ad placements is agreed upon ahead of time by all parties, with no possibility of free "organic impressions." Additionally, while both paid advertisements are technically temporary, when paid content is shared organically on Facebook, it becomes semipermanent. When the advertisement is shared to a person's profile, page, or group, it will stay there until it is deleted or until newer content pushes it down the timeline feed. This allowed our Facebook advertisements to remain influential even after the advertising campaign technically ended (profiles created after the paid campaign period were omitted in the current study).

Finally, the success of the Facebook campaign may also be due to the social influence of social media. When listening to the radio, one cannot easily connect with other listeners. However, social networking sites allow individuals to instantly connect and communicate with other people anywhere in the world. Parents of children with special needs, including ASD, have been shown to frequently turn to other parents as sources of information and to heavily use online support groups, such as those on Facebook [18-20]. Indeed, our ad was shared in many online autism support groups, yielding additional organic impressions to a targeted audience who may have already been accustomed to receiving health-related advice and information from other parents in this way [18]. Further, studies have shown that our social networks can influence a wide range of behaviors, including offline behaviors such as exercise frequency [28]. Participation in a research study may thus also be influenced by a person in an individual's social network sharing information about the study.

Limitations

Although we were able to demonstrate the value of paid advertisements on social media for research recruitment, this study has a number of limitations. First, participants were not asked at the time of enrollment how they heard about the study; therefore, it is possible that participants who enrolled during the campaigns heard about SPARK through other mechanisms mentioned earlier, such as flyers, word of mouth, or the clinic website. However, these "background" recruitment efforts were consistent for both campaigns and thus were not expected to differentially impact the results. Furthermore, participants completed the REDCap survey up to one year post-enrollment, which may have affected the accuracy of the participants' responses. However, the data from the questionnaire and the number of participants who enrolled during the campaigns are consistent with each other, with a clear advantage seen in social media recruitment in both cases.

Conclusions

Despite the rising prevalence of ASD in the United States and our limited understanding of its causes, there is a surprising lack of studies examining how to recruit families with ASD into research studies. Because ASD is a heterogeneous condition both phenotypically and genetically, the sample size for etiological studies must be large, and efficiently recruiting participants will continue to be of primary importance. Understanding how to quickly enroll eligible participants allows research to progress more rapidly, bolstering the likelihood of success in understanding and treating ASD and related conditions. We concluded that participant recruitment via social media—specifically Facebook advertising—was superior to radio outreach across multiple parameters (participant numbers, cost, and geographic reach) for a web-based ASD-focused study that included submission of saliva samples for genetic analysis. Research teams attempting to target individuals with ASD should consider Facebook and other social media platforms as a viable, cost-effective recruitment strategy, including projects

that have offline components (eg, clinical assessments and medical procedures).

Future efforts should examine whether the success in Facebook advertising described here can be replicated in other types of studies, such as those with wholly in-person procedures, and in different patient populations. Additionally, it will be important to similarly evaluate the success of other common recruitment strategies, such as printed materials (eg, brochures and fliers), community events, or provider referrals, across multiple parameters; also, it should be determined which combinations of recruitment strategies yield the greatest return on investment. Collectively, this type of research stands to inform best practices with regard to efficient, cost-effective recruitment strategies that ensure the successful completion of studies and subsequent advancement of scientific knowledge.

Acknowledgments

This work was supported by a grant from the Simons Foundation (SFARI #385052, RPG-K). We are grateful to all the families in SPARK, the SPARK clinical sites, and SPARK staff. We appreciate the access to phenotypic data on the Simons Foundation Autism Research Initiative (SFARI) Base. This research was also partially supported by the Intellectual and Developmental Disabilities Research Center (1U54 HD083092) at Baylor College of Medicine.

Conflicts of Interest

RPGK contracted with Yamo Pharmaceuticals to consult on clinical trial design.

References

1. Geschwind DH. Genetics of autism spectrum disorders. Trends Cogn Sci 2011 Sep;15(9):409-416 [FREE Full text] [doi: 10.1016/j.tics.2011.07.003] [Medline: 21855394]
2. Quesnel-Vallières M, Weatheritt RJ, Cordes SP, Blencowe BJ. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. Nat Rev Genet 2019 Jan 2;20(1):51-63. [doi: 10.1038/s41576-018-0066-2] [Medline: 30390048]
3. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
4. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015 Feb 04;12(1):77-83 [FREE Full text] [doi: 10.1177/1740774514558307] [Medline: 25475878]
5. Torgerson JS, Arlinger K, Käppi M, Sjöström L. Principles for Enhanced Recruitment of Subjects in a Large Clinical Trial. Control Clin Trials 2001 Oct;22(5):515-525. [doi: 10.1016/s0197-2456(01)00165-9]
6. Nasser N, Grady D, Balke CW. Commentary: Improving Participant Recruitment in Clinical and Translational Research. Academic Medicine 2011;86(11):1334-1335. [doi: 10.1097/acm.0b013e3182302831]
7. Schrott HG, Merideth N. Recruitment by screening entire communities. The Iowa Lipid Research Clinic Experience. Circulation 1982 Dec;66(6 Pt 2):IV23-IV26. [Medline: 7127714]
8. State of the Media: Audio Today. The Nielsen Group. 2014. URL: http://www.nielsen.com/us/en/insights/reports/2014/state-of-the-media-audio-today-2014 [accessed 2018-10-23]
9. Generali P, Kurtzman W, Rose B. What Happens When The Spots Come On: 2011 Edition. Coleman Insights. 2011. URL: http://colemaninsights.com/wp-content/uploads/2013/07/What-Happens-When-The-Spots-Come-On-2011-Edition.pdf [accessed 2018-07-01]
10. Boyle RG, Enstad C, Asche SE, Thoele MJ, Sherwood NE. Evaluating strategies and costs to recruit smokeless tobacco users. Addict Behav 2007 Dec;32(12):3088-3092. [doi: 10.1016/j.addbeh.2007.06.005] [Medline: 17602843]
11. Cambron JA, Dexheimer JM, Chang M, Cramer GD. Recruitment methods and costs for a randomized, placebo-controlled trial of chiropractic care for lumbar spinal stenosis: a single-site pilot study. J Manipulative Physiol Ther 2010 Jan;33(1):56-61. [doi: 10.1016/j.jmpt.2009.11.002] [Medline: 20114101]
12. Heard A, March R, Maguire P, Reilly P, Helmore J, Cameron S, et al. Recruitment strategies for an osteoporosis clinical trial: analysis of effectiveness. Australas J Ageing 2012 Sep;31(3):176-180. [doi: 10.1111/j.1741-6612.2011.00573.x] [Medline: 22950589]
13. Williams MT, Proetto D, Casiano D, Franklin ME. Recruitment of a hidden population: African Americans with obsessive-compulsive disorder. Contemp Clin Trials 2012 Jan;33(1):67-75 [FREE Full text] [doi: 10.1016/j.cct.2011.09.001] [Medline: 21983626]
14. Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, et al. Design and Conduct of an Internet-Based Preconception Cohort Study in North America: Pregnancy Study Online. Paediatr Perinat Epidemiol 2015 Jul;29(4):360-371 [FREE Full text] [doi: 10.1111/ppe.12201] [Medline: 26111445]

15. Blaisdell L, Siewers A, Gagnon PC, Farina-Henry E, Begley J. Use of multi-modal community outreach and media to support recruitment in the National Children's Study. J Commun Healthc 2014 Dec 04;7(4):303-310. [doi: 10.1179/1753807614y.0000000064]
16. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb 14;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
17. Smith A, Anderson M. Social Media Use in 2018. Pew Research Center: Internet & Technology. 2018. URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2018-07-02]
18. Ammari T, Schoenebeck S, Morris M. Accessing Social Support and Overcoming Judgment on Social Media among Parents of Children with Special Needs. In: Association for the Advancement of Artificial Intelligence. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 22-31 URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8032
19. Mackintosh VJ, Myers B, Goin-Kochel RP. Sources of information and support used by parents of children with autism spectrum disorders. J Dev Disabil 2005;12(1):41-51 [FREE Full text]
20. Mohd Roffeei SH, Abdullah N, Basar SKR. Seeking social support on Facebook for children with Autism Spectrum Disorders (ASDs). Int J Med Inform 2015 May;84(5):375-385. [doi: 10.1016/j.ijmedinf.2015.01.015] [Medline: 25701266]
21. Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ 2020 Mar 27;69(4):1-12 [FREE Full text] [doi: 10.15585/mmwr.ss6904a1] [Medline: 32214087]
22. Zamora I, Williams ME, Higareda M, Wheeler BY, Levitt P. Brief Report: Recruitment and Retention of Minority Children for Autism Research. J Autism Dev Disord 2016 Feb 24;46(2):698-703. [doi: 10.1007/s10803-015-2603-6] [Medline: 26404703]
23. SPARK Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. Neuron 2018 Feb 07;97(3):488-493 [FREE Full text] [doi: 10.1016/j.neuron.2018.01.015] [Medline: 29420931]
24. Lee H, Marvin AR, Watson T, Piggot J, Law JK, Law PA, et al. Accuracy of phenotyping of autistic children based on Internet implemented parent report. Am J Med Genet B Neuropsychiatr Genet 2010 Sep 15;153B(6):1119-1126 [FREE Full text] [doi: 10.1002/ajmg.b.31103] [Medline: 20552678]
25. Harris PA, Lane L, Biaggioni I. Clinical Research Subject Recruitment: The Volunteer for Vanderbilt Research Program www.volunteer.mc.vanderbilt.edu. J Am Med Inform Assoc 2005 Nov 01;12(6):608-613. [doi: 10.1197/jamia.m1722]
26. QuickFacts United States. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/US/PST045219 [accessed 2020-08-25]
27. SFARI Base. URL: https://base.sfari.org/ [accessed 2020-08-25]
28. Althoff T, Jindal P, Leskovec J. Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior. Proc Int Conf Web Search Data Min 2017 Feb;2017:537-546 [FREE Full text] [doi: 10.1145/3018661.3018672] [Medline: 28345078]

Abbreviations

ASD: Autism Spectrum Disorder
REDCap: Research Electronic Data Capture
SFARI: Simons Foundation Autism Research Initiative
SPARK: Simons Foundation Powering Autism Research for Knowledge

Edited by G Eysenbach; submitted 21.10.19; peer-reviewed by E Da Silva, A Kulanthaivel; comments to author 15.02.20; revised version received 13.03.20; accepted 20.03.20; published 24.09.20.

Please cite as:
Ahmed KL, Simon AR, Dempsey JR, Samaco RC, Goin-Kochel RP
Evaluating Two Common Strategies for Research Participant Recruitment Into Autism Studies: Observational Study
J Med Internet Res 2020;22(9):e16752
URL: http://www.jmir.org/2020/9/e16752/
doi:10.2196/16752
PMID:32969826

Original Paper

Xigua Video as a Source of Information on Breast Cancer: Content Analysis

Peng Pan[1], PhD; Changhua Yu[1], MD; Tao Li[1], BA; Xilei Zhou[1], MD; Tingting Dai[1], MD; Hanhan Tian[1], MD; Yaozu Xiong[1], MD

Department of Radiation Oncology, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Jiangsu province, Huaian, China

Corresponding Author:
Changhua Yu, MD
Department of Radiation Oncology
The Affiliated Huaian No 1 People's Hospital of Nanjing Medical University
Jiangsu province
Radiotherapy Department Building, 4th floor
Huaian
China
Phone: 86 517 8490 7286
Email: yu884443@sina.com

Abstract

Background: Seeking health information on the internet is a popular trend. Xigua Video, a short video platform in China, ranks among the most accessed websites in the country and hosts an increasing number of videos with medical information. However, the nature of these videos is frequently unscientific, misleading, or even harmful.

Objective: Little is known about Xigua Video as a source of information on breast cancer. Thus, the study aimed to investigate the contents, quality, and reliability of breast cancer–related content on Xigua Video.

Methods: On February 4, 2020, a Xigua Video search was performed using the keyword “breast cancer.” Videos were categorized by 2 doctors based on whether the video content provided useful or misleading information. Furthermore, the reliability and quality of the videos were assessed using the 5-point DISCERN tool and 5-point global quality score criteria.

Results: Out of the 170 videos selected for the study, 64 (37.6%) were classified as useful, whereas 106 (62.4%) provided misleading information. A total of 41.8% videos (71/170) were generated by individuals compared to 19.4% videos (33/170) contributed by health care professionals. The topics mainly covered etiology, anatomy, symptoms, preventions, treatments, and prognosis. The top topic was “treatments” (119/170, 70%). The reliability scores and global quality scores of the videos in the useful information group were high (P<.001). No differences were observed between the 2 groups in terms of video length, duration in months, and comments. The number of total views was higher for the misleading information group (819,478.5 vs 647,940) but did not reach a level of statistical significance (P=.112). The uploading sources of the videos were mainly health care professionals, health information websites, medical advertisements, and individuals. Statistical differences were found between the uploading source groups in terms of reliability scores and global quality scores (P<.001). In terms of total views, video length, duration, and comments, no statistical differences were indicated among the said groups. However, a statistical difference was noted between the useful and misleading information video groups with respect to the uploading sources (P<.001).

Conclusions: A large number of Xigua videos pertaining to breast cancer contain misleading information. There is a need for accurate health information to be provided on Xigua Video and other social media; health care professionals should address this challenge.

(J Med Internet Res 2020;22(9):e19668) doi:10.2196/19668

KEYWORDS

breast cancer; internet; Xigua Video; content analysis

Introduction

Cancer incidence and mortality have been increasing in China, which have raised major health concerns [1]. Breast cancer ranks fifth among newly diagnosed cancers and is the most common form among Chinese women [2].

By the end of June 2019, the number of internet users in China had reached 854 million, out of which mobile internet users accounted for 99.1% of the total netizen population. Moreover, 759 million online video users accounted for 88.8% of the total internet users [3]. With the increasing popularity of online information, people tend to use the internet as an important source of health care information. Previous studies have reported that 80% of internet users access health information online [4], as freely available video websites are popular channels of information dissemination.

Xigua Video is a short video platform of aggregative media Jinri Toutiao ("today's headlines") run by Beijing ByteDance Technology Co Ltd. It is estimated to be the largest site for sharing of short videos in China [5]. Xigua Video places emphasis on user-uploaded videos and has accumulated an average of 120 million daily views [6]. Although information from the website is easily accessible, users may be unable to judge the quality and accuracy of its contents [7-10].

Previous studies have examined the quality of online information on various health topics [11-19]. Unfortunately, many studies revealed poor-quality, biased, and commercial content, which may lead to dangerous consequences for users.

To date, data on the evaluation of Xigua Video as a source of health information remain lacking. To better understand how Chinese internet users discuss breast cancer on social media platforms such as Xigua Video, we conducted an analysis to evaluate the content, quality, and reliability of Chinese-language breast cancer–related information available on Xigua Video.

Methods

Data Collection

On February 4, 2020, a Xigua Video search was performed using the keyword "breast cancer" for videos. All videos uploaded up by the search date were included. In the case of duplicate videos, only 1 was taken into consideration. Videos were excluded if they were irrelevant or lacked accompanying audio. As a result, the study obtained 170 videos in total.

After collecting data on content, duration in months on Xigua Video, number of views, length, and comments, 2 physicians independently evaluated the overall quality of the videos. The kappa statistic for the 2 reviewers was κ=0.78, which indicated substantial agreement [20].

Creation of Content Categories

As the title often does not reflect the actual content of the video, the study considered topics for categorization. Content was categorized according to a previous YouTube study on colorectal cancer [21]. If a video covered more than 1 topic, then each topic was listed separately.

In addition, the videos were classified into 4 groups according to source, namely, health care professionals, health information websites, medical advertisements, and individuals.

Scoring and Classification of Videos

Videos were classified into 2 groups, namely, useful information and misleading information. Videos rated as useful contained scientifically accurate information about any aspect of breast cancer, such as symptoms, treatment, and prevention. Conversely, misleading videos contained inaccurate or unproven information about the disease. This classification method has been used in other studies [22,23].

Video reliability was scored using the 5-point DISCERN tool [24]. In addition, the overall quality of the videos was rated using the 5-point global quality score criteria [25]. In both instruments, higher scores indicate higher content quality of the video clip.

Statistical Analysis

Statistical analysis was performed using SPSS 22.0 (IMB Corp). Numerical variables were reported as mean (SD) or median (IQR) values. The Student *t* test, the Mann-Whitney *U* test, analysis of variance, and the Kruskal-Wallis test were applied for the comparison of numerical variables. The Dwass-Steel-Critchlow-Fligner post hoc test was conducted after the Kruskal-Wallis test. Categorical variables were stated as number (n) and percentage (%). In the comparison of categorical variables, chi-square and Fisher exact tests were used. A value of $P<.05$ was considered statistically significant.

Ethical Approval and Consent to Participate

This study did not require approval by the local research ethics board since only publicly available data were used.

Results

Topics and Uploading Sources

The videos were analyzed based on the topics covered therein. In all categories, the topic of treatments was the most frequently covered (119/170, 70%), followed by symptoms (56/170, 33%), prognosis (44/170, 26%), anatomy (34/170, 20%), prevention (25/170, 15%), and etiology (18/170, 11%).

A total of 41.8% (71/170) of the videos were posted on the website by individuals. Medical advertisements and health information websites were responsible for uploading 25.9% (44/170) and 12.9% (22/170) of the total videos, respectively. The videos contributed by health care professionals accounted for only 19.4% (33/170).

Information Reliability and Quality

As previously mentioned, the selected videos were classified into useful and misleading groups. Of the 170 selected videos, the number of videos containing misleading information was 106 (62.4%); 64 (37.6%) contained useful information. Table 1 presents the classification of the video characteristics. A statistically significant difference was determined in the useful information group with respect to the reliability and global quality scores ($P<.001$). However, no significant differences

were noted in terms of total views (*P*=.11), video length (*P*=.66), duration in months on Xigua Video (*P*=.051), and comments between the 2 groups (*P*=.73).

In addition, videos were compared according to their ownership. Table 2 presents the analysis of the video characteristics by uploading sources. Statistically distinctive differences between the uploading sources were observed for the reliability and global quality scores (*P*<.001). In terms of total views (*P*=.22), video length (*P*=.06), duration in months (*P*=.15), and comments (*P*=.47), no statistical differences were found. In the useful information group, 35.9% (23/64) and 20.3% (13/64) of the videos were generated by health care professionals and health information websites, respectively. On the contrary, 50% (53/106) of the videos in the misleading information group were generated by individuals. Data indicated statistical difference between the useful and misleading groups in terms of uploading sources (*P*<.001).

Table 1. Analysis of video characteristics by usefulness category.

Characteristics	Useful group	Misleading group	*P* value
Views, IQR	592,515.25	894,786	.11
Video length (s), IQR	234	264.5	.66
Duration in months on Xigua Video, mean (SD)	11.9 (7.6)	10.7 (5.9)	.051
Comments, IQR	37	42.5	.73
Reliability score, mean (SD)	2.34 (1.03)	1.45 (1.01)	<.001
Global quality score, mean (SD)	3.58 (1.24)	2.12 (0.99)	<.001

Table 2. Analysis of video characteristics by uploading sources.

Characteristics	Health care professionals	Health information websites	Medical advertisements	Individuals	*P* value
Videos (N=170), n (%)	33 (19.4)	22 (12.9)	44 (25.9)	71 (41.8)	N/A[a]
Reliability score, mean (SD)	2.85 (1.12)	2.55 (1.26)	2.07 (0.82)	1.65 (1.11)	<.001
Global quality score, mean (SD)	3.45 (1.06)	3.50 (1.34)	1.89 (0.92)	2.37 (1.123)	<.001
Total views, IQR	1,010,639	840,070.75	898,145	857,950	.22
Video length (s), IQR	246	245	209	260	.06
Duration on Xigua Video (months), mean (SD)	13.5 (8.1)	10.0 (7.1)	10.6 (5.6)	10.7 (6.1)	.15
Comments, IQR	36.5	62.3	46.8	41	.47
Misleading information (n=106), n (%)	10 (9.4)	9 (8.5)	34 (32.1)	53 (50.0)	<.001
Useful information (n=64), n (%)	23 (35.9)	13 (20.3)	10 (15.6)	18 (28.2)	N/A

[a]N/A: not applicable.

Discussion

Overview

Breast cancer is the most common form of cancer and accounts for 30.4% of total cancer incidence and 2.51% of total mortality due to cancer among women in China. Notably, the incidence rate of breast cancer has increased rapidly in women aged 20-55 years. Therefore, the effectiveness of strategies for prevention, early detection, and management of breast cancer is an essential aspect to be considered.

A broad range of information is available on the internet. The China Research Institute for Science Popularization analyzed netizens' behavior and requirements for information [26]. The results indicated that the science communication search index, which is based on the amount of time people spend on search engines, increased from 2.8 billion in 2014 to 9.2 billion in 2019. Young and middle-aged netizens were the main drivers of the trend with health and medical care as the top topic. With the rapid development of the internet, new media are providing important access to health and medical information. Previous studies reported YouTube as an information source for various forms of cancers [27-31]. Although YouTube is inaccessible in China for a number of reasons, many similar video websites deliver the same functionality, such as Jinri Toutiao, which was launched in 2012. The website has gained 200 million monthly active users [6]. Specifically, Xigua Video is the video platform of Jinri Toutiao. To the best of our knowledge, no study has assessed the contents, quality, and reliability of videos on the internet as a source of health information related to breast cancer in China.

Restricted by relevant Chinese laws, users' gender and age cannot be displayed. However, the study obtained useful information from other sources. According to the Baidu search engine's statistics on "breast cancer" search results, approximately two-thirds of users are women, and 90% are aged

20-50 years. Therefore, we infer that Jinri Toutiao users who search for information on breast cancer are mainly young and middle-aged female netizens.

Principal Findings

Our study found that nearly half of the videos generated by individuals covered the topic of treatment. Alarmingly, approximately two-thirds of the videos spread misleading information. Furthermore, the quality of the videos was low. In addition, the fact that viewers gave better ratings to poor-quality videos compared to the high-quality videos indicates that the majority of health seekers were incapable of recognizing low-quality medical information in videos.

Low Percentage of Good- or Excellent-Quality Videos

The topics of the videos observed in the study mainly covered etiology, anatomy, symptoms, prevention, treatments, and prognosis. The majority of the videos only contained 1 or 2 of the above-mentioned categories. In all the categories, treatments ranked first. The videos examined gained a total of 13 million views with a total watch time of more than 16 hours and 8598 comments. These data illustrate that videos on breast cancer are extremely popular. However, the contents of many videos frequently lack peer review, which limits the reliability and accuracy of the videos on Xigua Video. Although the reliability and global quality scores of the useful videos were high, the number of useful videos (64/170) was lower than that of the misleading videos (106/170). Notably, the number of total views was higher for the videos in the misleading information group (IQR 894-786) than that for the useful information group (IQR 592-515.25). These findings indicate that viewers may find it difficult to differentiate between useful and misleading information. A total of 41.8% (71/170) videos were produced by individuals, the majority (53/71, 74.6%) of which contained misleading information. Given that users who want to access information are mostly non–health care personnel, results indicate obtaining accurate information from Xigua videos is difficult.

In general, more reliable information can be obtained from videos uploaded by health care professionals. However, the proportion of these videos reached only 35.9% (23/64) in the useful information group. In other words, only one-third of useful videos had been uploaded by health care professionals. This result is similar to that reported by Mueller et al [4].

The most commonly watched videos were those containing misleading information, whereas the least watched videos were those from health care professionals. These results confirm that effective measures should be taken to reduce inaccurate (and possibly harmful) information on this video platform.

The results further indicated that videos from health care professionals have significantly higher reliability and global quality scores than those posted by individuals. As such, lay users find difficulty in distinguishing useful information from a massive number of videos. Moreover, the results indicated that ownership is an important element that can be used to assess the reliability of videos. If professionals such as doctors, health care organizations, and health information websites produced videos, then the content of those videos may be considered trustworthy [32]. The study demonstrated that videos produced by health care professionals were geared toward more educational purposes. This finding is in accord with those of other studies [33-35]. Hence, such professionals should utilize their expertise and contribute high-quality videos for patients as information sources on websites. Conversely, individuals contributed 18 videos (28.2%) to the useful information group. However, our results emphasizes that these videos may not be able to cover all aspects of breast cancer. Therefore, many videos can contain some valuable information despite providing overall misleading content.

Limitations

A limitation of this study is that only videos in the Chinese language were examined. Second, the results comprise a snapshot of information distribution to illustrate the quality of videos on the Xigua platform at one point in time. Therefore, these results may change as videos are added or removed over time.

Conclusions

Despite the rising incidence of breast cancer, the level of public awareness in China remains low. Chinese websites, such as Xigua Video, provide a new medium for disseminating cancer information to the public through videos. For the sake of public awareness, health care and medical professionals should adopt this technology and take effective actions to provide accurate information about breast cancer on these video websites. The lack of reliable and useful health information on the internet remains a significant problem, and health care professionals and governments urgently need to address this challenge.

Acknowledgments

We would like to thank Honglei Luo for critically reviewing the manuscript.

Authors' Contributions

PP and CY participated in the conceptualization of the paper. PP and XZ conducted the data searches on the internet. HT, TD, and YX extracted relevant and analysis data. TL and PP critically reviewed the manuscript for important intellectual content. PP structured and wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. National Bureau of Statistics of China. China Statistical Yearbook 2019. Beijing: China Statistics Press; 2019.
2. Chen W, Sun K, Zheng R, Zeng H, Zhang S, Xia C, et al. Cancer incidence and mortality in China, 2014. Chin J Cancer Res 2018 Feb;30(1):1-12 [FREE Full text] [doi: 10.21147/j.issn.1000-9604.2018.01.01] [Medline: 29545714]
3. China Internet Network Information Center (CNNIC). Statistical Report on Internet Development in China. 2019. URL: www.cac.gov.cn/wxb_pdf/0228043.pdf [accessed 2019-02-28]
4. Mueller SM, Jungo P, Cajacob L, Schwegler S, Itin P, Brandt O. The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers. J Med Internet Res 2019 Jan 16;21(1):e11935 [FREE Full text] [doi: 10.2196/11935] [Medline: 30664460]
5. iiMedia Report. 2019 H1 China Micro-video Innovation Trend Research Report, iiMedia Report. URL: https://report.iimedia.cn/repo13-0/38875.html [accessed 2019-09-11]
6. 23 Interesting Toutiao Facts and Statistics (2020). URL: https://expandedramblings.com/index.php/qutoutiao-statistics-facts/ [accessed 2020-02-11]
7. Hasamnis AA, Patil SS. YouTube as a tool for health education. J Educ Health Promot 2019;8:241 [FREE Full text] [doi: 10.4103/jehp.jehp_150_19] [Medline: 32002413]
8. Aldallal SN, Yates JM, Ajrash M. Use of YouTube™ as a self-directed learning resource in oral surgery among undergraduate dental students: a cross-sectional descriptive study. Br J Oral Maxillofac Surg 2019 Dec;57(10):1049-1052. [doi: 10.1016/j.bjoms.2019.09.010] [Medline: 31699466]
9. Nour MM, Nour MH, Tsatalou O, Barrera A. Schizophrenia on YouTube. Psychiatr Serv 2017 Jan 01;68(1):70-74. [doi: 10.1176/appi.ps.201500541] [Medline: 27524369]
10. de'Angelis N, Gavriilidis P, Martínez-Pérez A, Genova P, Notarnicola M, Reitano E, et al. Educational value of surgical videos on YouTube: quality assessment of laparoscopic appendectomy videos by senior surgeons vs. novice trainees. World J Emerg Surg 2019;14:22 [FREE Full text] [doi: 10.1186/s13017-019-0241-6] [Medline: 31086560]
11. Jain N, Abboudi H, Kalic A, Gill F, Al-Hasani H. YouTube as a source of patient information for transrectal ultrasound-guided biopsy of the prostate. Clin Radiol 2019 Jan;74(1):79.e11-79.e14. [doi: 10.1016/j.crad.2018.09.004] [Medline: 30322707]
12. Gokcen HB, Gumussuyu G. A Quality Analysis of Disc Herniation Videos on YouTube. World Neurosurg 2019 Feb 02;124:e799-e804. [doi: 10.1016/j.wneu.2019.01.146] [Medline: 30721774]
13. Oremule B, Patel A, Orekoya O, Advani R, Bondin D. Quality and Reliability of YouTube Videos as a Source of Patient Information on Rhinoplasty. JAMA Otolaryngol Head Neck Surg 2019 Mar 01;145(3):282-283 [FREE Full text] [doi: 10.1001/jamaoto.2018.3723] [Medline: 30605207]
14. Clarke Hillyer G, Basch CH, Guerro S, Sackstein P, Basch CE. YouTube videos as a source of information about mastectomy. Breast J 2019 Mar;25(2):349-350. [doi: 10.1111/tbj.13222] [Medline: 30734408]
15. Duke CH, Yin J, Zhang X, Blankenship EB, Akuse SE, Shah GH, et al. Adopting YouTube to Promote Health: Analysis of State Health Departments. Perm J 2019;23:1-6 [FREE Full text] [doi: 10.7812/TPP/18-094] [Medline: 30939268]
16. Basch CH, Hillyer GC, Garcia P, Basch CE. Content of widely viewed YouTube videos about celiac disease. Public Health 2019 Feb;167:147-151. [doi: 10.1016/j.puhe.2018.11.004] [Medline: 30682698]
17. Yiannakoulias N, Slavik CE, Chase M. Expressions of pro- and anti-vaccine sentiment on YouTube. Vaccine 2019 Apr 03;37(15):2057-2064. [doi: 10.1016/j.vaccine.2019.03.001] [Medline: 30862365]
18. Woo BKP. What Can We Learn From Diabetes-Related YouTube Videos? Can J Diabetes 2018 Aug;42(4):342. [doi: 10.1016/j.jcjd.2018.01.009] [Medline: 30072018]
19. Rayi A, Borad SJ, Kemper SE, Malhotra K. What information about sudden unexpected death in epilepsy (SUDEP) is available on YouTube? Epilepsy Behav 2019 Apr;93:125-128. [doi: 10.1016/j.yebeh.2018.10.017] [Medline: 30528465]
20. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
21. Zhang S, Yang Y, Yan D, Yuan B, Jiang X, Song C. Internet videos and colorectal cancer in mainland China: a content analysis. BMC Med Inform Decis Mak 2018 Dec 04;18(1):129 [FREE Full text] [doi: 10.1186/s12911-018-0711-x] [Medline: 30514290]
22. Gul M, Diri MA. YouTube as a Source of Information About Premature Ejaculation Treatment. J Sex Med 2019 Nov;16(11):1734-1740. [doi: 10.1016/j.jsxm.2019.08.008] [Medline: 31522984]
23. Esen E, Aslan M, Sonbahar B, Kerimoğlu RS. YouTube English videos as a source of information on breast self-examination. Breast Cancer Res Treat 2019 Feb;173(3):629-635. [doi: 10.1007/s10549-018-5044-z] [Medline: 30443880]
24. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
25. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
26. China Research Institute for Science Popularization. A report on Chinese netizens' search behavior for popular science. 2019. URL: https://www.cast.org.cn/col/col1280/index.html [accessed 2019-04-26]

27. Loeb S, Sengupta S, Butaney M, Macaluso JN, Czarniecki SW, Robbins R, et al. Dissemination of Misinformative and Biased Information about Prostate Cancer on YouTube. Eur Urol 2019 Apr;75(4):564-567. [doi: 10.1016/j.eururo.2018.10.056] [Medline: 30502104]
28. Hassona Y, Taimeh D, Marahleh A, Scully C. YouTube as a source of information on mouth (oral) cancer. Oral Dis 2016 Apr;22(3):202-208. [doi: 10.1111/odi.12434] [Medline: 26718020]
29. Sahin AN, Sahin AS, Schwenter F, Sebajang H. YouTube Videos as a Source of Information on Colorectal Cancer: What Do Our Patients Learn? J Cancer Educ 2019 Dec;34(6):1160-1166 [FREE Full text] [doi: 10.1007/s13187-018-1422-9] [Medline: 30242615]
30. Myrick JG, Oliver MB. Laughing and crying: mixed emotions, compassion, and the effectiveness of a YouTube PSA about skin cancer. Health Commun 2015;30(8):820-829. [doi: 10.1080/10410236.2013.845729] [Medline: 24877892]
31. Aydin MA, Akyol H. Quality of Information Available on YouTube Videos Pertaining to Thyroid Cancer. J Cancer Educ 2019 Mar 05:599-605. [doi: 10.1007/s13187-019-01502-9] [Medline: 30838529]
32. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
33. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Nov 14;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
34. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612-2621 [FREE Full text] [Medline: 11368735]
35. Hayanga AJ, Kaiser HE. Medical information on YouTube. JAMA 2008 Mar 26;299(12):1424-1425. [doi: 10.1001/jama.299.12.1424-b] [Medline: 18364483]

Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by C Basch, W Cheng; comments to author 20.05.20; revised version received 24.08.20; accepted 03.09.20; published 29.09.20.

Please cite as:
Pan P, Yu C, Li T, Zhou X, Dai T, Tian H, Xiong Y
Xigua Video as a Source of Information on Breast Cancer: Content Analysis
J Med Internet Res 2020;22(9):e19668
URL: http://www.jmir.org/2020/9/e19668/
doi:10.2196/19668
PMID:32883651

Original Paper

Public Disclosure on Social Media of Identifiable Patient Information by Health Professionals: Content Analysis of Twitter Data

Wasim Ahmed[1], BA, MSc, PhD; Reshma Jagsi[2], DPhil, MD; Thomas G Gutheil[3], MD; Matthew S Katz[4], MD

[1]Department of Marketing, Operations and Systems, Newcastle University Business School, Newcastle University, Newcastle upon Tyne, United Kingdom

[2]Department of Radiation Oncology, Center for Bioethics and Social Science in Medicine, University of Michigan, Ann Arbor, MI, United States

[3]Department of Psychiatry, Beth Israel Deaconess Medical Center, Mass. Mental Health Center, Harvard Medical School, Harvard University, Boston, MA, United States

[4]Department of Radiation Medicine, Lowell General Hospital, Lowell, MA, United States

Corresponding Author:
Wasim Ahmed, BA, MSc, PhD
Department of Marketing, Operations and Systems
Newcastle University Business School
Newcastle University
5 Barrack Rd
Newcastle upon Tyne, NE1 4SE
United Kingdom
Phone: 44 191 2081704
Email: Wasim.Ahmed@Newcastle.ac.uk

Abstract

Background: Respecting patient privacy and confidentiality is critical for doctor-patient relationships and public trust in medical professionals. The frequency of potentially identifiable disclosures online during periods of active engagement is unknown.

Objective: The objective of this study was to quantify potentially identifiable content shared on social media by physicians and other health care providers using the hashtag #ShareAStoryInOneTweet.

Methods: We accessed and searched Twitter's API using Symplur software for tweets that included the hashtag #ShareAStoryInOneTweet. We identified 1206 tweets by doctors, nurses, and other health professionals out of 43,374 tweets shared in May 2018. Tweet content was evaluated in January 2019 to determine the incidence of instances where names or potentially identifiable information about patients were shared; content analysis of tweets in which information about others had been disclosed was performed. The study also evaluated whether participants raised concerns about privacy breaches and estimated the frequency of deleted tweets. The study used dual, blinded coding for a 10% sample to estimate intercoder reliability using Cohen κ statistic for identifying the potential identifiability of tweet content.

Results: Health care professionals (n=656) disclosing information about others included 486 doctors (74.1%) and 98 nurses (14.9%). Health care professionals sharing stories about patient care disclosed the time frame in 95 tweets (95/754, 12.6%) and included patient names in 15 tweets (15/754, 2.0%). It is estimated that friends or families could likely identify the clinical scenario described in 242 of the 754 tweets (32.1%). Among 348 tweets about potentially living patients, it was estimated that 162 (46.6%) were likely identifiable by patients. Intercoder reliability in rating the potential identifiability demonstrated 86.8% agreement, with a Cohen κ of 0.8 suggesting substantial agreement. We also identified 78 out of 754 tweets (6.5%) that had been deleted on the website but were still viewable in the analytics software data set.

Conclusions: During periods of active sharing online, nurses, physicians, and other health professionals may sometimes share more information than patients or families might expect. More study is needed to determine whether similar events arise frequently and to understand how to best ensure that patients' rights are adequately respected.

(J Med Internet Res 2020;22(9):e19746) doi:10.2196/19746

KEYWORDS

Social Media; Twitter; Patient Information; Confidentiality; Health Professionals

Introduction

Background

Physicians, nurses, and other health professionals remain among the most trusted professionals in the United States because of their commitment to the well-being of others; they are a trusted source of health information and guidance [1]. Surveys have demonstrated the high trust in health care professionals of the US public with even higher levels of trust in other countries [1-3]. Still recited by many medical students as they become physicians, the Hippocratic Oath reflects the fundamental importance of patient privacy as a critical element of the doctor-patient relationship and a precondition for the trust of the public. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) requires deidentification of data to avoid sharing protected health information [4]. The US Patient's Bill of Rights also states that patients have the right to be able to talk privately with medical professionals and that personal information be protected. Thus, both ethical and legal reasons to maintain patient privacy exist.

Fulfilling physicians' obligations to protect the well-being and privacy of their patients is complicated in the age of the internet. Internet culture is very different from that of the medical profession, creating potential ethical problems with boundaries and privacy that did not exist when physicians interacted exclusively offline. In order to maintain the trust of the public and that of individual patients, physicians increasingly need to understand the limits and risks of disclosure of certain types of information online. Although concerns about unprofessional medical student and resident behavior online have been articulated before [5,6], the ethical risks of public disclosure, when narrative medicine intersects with social media, remain poorly defined [7,8].

Social media usage has become popular among medical professionals. A survey in 2014 by QuantiaMD [9] found that, of 4000 physicians surveyed, 90% noted that they used some form of social media for personal activities, and 65% used social media for professional reasons. In May 2018, thousands of individuals—including many health care professionals—shared health-related stories on Twitter using the hashtag *#ShareAStoryInOneTweet* in response to one physician's spontaneous tweet of a patient story that included this hashtag. Certain tweets included potentially identifiable information that could be considered a breach of confidence when disclosed without patient consent, risking harm to patients, physician's careers, and public trust in the profession.

An article in July 2018 [10] highlighted the importance of sharing stories but did not address the potential risks of sharing online. Over time, some viral tweets have been deleted, raising further concerns that the platform allowing easy disclosure might have led to posts that authors subsequently regretted. A notable example of a popular post (altered to avoid identification) was retweeted 13,491 times and liked by 55,994 people before being deleted:

> *I delivered a baby very underweight, weighing two pounds. They said he did not have a chance. I remained with him for a couple of days. Nine years later, he played his first football game last week.*

Hashtags can make online content searchable and discoverable online, regardless of time since publication [11]. The American Medical Association, Massachusetts Medical Society, and other organizations advise physicians to report unprofessional social media use [12,13]. What constitutes unprofessional behavior on social media is not clearly defined. To advance a common understanding and to facilitate subsequent discussion within the profession about what is appropriate, we sought to describe participation of physicians and other health professionals in this event, the reach of their postings, and the occurrence of potentially identifiable disclosures about patients.

Related Work

Early research on health professionals using Weblogs [14] examined 271 medical blogs, finding that individual patients were described in 114 blogs, and 45 blogs had enough information for patients to identify themselves. Scholars have questioned whether it would ever be ethical for medical professionals to write publicly about patients without their consent [15]. Previous work [16] where young doctors on Facebook were studied has also specifically highlighted privacy issues by finding that some of the private information shared by the doctors (the doctors' own private information shared by the doctors themselves) could bring the profession into disrepute. Previous work [17] has also noted that the use of social networking sites such as Twitter and Facebook by doctors can lead to complaints by patients.

Despite the importance of previous work [16,17], there appears to be a lack of empirical research on the use of popular hashtags for sharing patient stories by medical professionals. Understanding information sharing using hashtags, such as #ShareAStoryInOneTweet, is important because social media is becoming more ingrained in society, and potential privacy violations may exist in this context. Furthermore, as social media use increases, online disclosure of private information via social media is likely to remain an issue for health care systems around the world. However, recent research [18] has also highlighted the positive role medical professionals could play on social media, for instance, by countering medical misinformation.

The results of this study are likely to be of interest to those compiling guidelines for the use of social media by medical professionals.

Research Questions and Objectives

The overall research aim of this study was to develop a better understanding of the content shared with the hashtag #ShareAStoryInOneTweet.

The objectives of the study were to (1) identify unique tweets sent by doctors and other health care providers using the hashtag, (2) to develop an understanding of the characteristics of the doctors and health care providers using the hashtag, and (3) to categorize tweets into themes and identify the frequency of instances in which patients could be identified by themselves or by their family.

Methods

Cohort Definition

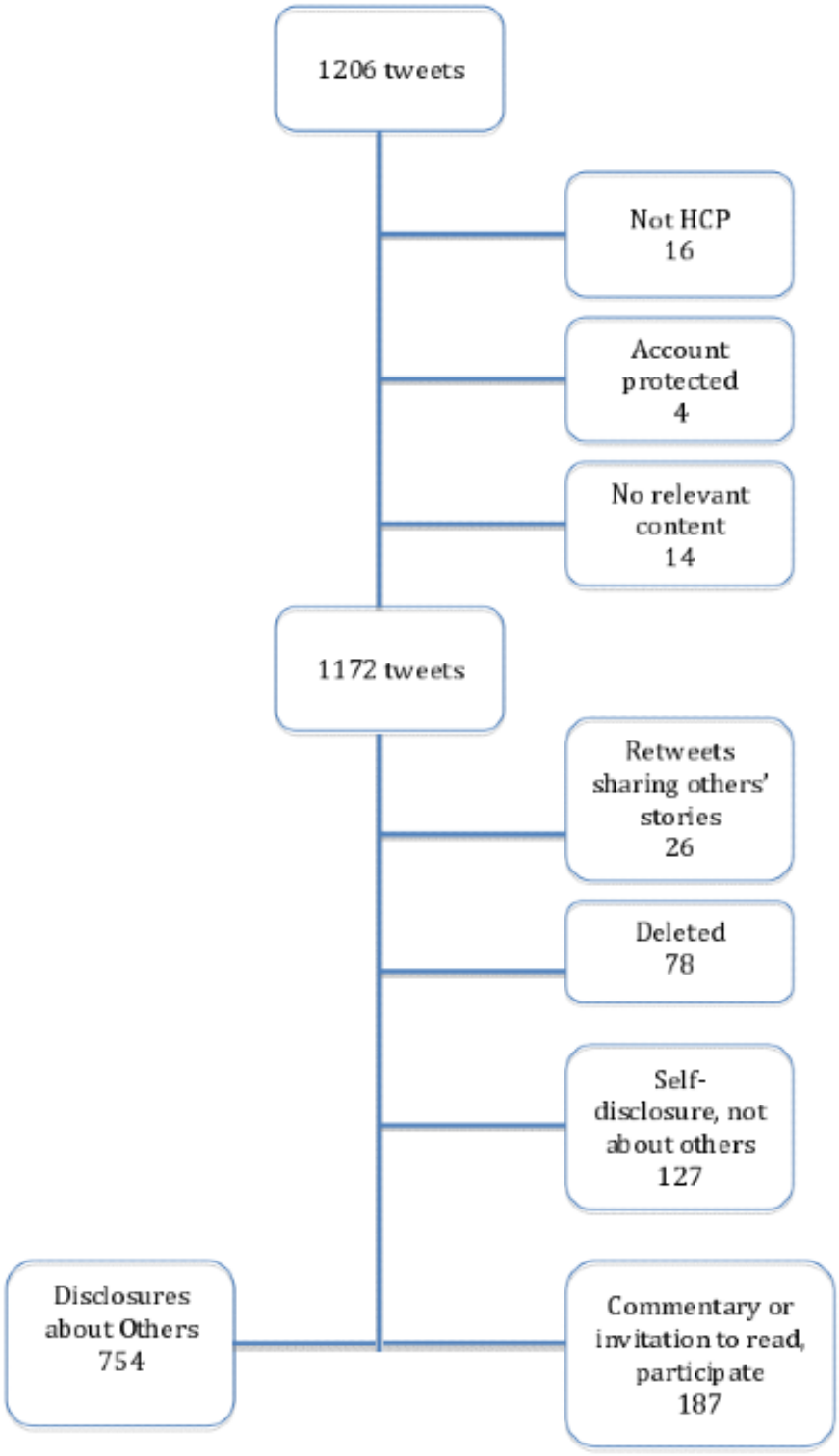
Because all information about the published content was publicly accessible, Lowell General Hospital approved this study as institutional review board exempt. To evaluate content in the #ShareAStoryInOneTweet phenomenon, Symplur Signals (Symplur LLC), a proprietary health care–focused database and analytics program collecting data on Twitter using its Enterprise application program interface, was utilized [19]. The first tweet with the hashtag occurred May 4, 2018. From May 4, 2018 through December 31, 2018, 45,040 tweets that included the hashtag #ShareAStoryInOneTweet were identified. The study focused on 43,374 tweets shared in the month of May 2018 (midnight May 1 to midnight June 1, Eastern Standard Time). The analysis was conducted in January 2019. Using the software program, we identified tweets from doctors, patients, and other health care stakeholders (eg, caregivers, pharmaceutical firms, academic, or research organizations), based upon public self-identification in their Twitter profiles (ie, by identifying information provided within their Twitter biography) [20]. There were 4871 tweets identified, of which, 1206 were unique (the remainder represented retweeting of prior postings).

Unique tweets were reviewed by reading text provided within the data set and then evaluating the URL and each account's public profile on Twitter's website as of March 2019. Tweets from students misclassified as health care providers (1.3%) (eg, those listing "future doctor" in profile), from blocked accounts (0.3%), or with no relevant content posted with the hashtag (1.2%) were excluded, leaving a total of 1172 tweets shared by physicians, nurses, or other health professionals (Figure 1). The study also excluded 26 retweets in which the authors used the hashtag to share someone else's disclosure rather than their own, and 127 that included the author's own illness experiences rather than those of others. The study also excluded 78 tweets (6.5%) with content found in the data set but deleted from Twitter when evaluated on the website.

Characteristics of the health professionals sharing these tweets was examined, using information publicly available in their online profiles, including profession, gender, and country. Physicians were also categorized by specialty as described in their profiles or as unknown if not stated. More detailed content analysis focused upon the tweets in which the health care professional shared the illness or clinical experience of someone other than themselves.

The study also evaluated tweets commenting upon the hashtag-related phenomenon or recommending participation to others. The study analyzed tweets individually rather than as content threads.

Figure 1. Tweets by doctors and other health care providers during #ShareAStoryInOneTweet in May 2018. HCP: health care providers.



Measures

To assess the magnitude of hashtag use, the study evaluated total number of tweets. We also evaluated the number of total participants, focusing upon physician, nurse, and other health care professionals. We calculated tweets per hour, number of tweets, and use of images both in aggregate and by health care stakeholder categories. In order to capture hourly tweet activity rates, we restricted the time frame to the first two weeks starting May 4th to focus on the viral period. We evaluated the potential reach of the tweets using the software's definition of *impressions*—follower count at the time of each tweet's publication online (eg, a doctor posting while having 500 followers was equal to 500 impressions).

For each tweet, we coded several measures: the tweet author's role in the other person's clinical care; whether the patient died or was dying; whether the author helped save the other person's life; inclusion of patient name; inclusion of a clinical image;

and inclusion of a specific age. We categorized the time frame of the event described within a tweet as within the past year, 1-2 years ago, 2-5 years ago, >5 years, or unknown.

Whether either a patient or the patient's family or friends would be able to identify the clinical scenario described in each tweet was categorized broadly in response to codebook question "Could patient or family potentially identify the clinical situation?" as *yes* (more likely than not to be identifiable) or *no* (not likely to be identifiable). If it was unclear, the code *indeterminate* was applied and was considered in analyses to be no.

One author (MK) assessed all tweets; a second author (WA) coded a 10% sample independently. Intercoder reliability and percentage agreement were assessed using ReCal [21]. The two authors then reached consensus on discrepancies and used this exercise to identify any areas where the first coder might systematically have erred.

Because the tweets could be discoverable in malpractice or tort suits, we also analyzed whether the author made comments with a negative opinion about the patient or family, or if the author acknowledged that a medical error occurred. We also assessed whether information was shared about vulnerable patients, as defined by the US Department of Health and Human Services [22].

We separately evaluated tweets commenting upon the hashtag to determine whether the authors had a favorable or unfavorable opinion of the viral sharing, or if they invited others to share stories or to participate. We also identified whether these tweets expressed any concern about privacy breaches.

Statistical Analysis

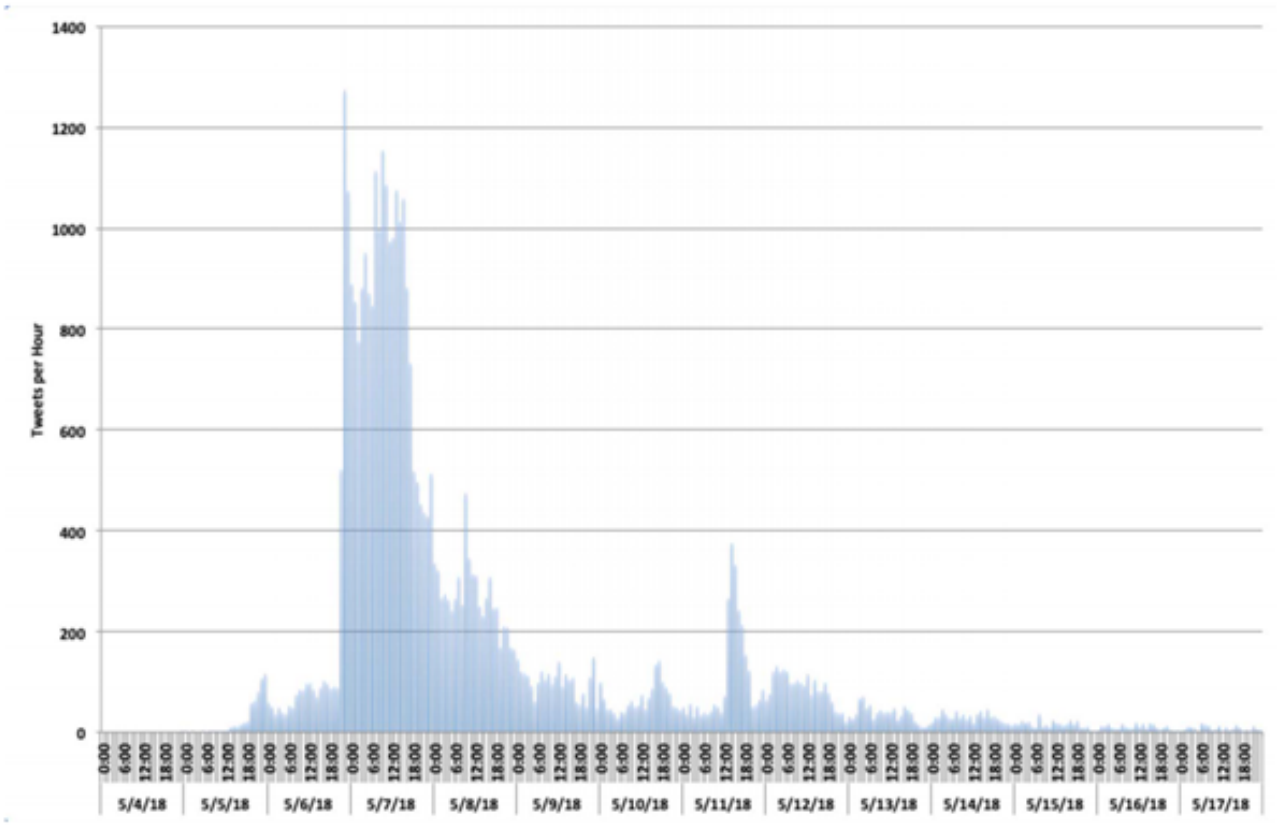
Overall activity and frequencies for stakeholder participation using Symplur. Frequencies, median, and mean endpoints for content analysis were calculated using Excel (for Mac 2011 version 14.7.2, Microsoft Inc). Cohen κ was used to measure interrater reliability [23].

Results

Tweet Volume

For May 2018, we identified 31,690 individuals who posted tweets with the hashtag, with a potential of 106.5 million views; 1725 (5.3%) individuals self-identifying as doctors and 861 (2.6%) as other health care providers shared tweets. At its peak, activity showed 1274 tweets per hour (Figure 2).

Figure 2. Tweets per hour including the hashtag #ShareAStoryInOneTweet in May 2018.



Tweets With Disclosures About Others

The characteristics of the health professionals sharing tweets with disclosures about others are presented in Table 1. Of the 656 health professionals, 384 (58.5%) were female; physicians—emergency medicine, family medicine or general practice, and hematology-oncology were the specialties most frequently represented—constituted the largest proportion of the tweeters (486/656, 74.1%), with nurses representing a minority (98/656, 14.9%); and most were in the United States (347/656, 52.9%), followed by Canada (99/656, 15.1%) and the United Kingdom (82/656, 12.5%).

The majority (659/754, 87.4%) involved the sharing of stories about direct patient care, rather than the author's role as a family caregiver or in another role (Table 2), and 13.6% (95/754) of tweets included a specific time frame. The patient's age was included in 163 of the 754 tweets (21.6%), and patient name was included in 15 of the 754 (2.0%). Of the 754, 11 tweets (1.5%) shared a clinical image, 152 tweets (20.2%) shared information about people in a vulnerable category. Only, 2 tweets (0.3%) mentioned patient consent to share within the tweet, one explicit and one inferred from past patient agreement to share a specific story. Based upon the number of likes, a minimum of 154,900 accounts viewed these 754 tweets.

Nearly half of the tweets (337/754, 44.7%) described a clinical scenario involving death or dying. Comments disclosing medical errors (6/754, 0.8%) or expressing a negative opinion about the patient or family were rare (4/754, 0.5%). Agreement between coders was 86.8%, and intercoder reliability Cohen κ=0.8 suggested substantial agreement [16]. Disagreements in coding occurred mostly between the categories no unclear, which led to the decision to combine the categories for further analysis. We estimated that almost one-third (242/754, 32.1%) of families or friends would likely find the content in the tweet identifiable. Among patients who were potentially still living, the study estimated that nearly half (162/348, 46.6%) contained likely identifiable information, of which 81 (50%) were likely identifiable by families and friends. The 754 tweets received a median of 2 retweets (range: 0-19; total 959) and 16 likes (range: 0-56; total 735).

Table 1. Characteristics of doctors and health care professionals sharing tweets disclosing information about others.

Characteristics	Value (N=656), n (%)
Gender	
Female	384 (58.5)
Male	266 (40.5)
Unknown	6 (0.9)
Profession	
Doctor	**486 (74.1)**
Anesthesia	22 (4.5)
Cardiology	25 (5.1)
Critical care	12 (2.5)
Emergency medicine	77 (15.8)
Family medicine-general practitioner	48 (9.9)
Gastroenterology	6 (1.2)
Hematology-oncology	39 (8.0)
Hospitalist	5 (1.0)
Infectious disease	5 (1.0)
Internal medicine	12 (2.5)
Neurosurgery	5 (1.0)
Obstetrics and gynecology	13 (2.7)
Palliative care	18 (3.7)
Pathology	9 (1.9)
Pediatrics	28 (5.8)
Psychiatry	7 (1.4)
Pulmonary medicine	7 (1.4)
Radiation oncology/clinical oncology	10 (2.1)
Radiology	6 (1.2)
Surgery	19 (3.9)
Trauma surgery	20 (4.1)
Unknown	36 (7.4)
Other	57 (11.7)
Nurse	**98 (14.9)**
Nurse, not otherwise specified	66 (67.3)
Critical care	16 (16.3)
Emergency medicine	8 (8.2)
Other	8 (8.2)
Nurse practitioner	12 (1.8)
Paramedic	18 (2.7)
Pharmacist	7 (1.1)
Physical therapy	10 (1.5)
Psychologist	5 (0.8)
Social worker	7 (1.1)
Speech therapy	5 (0.8)
Other	8 (1.2)

Characteristics	Value (N=656), n (%)
Country	
Australia	11 (1.7)
Canada	99 (15.1)
India	7 (1.1)
Ireland	16 (2.4)
Saudi Arabia	7 (1.1)
South Africa	4 (0.6)
Spain	4 (0.6)
United Kingdom	82 (12.5)
United States	347 (52.9)
Unknown	55 (8.3)
Other	24 (3.7)

Table 2. Content characteristics tweets with disclosures about others.

Content characteristic	Value (N=754), n (%)
Author role	
Health care professional	669 (88.7)
Patient	0 (0.0)
Caregiver	42 (5.6)
Other	43 (5.7)
Time frame described	
Within past year	5 (0.7)
1-2 years	6 (0.8)
2-5 years	5 (0.7)
> 5 years	79 (10.5)
Unknown/not described	659 (87.4)
Content	
Author involved in patient care	635 (84.2)
Dying patient or patient death	337 (44.7)
Saving a patient’s life	131 (17.4)
Include a clinical image	11 (1.5)
Include patient name	15 (2.0)
Provide specific patient age	163 (21.6)
Express negative opinion of patient or family	4 (0.5)
Mention medical error	6 (0.8)
Estimated likely	
Can family or friends identify situation described?	242 (32.1)
Can patient identify situation described?	
All tweets	183 (24.3)
Potentially living patients (n=348)	162 (46.6)
Vulnerable population	152 (20.2)

Tweets Relating to the Hashtag

Of 187 tweets actively part of the conversations without disclosures, 173 made some commentary: 6 tweets (3.2%) raised concerns about privacy or identifiable information in the tweets with disclosures, 1 (0.6%) tweet involved another critical comment, and 167 (96.5%) tweets were neutral or favorable. Of 187 tweets, 42 tweets (22.5%) invited others to read the hashtag's stream or contribute to it.

Discussion

Principal Findings

This retrospective study describes a physician-initiated event sharing health-related stories and information on Twitter by quantifying the global participation of health care professionals and the type of content shared. The tweeted stories became widely shared, attracting media attention and disseminating the information widely. Almost none (either explicitly or appear to) confirm consent to share information publicly on the popular social network. Nurses, physicians, and other health professionals commenting using the hashtag were more likely to express support for the event and encourage others to participate than they were to raise concerns about patient privacy breaches. However, recent research suggests that 12% of patients may have less trust in physicians describing patient stories on social media, even if shared respectfully [24].

The study showed a relatively high incidence of sharing stories including details that might make them potentially identifiable to patients themselves or to families and friends in a setting that involved a large number of health care professionals. This finding highlights a lack of awareness about the privacy issues intrinsically connected to interactions on social media. Early in the use of social media, most US state medical boards received at least one report of an episode of online professionalism violations for disciplinary action, including violations of patient confidentiality [25]. Although surveys of medical students and physicians suggest the incidence of unprofessional behavior among medical students is infrequent [26,27], this study indicates that in some circumstances health care professionals may share more information publicly than the public might expect. Privacy breaches risk potential negative effects on physician-patient relationships, professional disciplinary actions or torts, and eroding public trust.

The findings of this study differ from those of prior studies [16,17] of online medical professionalism at least partly because we analyzed, in detail, one specific event focused upon health-related disclosures. There is no indication this episode was planned, and the incidence of similar episodes is unknown. However, it was not an isolated event; for example, another prominent example involved physicians opposing gun violence, who used the hashtag #ThisIsOurLane on Twitter in November 2018 [28]. Physicians focused on policy issues, but some may have failed to recognize privacy concerns, publishing tweets with photographs similar in nature to prior social media content in other cases, resulted in professional termination [29,30]. Social media studies publishing tweets often permit reverse identification of the authors [31], and a survey suggests that most participants are somewhat or very uncomfortable with their tweets being quoted in a published research paper [32]. It was found that 6.5% of tweets archived in the software's archived data set were in fact deleted by health care professionals, indicating that some did not want their tweets to remain publicly visible. Even if deleted online, tweets may retain permanence and discoverability, when published in journals.

Most research evaluating online disclosures focus on the privacy paradox, in which people value their privacy but still share their own information. Surveys indicate people may value short-term social rewards of self-disclosure online more than long-term privacy concerns [33], and high social capital of social network users is associated with increased self-disclosures over time [34]. For people disclosing information about others, the research is more limited, but opinion leadership and female gender have been linked to less concern about others' privacy [35], consistent with the findings of this study. Health care professionals may be prone to these same tendencies, despite their training and education to maintain privacy. Generational differences in concerns about privacy online may also play a role [36], but assessment of this possibility was not within the scope of this study.

Based upon the temporal pattern of sharing, this hashtag-related event may be less similar to narrative medicine and writing and more similar to a brief episode of social contagion, in which viral sharing of content or emotions online may occur and involve more than simple, conscious risk–reward tradeoffs [37,38]. Unlike traditional peer-reviewed publication of a medical story in narrative medicine, tweeting occurs quickly and does not permit editing. The observation that 6.5% chose later to delete their contributions may suggest that some health care professionals who participated in the experience may have later viewed their behavior as a temporary lapse in judgment.

Another contributing factor may be a knowledge gap for physicians and other health care professionals on how to behave online. While many recognize the importance of online professionalism, curricula for use in formal medical education are only beginning to emerge and remain uncommon [39-41]. Of note, the ethical obligation to maintain confidentiality does not end with a patient's death [42]. The digital medium does not avoid the potential that disclosures about patients risk breaching confidentiality, undermining trust within that therapeutic relationship as well as public trust in the medical profession. The findings of this study suggest a potential need for evidence-based training in ethical digital communications skills for undergraduate, graduate, and continuing medical education. Professional societies could create resources that allow social media authors to document having obtained consent, so that disclosing identifiable patient information without consent does not inadvertently become normalized.

Limitations

This study had several limitations. First, this study examined a very specific event that may occur during very active periods of online engagement but likely overestimated the general incidence of online behaviors that could, in some cases, constitute violations of medical professionalism. Future research could analyze a broader collection of social media posts by

medical professionals. Second, the study could not assess the number of people actually seeing these tweets; only the number of likes was measured, and the potential reach was estimated. Third, the study only analyzed tweets from accounts that the software identified as health care professionals. The evaluation of all tweets in the cohort confirms the software rarely misclassified nonprofessionals into this group, but the study did not evaluate any other participants in the event to determine if the study could identify more participating health care professionals not categorized as doctors or nurses by the software, which could decrease or increase the incidence of potential privacy breaches. Fourth, by analyzing only tweets with the hashtag, the study potentially underestimated the frequency of others expressing concern about patient privacy. Fifth, given the brevity inherent to the medium of Twitter, it is possible that some authors did indeed have formal documentation of patients' consent to share their stories but that there was insufficient room to include due to character limits in each post. Finally, the assessment of identifiability in this study may differ from those in other studies, and we cannot exclude the possibility that some physicians and nurses tweeting what seemed to be identifiable stories consciously changed important details to deidentify. It was beyond the scope of this study to confirm whether any harm occurred.

Despite these limitations, the findings of this study clearly show that internet-based sharing raises potential pitfalls for medical professionalism. The internet provides nurses, physicians, and other professionals the opportunity to help or harm others on a global scale. Although internet culture may favor maximizing transparency, it can also pose the risk of directly contradicting health professionals' fiduciary duty: first, do no harm, including harm that may be inflicted by what we say.

Conclusion

The study identified a high incidence of potential privacy breaches online. More research is essential to confirm the findings of this study and determine how to ensure physicians, nurses, and other professionals adapt their behavior to maintain medical professionalism in the digital age. Our results suggest that some who were using the hashtag may not have appreciated that the information being shared might breach patients' privacy We recommend greater specification of professional ethical standards in this context along with evidence-based training in ethical digital communications skills for the undergraduate, graduate, and continuing medical education.

Conflicts of Interest

WA and TGG have nothing to disclose. MSK reports common stock ownership in Dr. Reddy's Laboratories, Healthcare Services Group, Mazor Robotics, and US Physical Therapy. RJ has stock options as compensation for an advisory board role in Equity Quotient, a company that evaluates culture in health care companies. RJ has received personal fees from Amgen and Vizient and grants for unrelated work from the National Institutes of Health, the Doris Duke Foundation, the Greenwall Foundation, the Komen Foundation, and Blue Cross Blue Shield of Michigan for the Michigan Radiation Oncology Quality Consortium. RJ has a contract to conduct an investigator initiated study with Genentech. RJ has served as an expert witness for Sherinian and Hasso and Dressman Benzinger LaVelle. RJ is an uncompensated founding member of TIME'S UP Healthcare and a member of the Board of Directors of ASCO.

References

1. Brenan M. Nurses again outpace other professions for honesty, ethics. Gallup. 2018 Dec 20. URL: https://news.gallup.com/poll/245597/nurses-again-outpace-professions-honesty-ethics.aspx [accessed 2020-08-25]
2. Blendon RJ, Benson JM, Hero JO. Public Trust in Physicians — U.S. Medicine in International Perspective. N Engl J Med 2014 Oct 23;371(17):1570-1572. [doi: 10.1056/nejmp1407373]
3. Huang EC, Pu C, Chou Y, Huang N. Public Trust in Physicians—Health Care Commodification as a Possible Deteriorating Factor: Cross-sectional Analysis of 23 Countries. INQUIRY 2018 Mar 05;55:004695801875917. [doi: 10.1177/0046958018759174]
4. Office for Civil Rights. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act. Health Information Privacy. 2016 Nov 06. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [accessed 2020-08-25]
5. Thompson LA, Dawson K, Ferdig R, Black EW, Boyer J, Coutts J, et al. The intersection of online social networking with medical professionalism. J Gen Intern Med 2008 Jul;23(7):954-957 [FREE Full text] [doi: 10.1007/s11606-008-0538-8] [Medline: 18612723]
6. Chretien KC. Online Posting of Unprofessional Content by Medical Students. JAMA 2009 Sep 23;302(12):1309. [doi: 10.1001/jama.2009.1387]
7. Ofri D. The Passion and the Peril. Academic Medicine 2015;90(8):1005-1006. [doi: 10.1097/acm.0000000000000672]
8. Wells DM, Lehavot K, Isaac ML. Sounding Off on Social Media. Academic Medicine 2015;90(8):1015-1019. [doi: 10.1097/acm.0000000000000668]
9. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
10. Rosenbaum L. Twitter Tailwinds — Little Capsules of Gratitude. N Engl J Med 2018 Jul 19;379(3):209-211. [doi: 10.1056/nejmp1806737]

11. Ahmed W. Public health implications of #ShoutYourAbortion. Public Health 2018 Oct;163:35-41. [doi: 10.1016/j.puhe.2018.06.010]
12. American Medical Association. Professionalism in the Use of Social Media. American Medical Association. 2016 Nov 14. URL: https://www.ama-assn.org/delivering-care/ethics/professionalism-use-social-media [accessed 2025-08-25]
13. MMS Physicians' Guide to Social Media. Massachusetts Medical Society. 2015. URL: http://www.massmed.org/Governance-and-Leadership/Committees,-Task-Forces-and-Sections/MMS-Physicians--Guide-to-Social-Media-(pdf)/ [accessed 2020-08-25]
14. Lagu T, Kaufman EJ, Asch DA, Armstrong K. Content of Weblogs Written by Health Professionals. J GEN INTERN MED 2008 Jul 23;23(10):1642-1646. [doi: 10.1007/s11606-008-0726-6]
15. Chretien KC, Kind T. Social Media and Clinical Care. Circulation 2013 Apr 02;127(13):1413-1421. [doi: 10.1161/circulationaha.112.128017]
16. MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. Med Educ 2010 Aug;44(8):805-813. [doi: 10.1111/j.1365-2923.2010.03720.x] [Medline: 20633220]
17. Rimmer A. Doctors' use of Facebook, Twitter, and WhatsApp is the focus of 28 GMC investigations. BMJ 2017 Sep 01:j4099. [doi: 10.1136/bmj.j4099]
18. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 6;22(5):e19458. [doi: 10.2196/19458]
19. Authentication. Symplur. URL: https://docs.symplur.com/reference#authentication [accessed 2020-08-25]
20. Utengen A. Health care Stakeholder Segmentation. Symplur. URL: https://help.symplur.com/en/articles/103684-health [accessed 2020-08-25]
21. Freelon D. ReCal: Intercoder reliability calculation as a web service. Int J Internet Science 2010;5(1):20-33. [doi: 10.4135/9781412963947.n228]
22. Vulnerable Populations. Office for Human Research Protections. URL: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/vulnerable-populations/index.html [accessed 2020-08-25]
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
24. Fatollahi JJ, Colbert JA, Agarwal P, Lee JL, Lehmann EY, Yuan N, et al. The Impact of Physician Social Media Behavior on Patient Trust. AJOB Empir Bioeth 2020 Oct 30;11(2):77-82. [doi: 10.1080/23294515.2019.1678533] [Medline: 31663810]
25. Greysen SR, Chretien KC, Kind T, Young A, Gross CP. Physician violations of online professionalism and disciplinary actions: a national survey of state medical boards. JAMA 2012 Mar 21;307(11):1141-1142. [doi: 10.1001/jama.2012.330] [Medline: 22436951]
26. Thompson LA, Black E, Duff WP, Paradise Black N, Saliba H, Dawson K. Protected health information on social networking sites: ethical and legal considerations. J Med Internet Res 2011 Jan 19;13(1):e8 [FREE Full text] [doi: 10.2196/jmir.1590] [Medline: 21247862]
27. Brynolf A, Johansson S, Appelgren E, Lynoe N, Edstedt Bonamy A. Virtual colleagues, virtually colleagues--physicians' use of Twitter: a population-based observational study. BMJ Open 2013;3(7):1-5 [FREE Full text] [doi: 10.1136/bmjopen-2013-002988] [Medline: 23883885]
28. Ranney ML, Betz ME, Dark C. #ThisIsOurLane — Firearm Safety as Health Care's Highway. N Engl J Med 2019 Jan 31;380(5):405-407. [doi: 10.1056/nejmp1815462]
29. ER Doc forgets patient information is private, gets fired for Facebook overshare. Above the Law. 2011 Apr 29. URL: https://abovethelaw.com/2011/04/er-doc-forgets-patient-info-is-private-gets-fired-for-facebook-overshare/ [accessed 2020-08-25]
30. 'New York Med' nurse fired for 'insensitive' Instagram shot. New York Post. 2014 Aug 14. URL: https://nypost.com/2014/07/08/new-york-med-nurse-katie-duke-fired-for-insensitive-instagram-shot/ [accessed 2020-08-25]
31. Ayers JW, Caputi TL, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. NPJ Digit Med 2018 Aug 2;1(1):30 [FREE Full text] [doi: 10.1038/s41746-018-0036-2] [Medline: 31304312]
32. Fiesler C, Proferes N. "Participant" Perceptions of Twitter Research Ethics. Social Media + Society 2018 Mar 10;4(1):205630511876336. [doi: 10.1177/2056305118763366]
33. Hallam C, Zanella G. Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards. Computers in Human Behavior 2017 Mar;68:217-227. [doi: 10.1016/j.chb.2016.11.033]
34. Trepte S, Reinecke L. The reciprocal effects of social network site use and the disposition for self-disclosure: A longitudinal study. Computers in Human Behavior 2013 May;29(3):1102-1112. [doi: 10.1016/j.chb.2012.10.002]
35. Koohikamali M, Peak DA, Prybutok VR. Beyond self-disclosure: Disclosure of information about others in social network sites. Computers in Human Behavior 2017 Apr;69:29-42. [doi: 10.1016/j.chb.2016.12.012]
36. Pereira S, Robinson JO, Peoples HA, Gutierrez AM, Majumder MA, McGuire AL, et al. Do privacy and security regulations need a status update? Perspectives from an intergenerational survey. PLoS One 2017 Sep 19;12(9):e0184525 [FREE Full text] [doi: 10.1371/journal.pone.0184525] [Medline: 28926626]
37. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Statist. Med 2012 Jun 18;32(4):556-577. [doi: 10.1002/sim.5408]

38. Alshamsi A, Pianesi F, Lepri B, Pentland A, Rahwan I. Beyond Contagion: Reality Mining Reveals Complex Patterns of Social Influence. PLoS One 2015 Aug 27;10(8):e0135740 [FREE Full text] [doi: 10.1371/journal.pone.0135740] [Medline: 26313449]
39. Langenfeld SJ, Vargo DJ, Schenarts PJ. Balancing Privacy and Professionalism: A Survey of General Surgery Program Directors on Social Media and Surgical Education. J Surg Educ 2016 Nov;73(6):e28-e32. [doi: 10.1016/j.jsurg.2016.07.010] [Medline: 27524278]
40. Gomes AW, Butera G, Chretien KC, Kind T. The Development and Impact of a Social Media and Professionalism Course for Medical Students. Teach Learn Med 2017 Mar 08;29(3):296-303. [doi: 10.1080/10401334.2016.1275971] [Medline: 28272900]
41. Flickinger TE, O'Hagan T, Chisolm MS. Developing a Curriculum to Promote Professionalism for Medical Students Using Social Media: Pilot of a Workshop and Blog-Based Intervention. JMIR Medical Education 2015 Dec 01;1(2):e17. [doi: 10.2196/mededu.4886]
42. Applebaum P, Gutheil T. Clinical Handbook of Psychiatry and the Law, Fourth Ed. In: Confidentiality and Privilege. Philadelphia, PA: Lippincott Wilkins & Williams; 2007:1-34.

Edited by G Eysenbach; submitted 30.04.20; peer-reviewed by MA Bahrami, C Jacob, A Louren, R Zowalla; comments to author 12.06.20; revised version received 06.07.20; accepted 23.07.20; published 01.09.20.

Please cite as:
Ahmed W, Jagsi R, Gutheil TG, Katz MS
Public Disclosure on Social Media of Identifiable Patient Information by Health Professionals: Content Analysis of Twitter Data
J Med Internet Res 2020;22(9):e19746
URL: https://www.jmir.org/2020/9/e19746
doi:10.2196/19746
PMID:32870160

Original Paper

Identifying Mobile Health Engagement Stages: Interviews and Observations for Developing Brief Message Content

Kara Burns[1], BSc (Hons), BASc, PhD; Rebekah Nicholas[2], BA (Hons), PhD; Amanda Beatson[1], BA, PhD; Marianella Chamorro-Koc[2], BA, MA, PhD; Alethea Blackler[2], BA (Hons), MA, PhD; Udo Gottlieb[1], BA (Hons), PhD

[1]School of Advertising, Marketing and Public Relations, QUT Business School, Queensland University of Technology, Brisbane, QLD, Australia
[2]QUT Design Lab, Creative Industries Faculty, Queensland University of Technology, Brisbane, QLD, Australia

Corresponding Author:
Kara Burns, BSc (Hons), BASc, PhD
School of Advertising, Marketing and Public Relations
QUT Business School
Queensland University of Technology
George St
Brisbane, QLD, 4000
Australia
Phone: 61 414294967
Email: drkaraburns@gmail.com

Abstract

Background: Interest in mobile health (mHealth) has increased recently, and research suggests that mHealth devices can enhance end-user engagement, especially when used in conjunction with brief message content.

Objective: This research aims to explore the *stages of engagement* framework for mHealth devices and develop a method to generate brief message content to promote sustained user engagement. This study uses the framework by O'Brien and Toms as a point of departure, where engagement is defined as the uptake or the use of an mHealth device. The framework is a linear repeatable process, including *point of engagement*, *period of engagement*, *disengagement*, and *re-engagement*. Each stage is characterized by attributes related to a person's technology experience. Although the literature has identified stages of engagement for health-related technology, few studies explore mHealth engagement. Furthermore, little research has determined a method for creating brief message content at each stage in this engagement journey.

Methods: Interviews and observations from 19 participants who used mHealth technologies (apps, devices, or wellness websites) in a solo capacity were recruited for sample group 1. In sample group 2, interviews, and observations from 25 participants using mHealth technologies in a group capacity through the Global Corporate Challenge were used. These samples were investigated at 3 time points in both research contexts. The results underwent deductive-inductive thematic analysis for the engagement stages' framework and attributes.

Results: In addition to the 4 stages identified by O'Brien and Toms, 2 additional stages, self-management and limited engagement, were identified. *Self-management* captures where users had disengaged from their technology but were still engaged with their health activity. *Limited engagement* captures where group mHealth users had minimal interaction with their mHealth technology but continued to engage in a group fitness activity. The results revealed that mHealth engagement stages were nonlinear and embedded in a wider engagement context and that each stage was characterized by a combination of 49 attributes that could be organized into 8 themes. Themes documented the total user experience and included technology usability, technology features, technology aesthetics, use motivations, health awareness, goal setting, social support, and interruptions. Different themes were found to have more relevance at different engagement stages. Knowing themes and attributes at all engagement stages allows technology developers and health care professionals to generate relevant brief message content informed by a person-centered approach.

Conclusions: This research extends an existing engagement stages framework and identifies attributes and themes relevant to mHealth technology users' total user experience and incorporates concepts derived from health, business studies, and information systems literature. In addition, we offer a practical 5-step process based on a person-centered approach to develop mHealth technology brief message content for sustained engagement.

(J Med Internet Res 2020;22(9):e15307) doi:10.2196/15307

KEYWORDS

mobile health; text messaging; social media; mobile phone; health communication

Introduction

Mobile Health

Since 2013, mobile health (mHealth) or the use of mobile computing and communication technologies in health care, public health, and personal wellness has gained significant interest, including the use of SMS through mobile devices. This is because of efforts to engage patients [1] and improve health outcomes at lower costs [2]. mHealth is considered to be a subset of the broader eHealth movement, which involves the digitization of health care processes [3]. Although eHealth and mHealth are potentially transformative for health care users, population health, and health care systems, the benefit of the significant economic investment in these modalities is undergoing scrutiny. For example, a 2011 review of the billions of dollars spent on eHealth systems globally has shown no evidence that implementing these technologies reduces health service costs [4].

When investigating clinical impact, early randomized control trials focused on mHealth feasibility. Current research has shown mHealth to be effective for improving health outcomes, especially when using brief messages such as those delivered by SMS through personal devices (eg, smartphones). Brief messages are cost-effective and can be delivered en masse in low-resource areas and can be personalized to target specific populations [5]. For example, women in low- and middle-income countries who received SMS support during pregnancy showed reduced chances of infant morbidity or mortality [6]. This is congruent with other SMS mHealth initiatives that have shown an increase in smoking cessation [7], improvements in the uptake of sexual health services using SMS reminders compared with controls [8], and reduced viral loads of patients with HIV when reminded to take antiretroviral medication [9].

For mHealth interventions that include a communication component, message content is critical to the effectiveness of the interventions. For example, in a 24-month randomized clinical trial for weight loss, targeted feedback from health care providers by SMS was shown to improve engagement behavior in self-reported energy and fat intake [10]. An important aspect of this study was that messages were personalized using the individual feedback provided by participants. Communication between health care providers and patients holds an immediate opportunity for improved clinical outcomes and patient engagement, emphasizing the importance of identifying suitable message content for this communication [5]. However, one of the key challenges faced by these health care services is how to understand patterns of engagement and keep users engaged [11].

mHealth Engagement

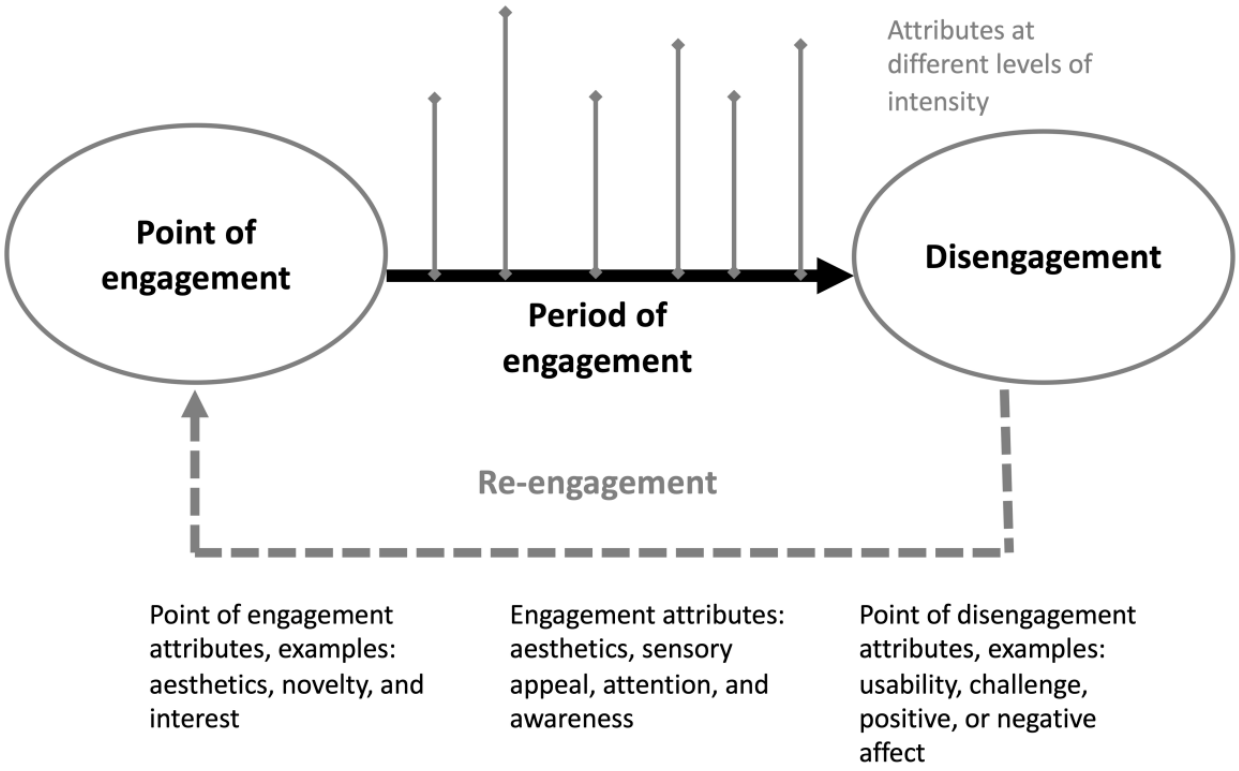
mHealth engagement is complex because of numerous definitions and the technology straddling the disciplines of health care, information systems, and business studies. In this research, engagement is defined as the uptake or use of an mHealth device [12]; however, limiting the exploration of mHealth to one discipline skews our understanding of the phenomena, resulting in a partial view of the impact of the technology and engagement.

Through information systems research, the usability and design of technology play a key role in the uptake of mHealth [13]. It is recommended that the usability and visual appeal of apps should be taken into consideration from the beginning of the design process [13]; however, there is also an opportunity for it to be critiqued based on user feedback throughout the life of the mHealth intervention to support sustained engagement.

The business literature suggests that a user-centered experience is a key factor that contributes to engagement in service contexts, and this can also be applied to mHealth technology [14]. It is evident that there are different views on user experience. This paper defines the user experience as the total experience that a person obtains from all interactions with a health care provider and an mHealth device. This includes the value they receive [15-17]. The total user experience can be evidenced by end-user thoughts, feelings, and behaviors related to their technology use [18,19]. Value is critical to engagement and can be defined as a total assessment of the perceived cost [20] versus benefits [21,22]. If the effort to engage is greater than the total perceived value of the engagement, users may not be interested in an mHealth device.

An interdisciplinary framework that incorporates health, information systems, and business does not exist in the present literature for mHealth engagement. However, O'Brien and Toms [12] developed a linear conceptual framework for defining technology user engagement that was applied in the context of web-based technology apps and incorporates attributes from both business and information systems. Through empirical inquiry, these authors demonstrated 4 stages of engagement, including point of engagement, period of engagement, disengagement, and re-engagement, that can be experienced in a linear sequence and repeated. Figure 1 provides the model of engagement stages, as proposed by O'Brien and Toms [12].

Figure 1. Proposed model of engagement by O'Brien and Toms (2008).



As illustrated earlier, the point of engagement is when the user encounters the technology for the first time, the period of engagement is the periodic use of the technology, disengagement is characterized by a period without using the technology, and re-engagement is a reintroduction of the technology to a previous user. Each stage in the process is characterized by themes and attributes. Attributes are the phenomena experienced by end users when engaging with mHealth technology, and these can be analyzed into themes. Attributes also signify the total user experience and are used as a baseline for exploration in this research. Importantly, this conceptualization provides a standard framework that can be flexibly applied in different contexts to explore the determinants of engagement in an mHealth context [23].

The framework of O'Brien and Toms has been explored in other digital health investigations, often underpinning studies that seek to conceptualize context-specific engagement [24]. Others simply apply the existing stages of the engagement framework [25,26]. Only one known study has explored additional stages by examining the notion of unengagement [11]. In this 2019 study of 486 health services' users who accessed web-based health services [11], it was recognized that the *nonengagement* or *disengagement* were examples of unengagement. Thus, to date, the stages and attributes in the model by O'Brien and Toms require critique and further investigation, specifically within the context of mHealth, where exploration is limited.

With that in mind and using the framework by O'Brien and Toms as a departure point, this study aimed to develop a person-centered approach to generate relevant brief message content relating to engagement stages to promote sustained user engagement. When combined with users' thoughts and observations of actual use, the framework by O'Brien and Toms can provide a way to investigate the stages of engagement for mHealth devices [12]. This is commensurate with the business literature and other health technology interventions that "demand a user-centered and iterative approach to development, using mixed methods and in-depth qualitative research to progressively refine the intervention to meet user requirements" [27]. Hence, this study asks, "How can knowledge about users' mHealth device engagement stages inform the development of brief message content intended to promote sustained user engagement?"

Methods

Study Design

To gain rich insight, a qualitative approach was used to investigate the framework by O'Brien and Toms [28]. Qualitative studies allowed the researcher to observe or record behavior in its natural setting [28]. Interviews were used to explore participants' state of mind in-depth [29], and face-to-face observations provided an understanding of participants' behaviors [30].

In total, 2 sample groups were used in this research to deepen the understanding of mHealth technology use, avoiding the use of a single case to support findings. Participants were considered eligible if they had used an mHealth technology for at least 2 weeks before the interviews and observations (excluding pregnancy apps), were aged >18 years, were able to communicate effectively in English, and provided informed consent. Interviews and observations across the 2 sample groups were conducted until data saturation occurred [31]. This occurred in a total of 34 participants.

Samples 1 and 2 were contextually different, comparing mHealth engagement with solo (single person) technology use and mHealth engagement in a social (group) context. In both samples, a screening questionnaire, interviews, and structured face-to-face observations were undertaken. Questions included perceptions of use, value, behaviors, and engagement. Participants were asked about their program usage (either an individual program such as a tracking app for sample 1 or the group program for sample 2). Example questions included How has the program impacted your behavior toward wellness?, Can you please show me what you value about the program and explain why you have selected this?, and Why do you think your usage has dropped and/or increased? The full interview protocol is outlined in Multimedia Appendix 1. Ethics approval was obtained from the researcher's home university (University Human Research Ethics Committee), and the research was conducted in line with standard ethical guidelines and the National Statement on Ethical Conduct in Human Research [32]. No incentives were offered in this study.

Sample Group 1

Participants who used mHealth technologies (eg, mHealth apps or wellness websites) in an individual context for at least 2 weeks before the first interview were purposively recruited using a promotional flyer and snowball sampling in a large metropolitan city in Australia. Upon recruitment, participants attended in-depth confidential interviews and structured face-to-face observations using their mHealth technology in a private room or in a public location of their choice. Interviews and observations of the participant using the device were recorded by a researcher on a video camera. Participants were then invited to attend a second session 6 to 8 weeks after the first session and a final interview and observations at 8 to 10 months after the first session.

Sample Group 2

Participants who used mHealth technologies in a group context were purposively recruited from the Global Corporate Challenge (GCC), a 100-day workplace team pedometer competition. The program is designed to promote workplace health and engagement around the world by encouraging employees to work in teams completing a minimum of 10,000 steps per day. After 2 to 4 weeks of engagement, participants attended in-depth confidential interviews and structured face-to-face observations using their mHealth technology in a private room or in a public location of their choice. These were recorded using a video camera by a researcher. All participants were invited to attend a second session 8-10 weeks after the first session and a final interview and observation 3-5 weeks after the GCC was complete.

Data Analysis for Engagement Stages and mHealth Content

Interviews and structured participant observations from samples 1 and 2 produced verbal and visual data that were analyzed by the principal investigator. Interviews were the primary source of the research findings, whereas observations were used to confirm the data collected from the interviews and to show how participants engaged with their mHealth technology.

Transcribed interviews were read several times, and deductive-inductive thematic analysis was undertaken using Atlas.ti software [33,34]. The stages of engagement and attributes from the framework by O'Brien and Toms [12] were applied to the interviews and observations. Deductive analysis was used to evaluate the conceptual framework, and a general inductive analysis was used to discover new attributes and important themes based on attributes related to each stage of engagement [35]. When a priori attributes and stages of engagement were not present in the data, they were removed; when new ones emerged, they were added to the schema. To reduce researcher subjectivity bias, 2 experienced health engagement researchers completed inter-rater reliability checks on the data. The inter-rater reliability calculated using the Cohen kappa coefficient was 0.92 in SPSS Statistics 26 (IBM Corp). This result suggests a strong level of agreement between the coders. Finally, mHealth engagement data through face-to-face observations confirmed participant use, thus triangulating the engagement states through both interviews and observations.

To establish message content, the occurrence and salience of themes at each engagement stage were established. Increased repetition of a theme at an engagement stage suggested increased salience of the corresponding message content. Thus, the most common themes and codes at each stage of engagement were identified and arranged in a hierarchy of the most relevant to be used for personalized communication.

Results

Participants

A total of 19 participants attended in-depth confidential interviews and structured participant observations using their mHealth technology with a researcher in sample group 1 (individual context) at the first period. Of the 19 participants, 12 (63%) were females and 7 (37%) were males, with age range of 18 to 49 years. The participants were invited to be interviewed and observed at 3 periods. At the second period, of the 19 participants, 15 (79%) participants completed the interviews and observations, with 11 (58%) participants agreeing to a third interview. In sample group 2, (group context) 25 participants were recruited from the GCC. Of the 25 participants, 18 (72%) were females and 7 (28%) were males, with an age range of 18 to ≥70 years. Again, each of the participants was invited to partake in the 3 periods. In total, 96% (24/25) participants completed both the second and third periods of data collection. Full participant details for sample 1 and sample 2 are provided in Table 1. Examples of technologies used across both samples included Fitocracy, Strava, Garmin Watch, Fitbit, or simple pedometers, and these were accessed via smart watches, mobile phones, and advanced trackers.

Table 1. Participant details for sample 1 and sample 2.

Participants	Samples	Gender	Age groups (years)	Length of time participants had been using their digital health technology at start of their interviews
1	Sample 1	Female	40-49	3-6 months
2	Sample 1	Female	30-39	6 months to 1 year
3	Sample 1	Female	18-29	<1 month
4	Sample 1	Male	40-49	1-3 months
5	Sample 1	Male	18-29	2 years
6	Sample 1	Female	18-29	1-3 months
7	Sample 1	Male	40-49	2 years
8	Sample 1	Female	40-49	2 years
9	Sample 1	Female	18-29	1-3 months
10	Sample 1	Female	40-49	1-3 months
11	Sample 1	Male	30-39	6 months to 1 year
12	Sample 1	Female	18-29	6 months to 1 year
13	Sample 1	Female	18-29	6 months to 1 year
14	Sample 1	Male	30-39	6 months to 1 year
15	Sample 1	Male	18-29	6 months to 1 year
16	Sample 1	Male	30-39	2 years
17	Sample 1	Female	40-49	1-3 months
18	Sample 1	Female	40-49	6 months to 1 year
19	Sample 1	Female	30-29	1 year
1	Sample 2	Female	30-39	3 years
2	Sample 2	Female	30-39	2 years
3	Sample 2	Female	30-39	1 year
4	Sample 2	Male	50-59	4 years
5	Sample 2	Female	18-29	1 year
6	Sample 2	Female	50-59	1 year
7	Sample 2	Female	50-59	1 year
8	Sample 2	Female	40-49	1 year
9	Sample 2	Female	60-69	>6 years
10	Sample 2	Female	50-59	2 years
11	Sample 2	Female	40-49	1 year
12	Sample 2	Male	50-59	3 years
13	Sample 2	Male	30-39	2 years
14	Sample 2	Female	30-39	2 years
15	Sample 2	Male	40-49	1 year
16	Sample 2	Female	40-49	2 years
17	Sample 2	Female	30-39	2 years
18	Sample 2	Male	50-59	3 years
19	Sample 2	Female	30-39	3 years
20	Sample 2	Female	30-39	1 year
21	Sample 2	Male	40-49	3 years
22	Sample 2	Female	50-59	4 years

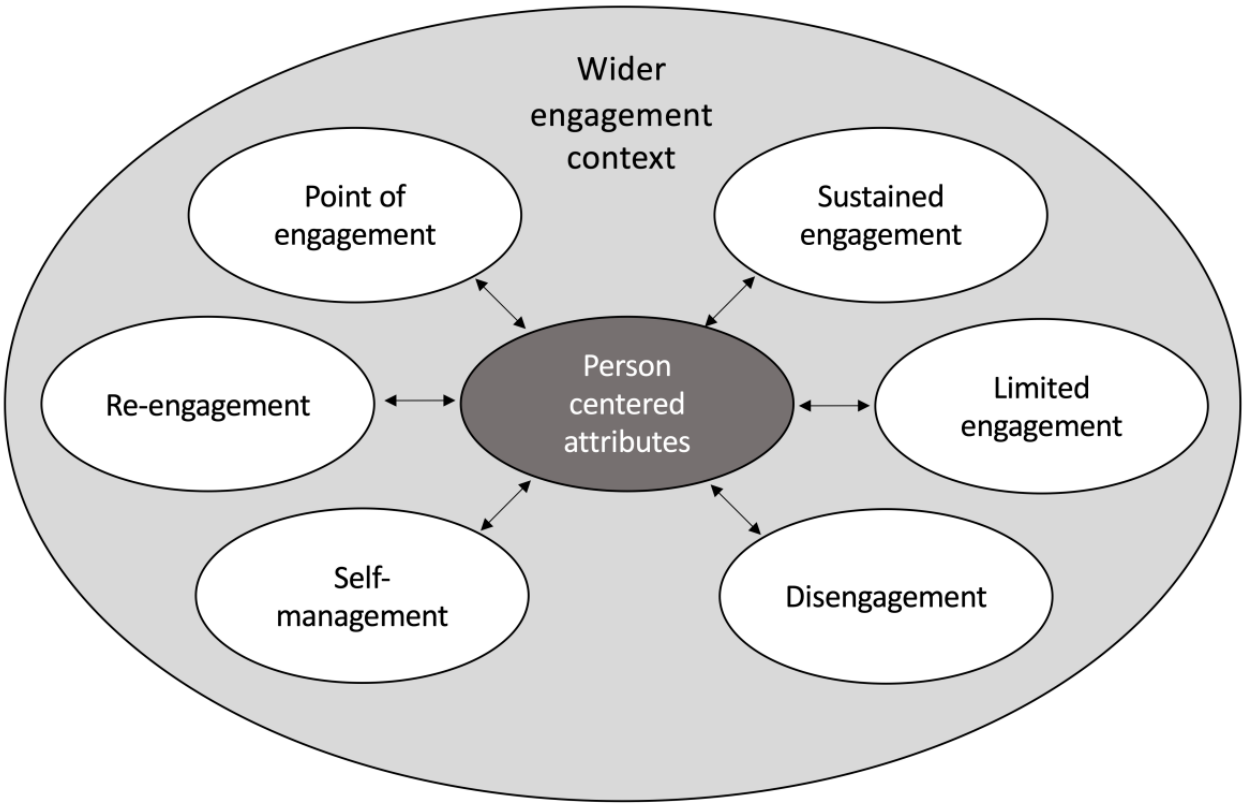
Participants	Samples	Gender	Age groups (years)	Length of time participants had been using their digital health technology at start of their interviews
23	Sample 2	Male	≥70	4 years
24	Sample 2	Female	40-49	4 years
25	Sample 2	Female	50-59	2 years

Engagement Stages

The framework by O'Brien and Toms [12] explored engagement in the context of web searching, web-based shopping, educational webcasting, and video games, finding 4 stages of engagement: point of engagement, period of engagement, disengagement, and re-engagement. Similarly, sample group 1 identified these 4 types of engagement plus a fifth engagement domain of *self-management*. Self-management refers to a stage where users had completely disengaged from their technology; however, they were still engaged with their health activity.

Furthermore, sample group 2 indicated an additional stage alongside self-management called *limited engagement*. This occurred when users had minimal interaction with their mHealth technology but engaged in the health activity to contribute to the group goals. Participants did not want to fully disengage as they felt the desire to stay in the program for others. Although limited engagement was discovered in the group context, it may also apply to solo engagement contexts. Thus, this research extends the framework by O'Brien and Toms [12] to include 6 stages of mHealth technology engagement, as seen in Figure 2. Importantly, these stages are not linear, with participants skipping stages based on motivation and accountability to engage, which are all features of a person's wider engagement context.

Figure 2. Mobile health engagement stages in this study (based on O'Brien & Toms 2008).



Themes at Each Engagement Stage

When developing message content for sustained engagement, this research determined 8 themes that represent the main discussions impacting the 6 engagement stages. Each engagement stage is characterized by a different combination of these themes and 49 attributes. Figure 3 indicates the themes relevant at each engagement stage. These themes are presented in a hierarchical order based on their level of importance for each engagement stage. For example, in the limited engagement stage, the 5 themes of interruptions, social support, specific goals, technology features, and use motivations contribute toward shaping the level of limited engagement for the participant. The most frequently discussed theme for limited engagement was interruptions, with the lowest being use motivations. This indicates that interruptions are the greatest consideration for users experiencing limited engagement with mHealth technology. Multimedia Appendix 2 provides a full

summary of the themes as they are expressed at each engagement stage.

Figure 3. Themes relevant at each engagement stage.

Point of engagement is determined by	Sustained engagement Is determined by	Limited engagement is determined by	Disengagement is determined by	Self-management is determined by	Re-engagement is determined by
• Technology • Feature • Use motivations • Social support • Technology usability • Health awareness • Specific goals	• Technology • Feature • Use motivations • Social support • Health awareness • Technology usability • Specific goals	• Interruptions • Social support • Specific goals • Technology feature • Use motivations	• Technology • Feature • Use motivations • Technology usability • Individual motivator • Interruptions • Social support	• Specific Goals • Interruptions • Use motivations • Technology feature	• Use motivations • Specific goals • Technology feature • Social support • Technology usability • Technology aesthetics • Health awareness

On the basis of Textbox 1 and Multimedia Appendix 3, the point of engagement is characterized by the user's attraction to the technical features of an mHealth device that helps them to fulfill their motivations to be healthy. At this stage, it is important that the technology is easy to use and promotes the support of a provider or a peer network. There was recognition that the mHealth device could improve users' health awareness and their ability to reach their goals; however, these are less important than the features of the technology. Technology aesthetics was not highly relevant at the point of engagement.

Textbox 1. Mobile health engagement stages and themes characterizing each stage.

- Point of engagement is characterized by technology feature, use motivations, social support, technology usability, health awareness, and goal setting
- Sustained engagement is characterized by technology feature, use motivations, social support, health awareness, technology usability, and goal setting
- Limited engagement is characterized by interruptions, social support, goal setting, technology feature, and use motivations
- Disengagement is characterized by technology feature, use motivations, technology usability, interruptions, and social support
- Self-management is characterized by goal setting, interruptions, use motivations, and technology feature
- Re-engagement is characterized by use motivations, goal setting, technology feature, social support, technology usability, technology aesthetics, and health awareness

Sustained engagement is characterized by the regular continued use of the mHealth device to support health and wellness activities. During this stage, technology features are important to achieve motivation, and this can be amplified through the support of a peer network or through provider messages. The continued use of an mHealth device encouraged users to have more health awareness. Usability and goal setting are still important in this stage but are less important than the features of the technology.

Limited engagement is characterized by users' interrupted use of the mHealth device, with participants driven by social motivations to continue with the health activity. Social support through a peer network or provider messages was considered by users as influential during this period to achieve their health goals. Interruptions experienced by users were often caused by a feature of the technology; however, social interactions provided the motivation to continue.

Disengagement is characterized by discontinuing the use of an mHealth device and participants stopping the health activity. This was often caused by the mHealth technology not being suitable for the user's needs and not providing motivation to continue. This mainly occurred because the mHealth device lacked features or was difficult to use. In addition, users did not receive the required social support from their device to develop sustained engagement. This was because of the inability of the mHealth technology to engage the user's peer network or support provider messages. Finally, other interruptions also caused disengagement.

Self-management is characterized by users who are able to stay motivated and achieve their goals without their mHealth technology. Interruptions caused by the technology features meant that an mHealth device was not appealing and no longer used. In addition, as users were already active and aware of their health status, they did not need their mHealth technology for health awareness.

Re-engagement is characterized by users wanting to be motivated and wanting to achieve their goals. During this period, the technology features and technology aesthetics or the visual appeal of a device were important in the decision to re-engage. When re-engaging with existing or new technology, users were

after specific technology features to help reach their health goals (eg, heart rate monitor and step counter). In addition, users were looking to re-engage with a mHealth technology that provides social support and helps them achieve health awareness.

Description of Each Theme

A total of 8 themes were established in interviews across the 2 samples: technology usability, features, technology aesthetics, use motivations, health awareness, goal setting, social support, and interruptions when engaging with mHealth technology. To develop message content, an understanding of the themes, their attributes, and the relevant stage of engagement is considered below.

Technology Usability

The usability of the mHealth wellness technology was discussed by most participants and was important at the points of engagement, sustained engagement, disengagement, and re-engagement stages. Fundamentally, this was related to functional congruence or the device performing in a way that was required by the wearer [36]. This finding is consistent with other studies that highlight the importance of usability [13,14]. Usability was specified as ease of use, wearability, portability, convenience, and connectivity. In addition, accuracy and customization were important features when participants engaged with mHealth technology at the point of engagement and to a lesser extent once they were engaged. Disengagement occurred when the device did not perform as expected; it could not be customized, resulting in a nonpersonalized experience or a feature of the technology failed. Many participants lamented that if they could not see their previous activity, this reduced motivation and increased effort to re-engage:

> *I like that it's simple to use, you just literally put it on your arm, load up on to your phone and just go, you don't have to do anything with it.* [Participant 9: female, aged 18-29 years, sample 1]

> *I like how the GCC and the Fitbit thing they sync with that device...I like that seamless syncing so that it already knows what's happened.* [Participant 20: female, aged 30-39 years, sample 2]

Technology Feature

The most relevant topic across both samples was the features of the mHealth technology, which was discussed by all participants and was present in each stage of engagement. Participants found value in the collection, monitoring, and review of health data, with it providing motivation to engage and use mHealth technology. Indeed, user-generated health data have been recognized to engage and empower patients in health care [37], and this also occurred in the wellness context. Gamification, messages, sounds, and dashboard tracking were features of devices that promoted continued use and the statistics provided through the Fitbit. However, if interruptions occur, for example, by a device having a short battery life, it could lead to disengagement and subsequent self-management. Although motivation to achieve goals was the predominant reason for re-engaging with technology, technology features also impacted the decision to re-engage:

> *...data, that was what led me to get this one because I wanted my heart rate and it gave me that and it gives me whatever data I want really on my sleep or my heart rate or exercise patterns, it can give me as much or as little as I want.* [Participant 1: female, aged 40-49 years, sample 1]

Technology Aesthetics

Despite the emphasis on design elements in the extant literature [14,15], this was only a minor theme discussed in interviews by one-third of the participants. Defined as an attractive interface, ergonomic design, and unobtrusive when worn, these attributes were only relevant at the re-engagement stage. Disengagement because of failure of a product feature affected future mHealth technology choices, and new offerings with improved design were considered to supersede predecessors. This was most evident in sample 1, where participants used the devices alone, whereas in sample 2, they were much more likely to re-engage based on perceptions of usability:

> *The new one is a bit thinner I think, this isn't too bad, I think it can still do the same as heart rate and all of that, but then I suppose after a while I might end up getting a thinner version.* [Participant 6: female, aged 18-29 years, sample 1]

Use Motivations

Although technology features were the most important aspect of mHealth adoption, motivation to use the technology was a heavily discussed theme at every stage of engagement. Motivation was considered key to sustained engagement, and although it is accepted that individual differences in motivational control exist [38], people generally have individual intrinsic motivators in sample 1, and in sample 2, extrinsic social motivators became more relevant. Being active, reducing stress, and having a balanced lifestyle were motivations to use an mHealth technology in the pursuit of achieving wellness and a long life. When using technology in sustained engagement, motivation was derived through technology features via encouragement to achieve challenges, competition through leader boards, and the sense of team obligations. In addition, achieving health goals left people feeling good about their technology experience and encouraged sustained use and re-engagement:

> *After the baby I wanted to get fit again and feel good about myself so either way I was going to be doing it. I just wanted something there, I could go back and have a look and see my fitness improving over time not just having to guess a particular thing.* [Participant 3: female, aged 18-29 years, sample 1]

> *...and on Facebook as well, we've got a chat group in messenger because all in our group are friends on Facebook, so sometimes we message each other on the weekend and say this morning I got 10,000 steps already or something like that. I think that when everyone else in the group sees that they feel like I'd better do something myself.* [Participant 15: male, aged 40-49 years, sample 2]

Health Awareness

Awareness was an attribute in the framework by O'Brien and Toms [26] and was prevalent in interviews, becoming the theme, health awareness, in this research. Engaging with an mHealth technology that results in greater awareness of health through the collection and analysis of health data is consistent with the quantified self-movement, defined as *self-knowledge through numbers* [39]. The theme was evident at the point of engagement and became more relevant in sustained engagement and was referenced again in re-engagement. Participants were motivated at the point of engagement by awareness of health and wellness goals and could objectively evaluate their performance when using the technology during sustained engagement, with re-engagement partly spurred by the awareness of previous health outcomes. This led to enhanced behavior recognition and increased personal accountability for health activities. This theme was not important at limited engagement, disengagement, or self-management, as these stages suggest low exposure to data collection and analysis:

> *So, I can actually see that I'm more active than what I originally thought because I thought I only do about 30 minutes of walking on the treadmill or whatever but now I'm realizing I'm doing an extra 6,500 steps.* [Participant 3: female, aged 18-29 years, sample 1]

Goal Setting

Although goal setting was noted at the point of engagement, sustained engagement, and limited engagement, it became the most relevant theme in self-management and important for re-engagement. Participants spoke of goals related to exercise, nutrition, sleep, weight, and heart rate before and during sustained use of the mHealth technology, although these were less important than technology features, use motivations, and technology usability. As technology was not present in self-management, goals became the focus of wellness activities. Finally, goals were an avenue to re-engage with their existing or new mHealth technology, with participants having more sensitivity to goals after their first experiences. This was because of an increased awareness of devices, the ability to self-quantify, and the previous experience of achieving goals:

> *It's simple to use and there's a lot of scope to how I can plan a run. I can get it to remind me every pretty much time, calories, the goal I reach.* [Participant 15: male, aged 18-29 years, sample 1]
>
> *...and I want to be fabulous at 50, so I thought this will encourage me to do a few extra steps and think about what I'm eating, so that's my goal for November.* [Participant 5: female, aged 18-29 years, sample 2]

Social Support

Social support was expressed in sample 1 as customer support and sample 2 as peer support. This was most relevant at the point of engagement, sustained engagement, limited engagement, and re-engagement. At the point of engagement, participants contacted customer support to understand a device feature for better use. Social support was relevant during sustained and limited engagement, including discussion of team interactivity and building a sense of community. Peer competition was an extrinsic motivator to do better, and when group goals were achieved, it enhanced the profile of the teams in the wellness community. When self-management was present and devices were no longer used, social support became less relevant; however, when re-engaging with technology, previous experiences of a social support community were influential in the uptake of new devices:

> *Maybe it's not so much to do with Strava but maybe it's just that there's a community of people that helps facilitate, maybe it's something to do with just that people can see what others are up to and follow them and share in their experiences.* [Participant 16: male, aged 30-39 years, sample 1]
>
> *I like the team. I like that idea of people being in a team together contributing, motivating each other.* [Participant 16: female, aged 40-49 years, sample 2]

Interruptions

Interruptions were a critical element in limited engagement, disengagement, and self-management stages. Interruptions were caused by an individual's engagement. Family factors, work time constraints, and illnesses impacted limited engagement. Technology failure included losing a device or being unable to access data, and this impacted disengagement and the decision to self-manage. Device features were also relevant, with low functionality resulting in a perception that devices were not holistic, leading to limited use and disengagement. Indeed, disengagement can occur in the limited engagement stage, even in the presence of social support, if an individual experiences significant interruptions:

> *Unfortunately, the battery goes very, very quickly and then these times where I can't hold the phone because I'm doing stuff so it's not monitoring my steps.* [Participant 12: female, aged 18-29 years, sample 1]
>
> *I think for me with kids I needed a program where I can go back three days and log what I'd done because I can get distracted or sidetracked.* [Participant 3: female, aged 18-29 years, sample 1]

Developing Message Content

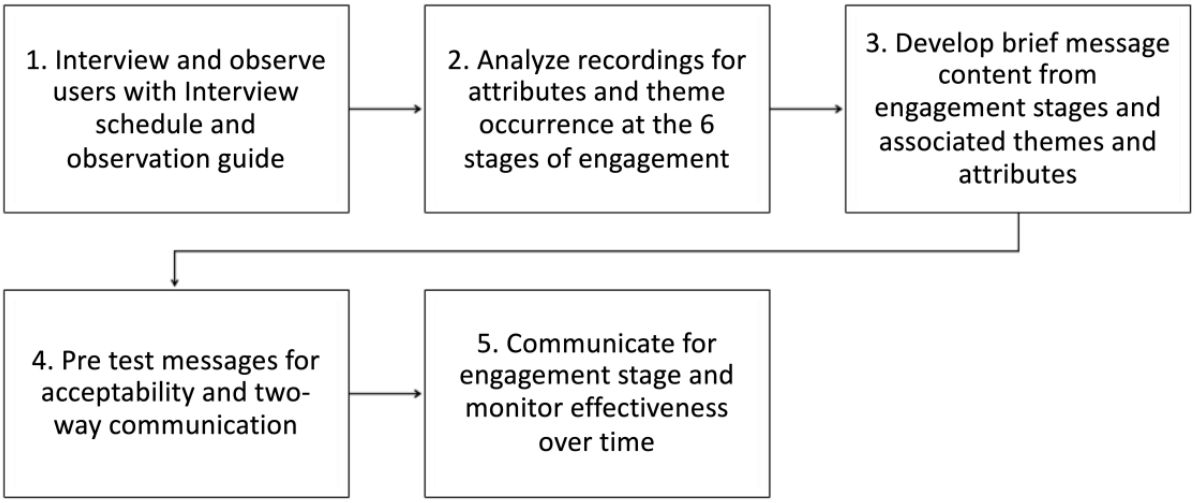
Multimedia Appendix 3 illustrates the full list of 49 attributes that emerged from the research relevant to each of the mHealth themes. These attributes highlight the key factors that determine the expression of each theme. By identifying the themes important at each engagement stage and then the relevant attributes, health care providers and mHealth developers can determine the type of content in messages that aim to sustain engagement.

For example, at the point of engagement, the most important theme that emerged was the technology feature. The attributes related to this theme are health data collection, dashboard tracking, messages, sounds, and device battery life. If health care providers such as insurers or personal trainers want to attract clients, the content of messages could focus on technology features such as the ability to track the user's data (dashboard tracking) or the use of health information (health

data collection). When users demonstrate limited engagement, messages can focus on the challenges of interruptions to health programs, such as competing the demands of family and providing some tips for how they can manage these difficult times. If users show re-engagement behaviors, messages could focus on reminding users of personal motivations for health management, focusing on specific small goals they may want to achieve.

Therefore, with engagement stages defined in this research and identification of key themes and attributes shaping people's engagement over time, this research recommends a 5-step process to help co-design appropriate content around these engagement stages for mHealth users. This co-designed 5-step process will help ensure that messages are meaningful with the aim of improving mHealth device engagement over time (Figure 4).

Figure 4. A five-step process to develop effective brief message content.



The 5-step process highlights the importance of first understanding the components necessary to group users into relevant engagement stages. By interviewing and observing users (step 1), organizations can identify which of the 6 engagement stages the mHealth user is at (step 2). This allows targeted communication relevant to the individual and their engagement stage to be developed, encouraging sustained engagement (step 3). The relevance of the message content can be drawn from the attributes identified. Step 4 involves pretesting the messages for acceptability by the user and to ensure the development of a two-way communication between the health care provider and the user. Without ensuring the suitability and targeting of the message content, health care providers run the risk that the target user will not pay attention to the message [40]. The final stage (step 5) involves the delivery of suitable messaging representative of the engagement stage the user is at, including monitoring the message success. This monitoring could come in the form of interactive responses via text messages, rebooking with health care staff, user-generated content entered into an app, or other such responses indicating engagement. This process is applicable across a broad application of mHealth relevant contexts, including but not limited to wellness initiatives and clinical contexts; however, programs require targeting and tailoring to an individual's context to ensure maximum communication success.

Discussion

The use of mHealth technology for patient engagement is enhanced by communication between health care providers and the technology user. To date, an engagement framework that can be flexibly applied in different contexts to develop message content has been under-researched in the extant literature. This research presents a model of 6 engagement stages: point of engagement, sustained engagement, limited engagement, disengagement, self-management, and re-engagement. Stages are characterized by 8 themes and 49 attributes, which signify the total user experience.

Importantly, the 6 engagement stages are not linear and have been reconceptualized using a person-centered approach, taking into account a user's wider engagement context. In a group context, mHealth devices can instigate limited engagement where the extrinsic motivation of the group performance is the only reason to engage. If limited engagement is recognized, however, message content can be tailored to increase the intrinsic motivation of goal setting to achieve sustained engagement. Similarly, after episodic wellness initiatives such as the GCC cessation, participants can still be engaged in health activities without the technology present. To pique interest and re-engagement in devices, technology aesthetics and features can be communicated to ex-users. Finally, people who disengage because of interruptions and find it difficult to re-engage may benefit from motivating message content that promotes limited engagement and social support as a goal before sustained engagement.

This research extends previous investigations into engagement stages, as initially proposed by O'Brien and Toms [12]. Although previous research has investigated the stages of engagement from a business and information systems

perspective [12], this study is one of the first investigations to specifically explore and extend these stages for an mHealth context. Within a health context, where the stages of engagement proposed by O'Brien and Toms [12] have been applied, there have been no developments in the existing stages or attributes, acknowledging the unique context [29,30]. Keeling et al [11] are the only researchers to further develop the stages proposed by O'Brien and Toms for a web-based health context; however, these researchers focused on investigating into digital engagement to understand why users do not engage with web-based health services. Although these researchers have contributed to knowledge about disengaged users, they observed that further research is needed into mHealth apps, as health conditions alter with time and research into touchpoints over time would be advantageous to better understand engagement. This study, to an extent, fulfills these shortcomings identified by Keeling et al [11].

The development of brief message content for mHealth engagement is a nascent research area, with existing studies focusing on specific providers or medical conditions. Indeed, many studies approach message development by creating content that emphasizes the providers' needs, rather than taking a person-centered approach, which involves understanding the wider context for the user [41], as undertaken in this research.

The standard approach to develop brief message content for mHealth is to align message content with an appropriate theory, develop content based on a desired behavior, and then pretest the messages on users, often rating acceptability using a validated survey [42]. Although message pretesting is considered a priority to "understand and learn about the specific audience's preferences in technology use, language, and health needs" [40], co-designing message content with end users is critical for acceptability and workflow integration [43]. Methods to create targeted messages for mHealth engagement using a person-centered approach are limited, although there is recognition that researchers need to consider formative qualitative investigation with the target population before message development [44], as was conducted in this research.

This study extends the shortcomings of previous research by developing a 5-step process to develop co-designed message content [40]. The attributes, themes, and engagement stages identified in this study can be used to develop personalized two-way message content that can be verified by message pretesting. If a health care provider, such as a weight loss clinic, wants to focus on a user who is demonstrating behaviors associated with the disengagement stage, they can ask about interruptions in brief message content and encourage re-engagement with an existing or new device with different features to achieve health goals. Attributes that are important to goal setting may include exercise, nutrition, sleep, weight, or heart rate. The proposed process highlights the importance of co-designing messages to ensure the best response and acceptance by users.

Conclusions

On the basis of current evidence, brief messages are cost-effective and can be personalized and delivered at scale in multiple settings to improve sustained health engagement. This research demonstrates that end-user mHealth engagement is complex, nonlinear, can be social or solo, and is characterized by 6 stages. Each stage is defined by themes and attributes that signify the total user experience. Fundamentally, this research calls for the consideration of interdisciplinary frameworks that incorporate health, information systems, and business to avoid the partial view of this phenomenon and to enhance engagement over time.

The limitations of this research are the small sample size from a single first-world country. Future research could measure the existing, and explore for additional, engagement stages using a larger sample and in different cultures. This would allow for more detailed investigations between groups, such as those users at different stages in the engagement journey, those displaying different demographic characteristics, or those using different types of technologies.

Further research could also include larger quantitative investigations that could be used to overcome the more exploratory and perhaps limited representative sample in this study. Future quantitative research could investigate interventions at each engagement stage or use experimental design to test the suitability of message content across these stages. The development of a short identification survey could also be undertaken, focusing on the 6 engagement stages. This would enable quick categorization of users and the stage they are at in their engagement journey, thus limiting the reliance on more longer time-intensive interviews for organizations that could result in more easily obtained targeted communication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview and observation protocol.
[DOCX File , 16 KB - jmir_v22i9e15307_app1.docx]

Multimedia Appendix 2
Expression of each theme at each engagement stage.
[DOCX File , 21 KB - jmir_v22i9e15307_app2.docx]

Multimedia Appendix 3
Brief communication content attributes.
[DOCX File , 83 KB - jmir_v22i9e15307_app3.docx]

References

1. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. J Am Med Assoc 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
2. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
3. Von Knoop C, Lovich D, Silverstein MB, Tutty M. Vital Signs: E-Health in the United States. The Boston Consulting Group. 2003. URL: https://mkt-bcg-com-public-images.s3.amazonaws.com/public-pdfs/legacy-documents/file14063.pdf [accessed 2020-08-20]
4. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
5. Long H, Bartlett YK, Farmer AJ, French DP. Identifying brief message content for interventions delivered via mobile devices to improve medication adherence in people with type 2 diabetes mellitus: a rapid systematic review. J Med Internet Res 2019 Jan 9;21(1):e10421 [FREE Full text] [doi: 10.2196/10421] [Medline: 30626562]
6. Lund S, Rasch V, Hemed M, Boas IM, Said A, Said K, et al. Mobile phone intervention reduces perinatal mortality in zanzibar: secondary outcomes of a cluster randomized controlled trial. JMIR Mhealth Uhealth 2014 Mar 26;2(1):e15 [FREE Full text] [doi: 10.2196/mhealth.2941] [Medline: 25098184]
7. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
8. Burns K, Keating P, Free C. A systematic review of randomised control trials of sexual health interventions delivered by mobile technologies. BMC Public Health 2016 Aug 12;16(1):778 [FREE Full text] [doi: 10.1186/s12889-016-3408-z] [Medline: 27514851]
9. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
10. Ambeba EJ, Ye L, Sereika SM, Styn MA, Acharya SD, Sevick MA, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43 [FREE Full text] [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
11. Keeling DI, de Ruyter K, Mousavi S, Laing A. Technology push without a patient pull: examining digital unengagement (DU) with online health services. Eur J Mark 2019 Sep 9;53(9):1701-1732. [doi: 10.1108/ejm-10-2017-0692]
12. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
13. Alagoz F, Valdez AC, Wilkowska W, Ziefle M, Dorner S, Holzinger A. From cloud computing to mobile Internet, from user focus to culture and hedonism: the crucible of mobile health care and wellness applications. In: Proceedings of the 5th International Conference on Pervasive Computing and Applications. 2010 Presented at: ICPCA'10; Dec 1-3, 2010; Maribor, Slovenia. [doi: 10.1109/ICPCA.2010.5704072]
14. Wirtz J, den Ambtman A, Bloemer J, Horváth C, Ramaseshan B, van de Klundert J, et al. Managing brands and customer engagement in online brand communities. J Serv Manag 2013 Jun 14;24(3):223-244. [doi: 10.1108/09564231311326978]
15. Zainuddin N, Previte J, Russell-Bennett R. A social marketing approach to value creation in a well-women's health service. J Mark Manag 2011 Feb 23;27(3-4):361-385. [doi: 10.1080/0267257x.2011.547081]
16. Maklan S, Klaus P. Customer experience: are we measuring the right things? Int J Mark Res 2011 Nov;53(6):771-772. [doi: 10.2501/ijmr-53-6-771-792]
17. Verhoef PC, Lemon KN, Parasuraman A, Roggeveen A, Tsiros M, Schlesinger LA. Customer experience creation: determinants, dynamics and management strategies. J Retail 2009 Mar;85(1):31-41. [doi: 10.1016/j.jretai.2008.11.001]
18. Voss C, Roth AV, Chase RB. Experience, service operations strategy, and services as destinations: foundations and exploratory investigation. Product Oper Manag 2008 May;17(3):247-266. [doi: 10.3401/poms.1080.0030]
19. Lemke F, Clark M, Wilson H. Customer experience quality: an exploration in business and consumer contexts using repertory grid technique. J Acad Mark Sci 2011;39(6):846-869. [doi: 10.1007/s11747-010-0219-0]
20. Zeithaml VA. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J Mark 1988 Jul 1;52(3):2-22. [doi: 10.1177/002224298805200302]
21. Woodall T. Conceptualising 'value for the customer': an attributional, structural and dispositional analysis. Acad Mark Sci Rev 2003;12(12):1-42 [FREE Full text]
22. Payne A, Holt S. Diagnosing customer value: integrating the value process and relationship marketing. Br J Management 2001 Jun;12(2):159-182. [doi: 10.1111/1467-8551.00192]

23. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
24. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
25. Lohse K, Shirzad N, Verster A, Hodges N, van der Loos HF. Video games and rehabilitation: using design principles to enhance engagement in physical therapy. J Neurol Phys Ther 2013 Dec;37(4):166-175. [doi: 10.1097/NPT.0000000000000017] [Medline: 24232363]
26. Short CE, Vandelanotte C, Dixon MW, Rosenkranz R, Caperchione C, Hooker C, et al. Examining participant engagement in an information technology-based physical activity and nutrition intervention for men: the manup randomized controlled trial. JMIR Res Protoc 2014 Jan 3;3(1):e2 [FREE Full text] [doi: 10.2196/resprot.2776] [Medline: 24389361]
27. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
28. O'Brien HL, Toms EG. Examining the generalizability of the User Engagement Scale (UES) in exploratory search. Inf Process Manag 2013 Sep;49(5):1092-1107. [doi: 10.1016/j.ipm.2012.08.005]
29. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, California: Sage Publications; 2017.
30. Dahlke S, Hall W, Phinney A. Maximizing theoretical contributions of participant observation while managing challenges. Qual Health Res 2015 Aug;25(8):1117-1122. [doi: 10.1177/1049732315578636] [Medline: 25810466]
31. Boddy CR. Sample size for qualitative research. Qual Market Res Int J 2016 Sep 12;19(4):426-432. [doi: 10.1108/qmr-06-2016-0053]
32. National Statement on Ethical Conduct in Human Research 2007. The National Health and Medical Research Council. 2018. URL: https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#block-views-block-file-attachments-content-block-1 [accessed 2019-06-10]
33. Green J, Willis K, Hughes E, Small R, Welch N, Gibbs L, et al. Generating best evidence from qualitative research: the role of data analysis. Aust N Z J Public Health 2007 Dec;31(6):545-550. [doi: 10.1111/j.1753-6405.2007.00141.x] [Medline: 18081575]
34. Neuman WL, Kreuger LW. Social Work Research Methods: Qualitative and Quantitative Applications. Boston, Massachusetts: Allyn and Bacon; 2003.
35. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 1;27(2):237-246. [doi: 10.1177/1098214005283748]
36. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Indust Manag Data Syst 2015 Oct 19;115(9):1704-1723. [doi: 10.1108/imds-03-2015-0087]
37. Burns K, McBride CA, Patel B, FitzGerald G, Mathews S, Drennan J. Creating consumer-generated health data: interviews and a pilot trial exploring how and why patients engage. J Med Internet Res 2019 Jun 13;21(6):e12367 [FREE Full text] [doi: 10.2196/12367] [Medline: 31199312]
38. Mitzner TL, McBride SE, Barg-Walkow LH, Rogers WA. Self-management of wellness and illness in an aging population. Rev Hum Factors Ergon 2013 Sep 26;8(1):277-333. [doi: 10.1177/1557234x13492979]
39. Ajana B. Digital health and the biopolitics of the quantified self. Digit Health 2017;3:2055207616689509 [FREE Full text] [doi: 10.1177/2055207616689509] [Medline: 29942580]
40. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17(Suppl 1):82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
41. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
42. Diez-Canseco F, Zavala-Loayza JA, Beratarrechea A, Kanter R, Ramirez-Zea M, Rubinstein A, et al. Design and multi-country validation of text messages for an mhealth intervention for primary prevention of progression to hypertension in Latin America. JMIR Mhealth Uhealth 2015 Feb 18;3(1):e19 [FREE Full text] [doi: 10.2196/mhealth.3874] [Medline: 25693595]
43. Barnett ML, Mehrotra A, Jena AB. Adverse inpatient outcomes during the transition to a new electronic health record system: observational study. Br Med J 2016 Jul 28;354:i3835 [FREE Full text] [doi: 10.1136/bmj.i3835] [Medline: 27471242]
44. Ybarra ML, Holtrop JS, Bağci Bosi AT, Emri S. Design considerations in developing a text messaging program aimed at smoking cessation. J Med Internet Res 2012 Jul 24;14(4):e103 [FREE Full text] [doi: 10.2196/jmir.2061] [Medline: 22832182]

Abbreviations

GCC: Global Corporate Challenge
mHealth: mobile health

Edited by G Eysenbach; submitted 07.07.19; peer-reviewed by N Brew Sam, S Atanasova, T Irizarry; comments to author 03.10.19; revised version received 23.12.19; accepted 26.07.20; published 22.09.20.

Please cite as:
Burns K, Nicholas R, Beatson A, Chamorro-Koc M, Blackler A, Gottlieb U
Identifying Mobile Health Engagement Stages: Interviews and Observations for Developing Brief Message Content
J Med Internet Res 2020;22(9):e15307
URL: http://www.jmir.org/2020/9/e15307/
doi:10.2196/15307
PMID:32960181

Original Paper

The Influence of Three Modes of Human Support on Attrition and Adherence to a Web- and Mobile App–Based Mental Health Promotion Intervention in a Nonclinical Cohort: Randomized Comparative Study

Melanie Elise Renfrew[1], BEd, Grad Dip; Darren Peter Morton[1], MAppSci, PhD; Jason Kyle Morton[1], BEd, BSci, PhD; Jason Scott Hinze[1], BEd, MLMEd, PhD; Geraldine Przybylko[1], BCom, MBA, MPH; Bevan Adrian Craig[1], BN, Grad Dip, PhD

Lifestyle and Health Research Centre, Avondale University College, Cooranbong, Australia

Corresponding Author:
Melanie Elise Renfrew, BEd, Grad Dip
Lifestyle and Health Research Centre
Avondale University College
582 Freemans Drive
Cooranbong, NSW
Australia
Phone: 61 405445151
Email: melanie.renfrew@avondale.edu.au

Abstract

Background: The escalating prevalence of mental health disorders necessitates a greater focus on web- and mobile app–based mental health promotion initiatives for nonclinical groups. However, knowledge is scant regarding the influence of human support on attrition and adherence and participant preferences for support in nonclinical settings.

Objective: This study aimed to compare the influence of 3 modes of human support on attrition and adherence to a digital mental health intervention for a nonclinical cohort. It evaluated user preferences for support and assessed whether adherence and outcomes were enhanced when participants received their preferred support mode.

Methods: Subjects participated in a 10-week digital mental health promotion intervention and were randomized into 3 comparative groups: standard group with automated emails (S), standard plus personalized SMS (S+pSMS), and standard plus weekly videoconferencing support (S+VCS). Adherence was measured by the number of video lessons viewed, points achieved for weekly experiential challenge activities, and the total number of weeks that participants recorded a score for challenges. In the postquestionnaire, participants ranked their preferred human support mode from 1 to 4 (S, S+pSMS, S+VCS, S+pSMS & VCS combined). Stratified analysis was conducted for those who received their first preference. Preintervention and postintervention questionnaires assessed well-being measures (ie, mental health, vitality, depression, anxiety, stress, life satisfaction, and flourishing).

Results: Interested individuals (N=605) enrolled on a website and were randomized into 3 groups (S, n=201; S+pSMS, n=202; S+VCS, n=201). Prior to completing the prequestionnaire, a total of 24.3% (147/605) dropped out. Dropout attrition between groups was significantly different (*P*=.009): 21.9% (44/201) withdrew from the S group, 19.3% (39/202) from the S+pSMS group, and 31.6% (64/202) from the S+VCS group. The remaining 75.7% (458/605) registered and completed the prequestionnaire (S, n=157; S+pSMS, n=163; S+VCS, n=138). Of the registered participants, 30.1% (138/458) failed to complete the postquestionnaire (S, n=54; S+pSMS, n=49; S+VCS, n=35), but there were no between-group differences (*P*=.24). For the 69.9% (320/458; S, n=103; S+pSMS, n=114; S+VCS, n=103) who completed the postquestionnaire, no between-group differences in adherence were observed for mean number of videos watched (*P*=.42); mean challenge scores recorded (*P*=.71); or the number of weeks that challenge scores were logged (*P*=.66). A total of 56 participants (17.5%, 56/320) received their first preference in human support (S, n=22; S+pSMS, n=26; S+VCS, n=8). No differences were observed between those who received their first preference and those who did not with regard to video adherence (*P*=.91); challenge score adherence (*P*=.27); or any of the well-being measures including, mental health (*P*=.86), vitality (*P*=.98), depression (*P*=.09), anxiety (*P*=.64), stress (*P*=.55), life satisfaction (*P*=.50), and flourishing (*P*=.47).

Conclusions: Early dropout attrition may have been influenced by dissatisfaction with the allocated support mode. Human support mode did not impact adherence to the intervention, and receiving the preferred support style did not result in greater adherence or better outcomes.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR): 12619001009101; http://www.anzctr.org.au/ACTRN12619001009101.aspx

(J Med Internet Res 2020;22(9):e19945) doi:10.2196/19945

KEYWORDS

human support; adherence; attrition; engagement; web-based mental health; health promotion; eHealth; SMS; videoconferencing

Introduction

The burden of mental distress is pervasive globally and includes common mental health disorders such as depression and anxiety. Approximately 300 million people worldwide are affected by depression—the principal cause of global disability [1]. Depression is frequently comorbid with other diseases and severely compromises effective functioning for individuals, negatively impacting family and work environments [1]. Furthermore, indicators suggest that even the general population is increasingly experiencing mental distress, including severe stress, anxiety, depressive symptoms, a sense of isolation, and feeling overwhelmed [2,3].

While a growing repertoire of digital interventions is improving accessibility to treatment options for people with common mental health disorders, there is also an urgent need for easily accessible mental health promotion interventions (MHPIs) to improve the mental well-being of nonclinical population groups. MHPIs that focus on enhancing psychological well-being may provide an important buffer against mental distress, potentially attenuating the mental health burden. Furthermore, lifestyle-focused MHPIs might also ameliorate symptoms for those who have already been diagnosed with a common mental health disorder [4].

Innovative web- and mobile app–based technologies allow MHPIs to be disseminated widely and cost effectively to maximize accessibility. However, despite the many advantages of digital interventions, both high dropout attrition (ie, participants who drop out early or who are lost to follow-up) and nonusage attrition (ie, nonadherence) are persistent problems [5-7].

In his formative publication titled “The Law of Attrition,” Eysenbach [8] called for the methodical study of attrition in eHealth interventions because, unlike drug trials, it is usually easy for participants to both join and withdraw from a digital intervention, especially when they are not critical to life—sometimes described as “easy-in” and “easy-out” [9]. Maximizing adherence, defined as “the degree to which the user followed the program as it was designed” [10], is a complex challenge for researchers and health care providers. Additionally, definition differences and measurement heterogeneity between studies are problematic [11-13]. Nevertheless, many factors affecting adherence have been identified.

Influences on adherence are multifactorial, and consistency in adherence patterns have been elusive [12,14]. A 2018 review of theoretical perspectives on adherence suggested the need for interdisciplinary collaboration to better understand patterns of adherence due to a diverse range of technological, environmental, and individual influences [12]. For instance, technological factors may include website design, persuasive systems design [15,16], behavior change techniques [17], human support factors [6,8,13,18,19], personalized content (ie, tailoring) [13], frequent updates and dialogue support (ie, praise, rewards, and reminders) [18], and gamification techniques [20]. Environmental influences consist of factors such as socioeconomic status, employment status, education level [21], internet or computer accessibility [22], literacy [21,23], culture, the health care system, family and community support [24], and time availability [13,22]. Examples of individual factors include whether a person self-selects into a study and invests effort [25], planning and self-efficacy [21,26], compatibility with personal values [8], motivation factors [6,22], focus on immediate benefit rather than long-term goals, perceived treatment credibility [13,27], receiving preferences [28], health status, psychological vulnerability [21-23], user expectations [6,13,23], gender [13,21], and age [17,21].

Human Support and Adherence

Clinical Settings

Despite the broad range of factors listed, adherence has frequently been positively associated with human support (ie, guidance) in clinical settings. A qualitative systematic review of 64 studies reported that adherence was improved by support of counselors, peers, and phone and email contact [24]. A 2012 systematic review and meta-analysis concluded that supported interventions yielded better retention and outcomes [29]. Other randomized controlled trials (RCTs) have also reported similar links between human support and adherence [14,30], though greater adherence does not always translate to better outcomes; and adherence may be problematic for individuals with depression, irrespective of the support received [30].

A 2017 scoping review [31] analyzed 19 RCTs from 2000 to 2016 that considered human support factors in internet-based interventions for depression and anxiety. The review identified 7 different human support factors (guided vs unguided, therapist expertise, human vs automated, scheduled vs unscheduled, support mode, synchronicity, and support intensity) and analyzed them for improvement in clinical outcomes and adherence. While just one human support factor (scheduled support) was associated with significantly improved outcomes, results were mixed in relation to adherence, with human support improving adherence in only 4 out of 9 studies [31].

Recent web-based interventions for common mental health disorders, comparing supported and unsupported arms, have found that well-designed self-guided interventions achieve significant improvements in outcomes and maintain high adherence rates irrespective of support provided [32-34]. Notably, in some cases, treatment satisfaction was higher in supported arms [34,35] and some participants perceived support as necessary to success [36].

Human support requirements to encourage adherence may be vastly different in nonclinical populations compared to clinical cohorts who experience symptoms that may preclude them from engaging with an intervention. Therefore, it is vital to determine if human support adds value to an intervention for a nonclinical cohort because unsupported interventions can be administered at a lower cost and be more easily distributed in a scalable manner.

Nonclinical Settings

Studies evaluating the influence of human support on adherence to MHPIs among nonclinical populations are scant in comparison to clinical cohorts. A study involving a mindfulness intervention, targeting college students and young working adults, found that despite telephone or email support, adherence was poor and nonadherers had poorer mental well-being and lower energy and treatment expectancy [37]. The researchers suggested a greater need for collaboration between health professionals and information technology experts to improve the "personalization" of digital interventions to enhance adherence [37]. A Swedish study that implemented an internet-based relaxation program found that human support did not affect treatment outcomes or adherence [27,35]. Outcomes were positively associated with completing the homework (ie, behavioral tasks) but not engagement with the online aspect of the program. Early attrition was predicted by low belief in the treatment; and nonadherence was associated with increased stress symptoms, lower levels of intrinsic motivation, and a greater focus on immediate consequences of behavior as opposed to long-term gains. Conversely, adherence was predicted (positively) by education level and treatment credibility.

In a pooled analysis of 3 web-based studies [14], researchers investigated the influence of 3 types of support: content-focused (ie, personalized email feedback), adherence-focused (ie, monitoring adherence and sending reminders), and administrative support (ie, access to contact details to ask for technical assistance). Those who received content- and adherence-focused support completed more modules than those who received only administrative support. However, the researchers concluded that even after taking human support and other demographic variables into consideration, most interindividual variations in nonadherence remained largely unsolved [14]. A web-based mindfulness and stress management RCT [38] compared the effects of no support, group support only, and group support with added clinician support on engagement and outcomes. Group support improved outcomes and adherence, but extra clinician support added no benefit. Notably, although the program was web-based, support was provided face-to-face in the workplace.

We have previously reported the mental health outcomes of this study and found that no difference was observed between groups; improvements in outcomes were obtained irrespective of the differing modes of human support offered [39]. This study focused on examining attrition and adherence patterns between the groups. We also evaluated user preferences for human support and assessed whether adherence and mental health outcomes were enhanced when participants received their preferred mode of human support. The outcomes of the study contribute toward understanding the value of human support on adherence to a web- and mobile app–based MHPI targeting a nonclinical cohort. This is vital information for researchers and clinicians in order to inform the optimal delivery of digital MHPIs in a cost-effective, accessible, and scalable manner.

Methods

This section provides a brief summary of methods used. For a detailed explanation of the methods, refer to our previous article that reported the influence of human support mode on mental well-being outcomes [39].

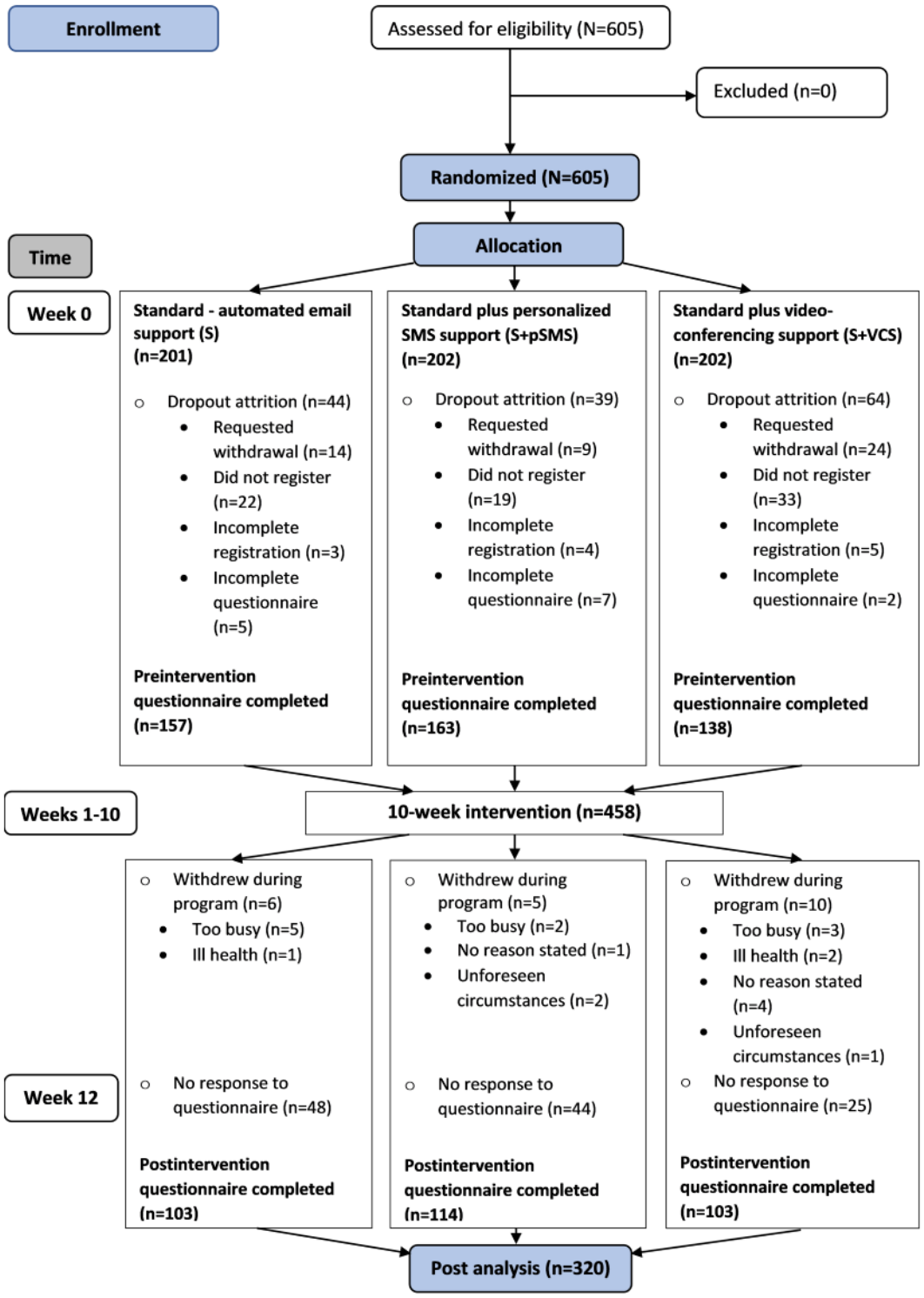
Study Design

The multiarm randomized comparative study design included 3 intervention groups that varied by human support mode: standard (S), comprising fully automated emails only; standard plus personalized SMS text messaging support (S+pSMS); standard plus videoconferencing support (S+VCS).

Recruitment and Randomization

The study was advertised to members of a faith-based organization using a combination of methods (eg, email, social media, bulletins, and magazines), and interested individuals were directed to a website to apply. After randomization, eligible participants were emailed their group allocation, a participant information document, and instructions on how to complete registration onto the electronic learning management system (eLMS), signaling consent. Figure 1 demonstrates the flow of participants through the study.

Figure 1. Flow of participants.



Intervention

The 10-week interdisciplinary intervention introduced participants to a range of evidence-based strategies for enhancing mental well-being, from the disciplines of lifestyle medicine and positive psychology. Participants logged onto the eLMS or mobile app to access the intervention, which included video content and a place to log daily and weekly experiential challenges (see screenshots in Multimedia Appendix 1). The website and mobile app were designed using a range of persuasive system design principles to increase engagement [15]. Each week, participants were encouraged to view one lesson, perform small daily challenges, and complete 1 larger weekly challenge. A detailed week-by-week summary of the content and experiential challenge activities can be found in Multimedia Appendix 2.

Human Support

The study was structured to compare the value of increasingly greater levels of human support on adherence. Principles of the adherence-focused supportive accountability model [6] underpinned interactions between the support coach and participants. As the lowest form of support, one automated email was sent to the participants each week, consisting of the following: their first name as a salutation, a couple of sentences to encourage them to view the content, and a link to an introductory video by the presenter. A reminder email was sent if a participant had not logged experiential activity for 3 days, or viewed video content for 8 days. The second level of support involved SMS messages that were written and sent from a support coach who focused on ensuring content was process-focused; messages were sent regularly, and tone exuded positivity. Messages addressed the recipient by first name and were signed by the support coach. They were sent 3 times weekly for the first 3 weeks, and twice weekly for the remaining 7 weeks. In the highest level of support, the opportunity to develop key tenets of the supportive accountability model (eg, bond and legitimacy) was provided through weekly videoconferencing sessions (9 weekly time slots). Focus was given to facilitating a supportive, respectful environment where participants could share experiences safely. The facilitator sought to build trust through benevolent engagement with participants.

Measurements

Dropout Attrition

Dropout attrition was measured as the total number of randomized participants who did not complete either the preintervention (ie, early dropout) or the postintervention (ie, lost to follow-up) questionnaire.

Adherence Measures

Primary Adherence—Videos Viewed

Primary adherence was measured as the total number of weekly videos viewed, out of a total of 10. A video was considered "viewed" when at least 80% of the presentation had been played.

Secondary Adherence—Experiential Challenge Activities

Participants were encouraged to accumulate points through daily and weekly challenge activities that involved putting the new learning into practice. Engagement with challenges was measured in 2 ways—the total weekly challenge score and the total number of weeks out of 10 in which an engagement with a challenge was recorded. Each daily challenge was worth 10 points (ie, a total of 70 points weekly), and weekly challenges were worth 30 points. Therefore, participants could score a maximum of 100 points each week throughout the 10-week intervention, thereby accumulating a total of 1000 points to be considered fully adherent for the experiential challenge activities.

Videoconference Adherence

Attendance records were kept for the videoconference support sessions for those in the S+VCS group. Each participant was invited to attend 1 session each week, out of a possible 9 available time slots. Participants were able to receive a maximum adherence score of 10 over the 10-week intervention.

Well-Being Measures

A self-report questionnaire termed the "7 Dimensions of Wellness Index" was completed by all participants on the web-based eLMS or on the mobile app, at preintervention (Week 0) and postintervention (Week 12) (Figure 1). Demographic and lifestyle-associated questions were combined with freely accessible validated instruments to measure the participant's well-being in several domains: physical, emotional, social, vocational, intellectual, spiritual, and environmental. Validated instruments utilized for this study on mental well-being included 2 subdomains from the 36-item Short Form Health Survey (SF-36) (ie, mental health and vitality) [40]; the 21-question Depression Anxiety and Stress Scales (DASS-21) to measure depression, anxiety, and stress [41]; the Diener flourishing scale [42]; and the Diener Satisfaction With Life Scale (SWLS) [43,44]. Well-being measures are reported on in detail in the aforementioned article [39].

Preferences

Participants rated 4 different human support options from most preferred to least preferred (1=most preferred, 4=least preferred) as part of the postintervention questionnaire. Support options were as follows: standard (automated email only), standard plus SMS only, standard plus VCS only, and standard plus SMS and VCS combined (not offered in the study). Stratified analyses were conducted to measure adherence and outcomes for participants who received their desired human support preference.

Statistical Analysis

Analyses were conducted using SPSS, version 25 (IBM Corp). Preintervention-to-postintervention changes were calculated using paired *t* tests. Descriptive statistics, involving means and standard deviations, measured patterns of adherence. Analysis of variance (ANOVA) was used to compare adherence between groups. Fisher exact test was utilized to analyze relationships between categorical variables, and Cohen *d* measured effect size.

Ethics and Informed Consent

Avondale Human Research Ethics Committee granted ethics approval for the study (Approval No. 2018.09), and it was registered with the Australian and New Zealand Clinical Trial Registry (ANZCTR12619001009101). An email containing an "Information Statement" and a "Participant Consent Form" was sent to all prospective participants. The email explained that choosing to register on the eLMS would signify informed consent.

Results

Participants

Potential participants (N=605) enrolled on the information website and agreed to participate in the study irrespective of randomization allocation. Subjects were randomized into 3 arms (S, n=201; S+pSMS, n=202; and S+VCS, n=202) (Figure 1).

During a 1-week period, after being notified of group allocation, and before the intervention commenced, 24.3% (147/605) of participants dropped out. Early dropout attrition between groups was significantly different (*P*=.009). A total of 21.9% (44/201) withdrew from the S group, 19.3% (39/202) from the S+pSMS group, and 31.6% (64/202) from the S+VCS group. Table 1 shows that dropout was significantly different between groups when the S+VCS group was compared; differences were not significant between the S and S+pSMS groups.

A total of 75.7% (458/605) of randomized participants registered on the eLMS and completed the initial questionnaire (S, n=157; S+pSMS, n=163; S+VCS, n=138). Notably, there were no significant differences between the groups in age (*P*=.19), gender (*P*=.82), education (*P*=.16), or ethnicity (*P*=.34). Of the registered participants, 69.9% (320/458) completed the postintervention questionnaire (S, n=103; S+pSMS, n=114; S+VCS, n=103), resulting in 30.1% (138/458) being lost to follow-up; there was no difference between groups (*P*=.24).

Table 1. Dropout attrition between groups.

Group comparison	*P* value
All groups	.009
S[a] and S+VCS[b]	.03
S+pSMS[c] and S+VCS	.004
S and S+pSMS	.52

[a]S: standard (automated emails only).

[b]S+VCS: standard plus videoconferencing support.

[c]S+pSMS: standard plus personalized SMS.

Adherence

Primary Adherence—Videos Viewed

As shown in Table 2, the number of videos viewed was not significantly different between the groups (*P*=.42). Almost half of all participants in each group were fully adherent (ie, watched all 10 video sessions), with less than 10% of individuals in each group viewing no videos. Furthermore, the percentage of participants at any level of adherence did not differ significantly between groups.

Table 2. Primary adherence percentages and between-group comparisons.

Videos viewed	S[a] (n=103)	S+pSMS[b] (n=114)	S+VCS[c] (n=103)	Between-group difference *P* value
Number of videos viewed, n (%)				
10	47 (44.6)	54 (47.4)	50 (48.5)	.58
8-9	4 (3.8)	6 (5.3)	5 (4.9)	.38
5-7	12 (11.6)	15 (13.2)	18 (17.5)	.54
1-4	33 (32.0)	31 (27.2)	21 (20.4)	.55
0	7 (6.8)	8 (7.0)	9 (8.7)	.95
Videos viewed, mean (SD)	6.0 (4.04)	6.5 (3.87)	6.8 (3.75)	.42

[a]S: standard (automated emails only).

[b]S+pSMS: standard plus personalized SMS.

[c]S+VCS: standard plus videoconferencing support.

Secondary Adherence—Challenges

No significant differences were recorded between the groups in the mean challenge points scored (*P*=.71*)* or in the mean number of weeks in which challenge scores were recorded *(P*=.66) (Table 3). There was an overall lack of adherence to experiential challenges, as indicated by mean challenge scores less than 400 out of a possible 1000 points. Additionally, challenge scores were logged in fewer than half of the weeks for the 10-week intervention.

Figure 2 shows weekly mean challenge scores by group, and Figure 3 portrays what percentage of participants were logging challenges each week over the 10-week intervention. The mean number of challenge points scored was greatest in the first 3 weeks for the 3 groups. Additionally, scores declined in a graded manner from weeks 3 to 10 in each group, as illustrated in Figure 2. During the first 3 weeks of the intervention, between 59%-72% (67 to 74) participants in each group were logging challenges. The number of participants logging challenges was greatest in week 1 for the S group, highest in week 2 for the S+VCS group, and highest in week 3 for the S+pSMS group, although between-group differences were not significant. Figure 3 illustrates the steady decline in participants logging challenges from lessons 3 to 10 for all groups.

Table 3. Secondary adherence scores for experiential challenge activities.

Challenge activities	S[a] (n=103)	S+pSMS[b] (n=114)	S+VCS[c] (n=103)	*P* value
Challenge points (out of 1000), mean (SD)	368.7 (361.6)	340.2 (339.0)	377.5 (354.0)	.71
Number of weeks in which challenge scores were logged (out of 10), mean (SD)	4.5 (3.7)	4.4 (3.4)	4.8 (3.6)	.66

[a]S: standard (automated emails only).

[b]S+pSMS: standard plus personalized SMS.

[c]S+VCS: standard plus videoconferencing support.

Figure 2. Group-based mean challenge scores over 10 weeks. S: standard—automated emails; S+pSMS: standard plus personalized SMS; S+VCS: standard plus videoconferencing support.

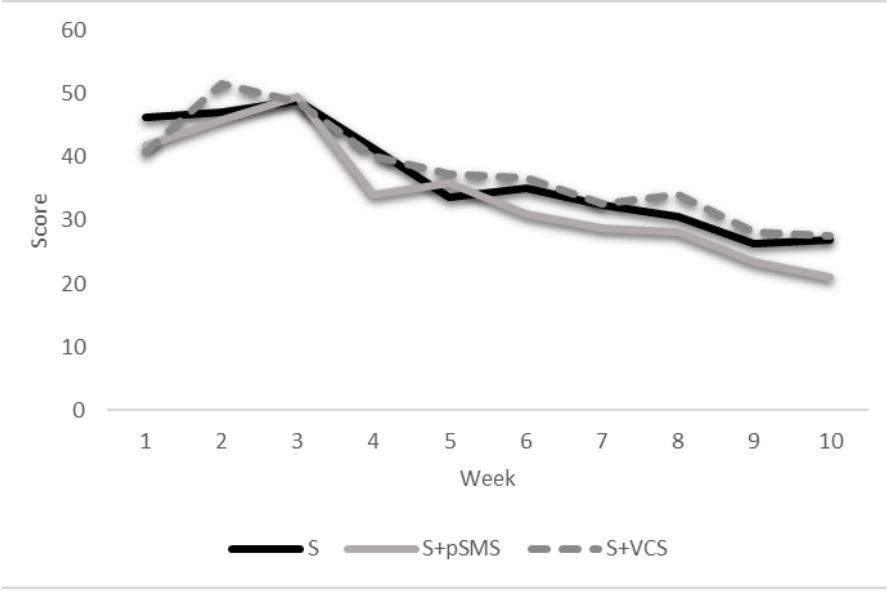
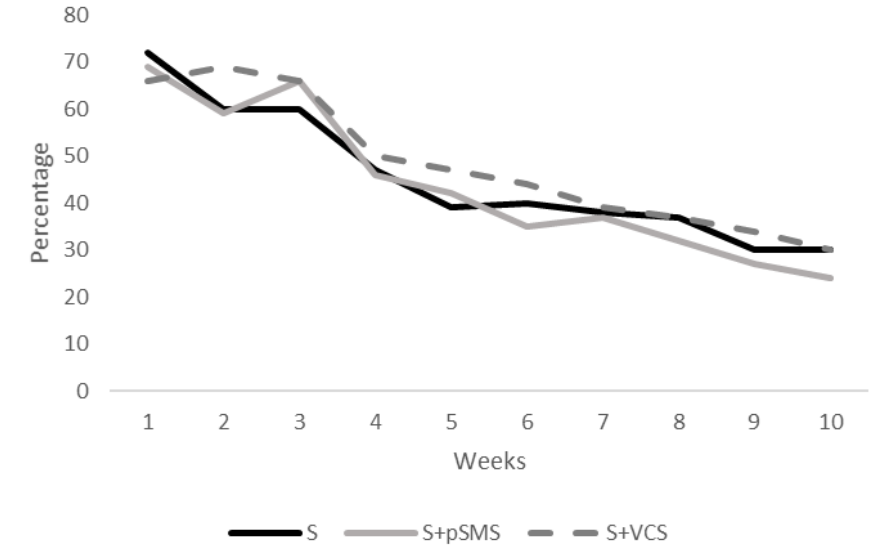


Figure 3. Percentage of participants who logged challenges over 10 weeks. S: standard—automated emails; S+pSMS: standard plus personalized SMS; S+VCS: standard plus videoconferencing support.



Videoconference Adherence

Participation in videoconferencing support was low, with 35.9% (37/103) of participants (in the S+VCS group) not attending any support sessions and only 18.4% (19/103) attending 7 or more out of 10 sessions.

Preferences

Table 4 specifies the preferred mode of human support by group. A total of 88.7% (284/320) of participants indicated a valid human support preference. Almost half of participants within each group indicated a preference for combined support (ie, automated emails, personalized SMS, and VCS); however, the combined support alternative was not offered as part of the study. In comparison, the preference for solely S+VCS support was low across all the groups, with only between 7% to 8% choosing it from each group. Table 4 also indicates that 24% (22/91) in the S group, 27% (20/96) in the S+pSMS group, and 8% (8/97) in the S+VCS group received their preferred mode of support.

Secondary analysis was conducted to determine whether those who received their first preference for support mode scored better adherence or outcomes than those who did not receive their first choice. Participants (n=128) who chose the combined support as their first preference (an option not included in the comparative study) were excluded from the analysis, and the remaining participant data (n=156) were analyzed. Table 5 outlines the results between those who did not receive their first preference in human support mode and those who did. There was no significant difference between the 2 groups in either adherence or well-being measures.

Table 4. Preferences for modes of human support.

Support mode preference	S[a] (n=91)	S+pSMS[b] (n=96)	S+VCS[c] (n=97)
Standard, n (%)	22 (24)	21 (22)	18 (19)
Standard plus SMS support, n (%)	25 (27)	26 (27)	23 (24)
Standard plus VCS[d] support, n (%)	6 (6)	7 (7)	8 (8)
Standard plus SMS & VCS support, n (%)	38 (42)	42 (44)	48 (49)

[a]S: standard (automated emails only).

[b]S+pSMS: standard plus personalized SMS.

[c]S+VCS: standard plus videoconferencing support.

[d]VCS: videoconferencing support.

Table 5. The preferred human support mode's effect on adherence and outcomes.

Variables	Preference		Between-group difference	
	No (n=1000), mean (SD)	Yes (n=56), mean (SD)	*P* value	Cohen *d*
Adherence				
Videos watched (out of 10)	6.1 (3.9)	6.5 (4.0)	.91	0.09
Challenge points (out of 100)	339.9 (352.2)	388.6 (385.9)	.27	0.13
Outcomes				
Mental health	8.4 (14.4)	9.8 (13.3)	.86	0.10
Vitality	9.5 (15.2)	10.6 (13.8)	.98	0.08
Depression	–1.3 (3.4)	–1.6 (2.5)	.09	0.10
Anxiety	–0.6 (1.9)	–1.2 (2.6)	.64	0.26
Stress	–1.0 (3.5)	–1.6 (2.8)	.55	0.19
Life satisfaction	2.2 (4.6)	1.9 (4.8)	.50	0.06
Flourishing	0.14 (.53)	0.21 (.53)	.47	0.13

Discussion

Principal Results

Higher early dropout attrition occurred in the group which was allocated videoconferencing as a mode of human support. However, the mode of human support made no impact on attrition or adherence after the commencement of the intervention. Moreover, for participants who received their first preference in human support, compared to those who did not, no differences were observed in adherence. Preference for VCS support was low, yet almost half of the participants indicated they would prefer all forms of human support, though this was not an option in this study.

In this study, after the initial email notified participants of their group allocation, a disproportionate number of participants allocated to the S+VCS group withdrew from the study, with a significant between-group difference (*P*=.009). The early dropout attrition may have been influenced by dissatisfaction

with the assigned support mode. Figure 1 shows that, in the S+VCS group, 57 participants either requested withdrawal or failed to register compared to 36 in the email group (S), and 28 in the S+pSMS group. Of the 24 participants who requested withdrawal from the S+VCS group, 10 stated that time was a factor, 8 provided no specific reason, 4 stated personal illness, 1 mentioned work commitments, and 1 participant stated they did not want to be involved in videoconferencing.

No differences in adherence between the groups were detected; and as documented in a previous report [39], we observed no differences between the groups in the well-being measures either. However, in observing that almost half (150/320, 46.9%) of the entire cohort were 100% adherent (ie, viewed all 10 videos), we conducted stratified analysis to compare well-being measures of those who were fully adherent (n=150) with those who were not (n=170). Significantly greater improvements were observed in the fully adherent group for life satisfaction (P=.011; d=0.15) and flourishing scores (P=.012; d=0.15), yet effect sizes were small.

While the primary aim of human support is usually to foster greater adherence [14,45,46], support did not influence adherence behavior for the participants in this study. While earlier research among clinical cohorts have indicated that supported interventions yielded better adherence and outcomes [29,47,48], numerous studies have also shown that human support made no difference [31,32,34,49-53]. Considering the extensive repertoire of previously identified factors affecting adherence, it is plausible that, in this nonclinical group, a wide range of variants may have influenced adherence at the individual level [14], reducing the potential impact of human support as a single factor. Therefore, exploring participants' perceptions regarding influences on adherence is an important topic for further investigation.

The S+VCS was the least preferred support mode (7%), and low VCS attendance reflected the low preference for this type of support. Exploring reasons for the disinterest in videoconferencing as a mode of support may be a topic for further research, as videoconferencing has been used successfully in other group contexts [54-58] and has proven to be useful so long as technical support was provided [54]. Albeit, in a recent German study, 64% of patients were resistant to videoconferencing as a method of communication with health professionals. Notably, less than 1% reported previous experience with its use [59].

The S+VCS group allocation required an extra time commitment on behalf of participants compared to the other groups. Both email and SMS support required no effort by the participant—support was "pushed" to devices [7]. Conversely, videoconference support attempted to "pull" participants to an extra event, requiring effort and time to gain benefit from the support offered. Conceivably, the time and energy required to engage with that mode of support, loss of anonymity, technological barriers, unfamiliarity with videoconferencing software, or concerns about the group interaction may have been a barrier that facilitated significantly greater dropout. Additionally, further research should investigate if the preference for S+VCS would have been higher if support had been provided on an individual rather than a group basis.

For participants remaining in the S+VCS group, many demonstrated low engagement with the videoconferencing support. Consequently, by not engaging in the VCS support, they experienced a similar level of support to those in the S group (ie, automated emails only). This hindered the ability to draw meaningful between-group comparisons regarding the influence of human support offered through video conferencing on adherence.

Receiving their first preference in human support mode did not translate to participants reporting better adherence or outcomes in this study. We were unable to locate comparable studies for nonclinical groups. However, previous research among clinical cohorts has demonstrated that receiving support preferences may impact patient perceptions about the usefulness of an intervention [60] and improve adherence [28,61]. While a meta-analysis revealed that patients receiving their preferences demonstrated improved treatment satisfaction, adherence, and outcomes with moderate effect sizes [28], other research found that receiving the preferred option does not always impact outcomes [61,62]. A 2019 study of patients with anxiety and depression compared adherence and outcomes when patients chose their preferred support [63]. Interestingly, 78% chose the maximal support option, regular weekly support, and just 22% chose optional support (ie, support by request only). Yet, both groups achieved similar improvements in anxiety and depression scores and there were no differences in adherence. Contact between participants and therapist was much less for those who chose optional support, suggesting that similar results can be achieved with less time and cost investment.

Although combined human support (ie, access to automated emails, SMS, and videoconferencing) was not offered in the study, it is interesting that almost half of the participants in every group chose this as their preferred support option. While the reasons for this are unclear, it is hypothesized that it is a common trait of human nature to want access to everything possible, even though the available options are not necessarily utilized or needed. Further research exploring participant perceptions regarding human support modes and preferences is warranted.

Strengths and Limitations

The intervention and its implementation were supported by established theory. While the intervention itself was underpinned by the theory of planned behavior [64], the design components of the intervention for the web- and mobile app–based platforms were informed by the persuasive design model [15] and human support elements reflected principles of the supportive accountability model [6]. The use of a sole facilitator as a support person provided consistency in the participant's experience with regards to technical assistance, messaging, and videoconference facilitation. The study attracted a large cohort with a broad range of ages (18-81 years); and adherence data was easily collected through the eLMS and mobile app, avoiding the possibility of human error. Despite the addition of human support elements, the administration of the intervention using the eLMS and the mobile app provided acceptable scalability,

portability, and accessibility, demonstrating the potential for broad-based mental health promotion.

Several limitations should be noted. The study attracted mainly White, well-educated women. While this is often seen in digital health interventions [65,66], it limits the ability to generalize more widely. Data gathering relied on self-reporting, which may be subject to bias, and self-selection into the study may have resulted in a cohort skewed by factors such as technological ability and motivation to achieve better mental health. Bias may have been introduced by the disproportionate number of participants who withdrew from the S+VCS group after being notified of their group allocation, reducing the validity of the randomization process. In asking participants to rank human support preferences, analyses were limited by including a preference option in the questionnaire that was not included in the study (ie, combined support involving email, SMS, and VCS). Almost half of the participants ranked combined support as their first preference and were consequently excluded from further investigation, reducing the power of the analyses. Additionally, some factors that would have been useful for group comparison were not measured. For example, asking participants to indicate their level of engagement with SMS and emails, and collecting data regarding preference for use of the mobile app compared to the web-based experience, would have provided an indication of engagement with the human support offered.

Conclusions

The findings of this study indicate that a web- and mobile app–based MHPI for a nonclinical cohort can be designed and implemented to maximize accessibility, scalability, and adherence without the additional cost of human support. While early dropout attrition may have been influenced by displeasure with allocated support, adherence to a 10-week MHPI for a healthy cohort was not impacted by differing modes of human support. Engagement with videoconference support was suboptimal, hindering the ability to draw meaningful between-group comparisons. However, SMS support demonstrated no added value compared to automated email support. Adherence was not impacted by participants receiving their first preference for support. Future research should explore participant perspectives on adherence behaviors.

Acknowledgments

The Seventh-day Adventist Church of the South Pacific Division funded the development of the eLMS and mobile app. They provided the intervention cost-free to participants and offered technical support throughout the duration of the study.

Conflicts of Interest

DM operates a “profit-for-purpose” trust that administers the delivery of a version of the intervention; no personal remuneration is received. GP is employed by the South Pacific Division of the Seventh-day Adventist Church, which administers the intervention among members of the organization, including the participants of this study. No authors have a financial interest in the initiative, and there are no other conflicts of interest to declare.

Multimedia Appendix 1
Website and app screenshots.
[PDF File (Adobe PDF File), 1454 KB - jmir_v22i9e19945_app1.pdf]

Multimedia Appendix 2
Detailed intervention overview.
[PDF File (Adobe PDF File), 208 KB - jmir_v22i9e19945_app2.pdf]

References

1. Depression. World Health Organization. URL: http://www.who.int/mediacentre/factsheets/fs369/en/ [accessed 2020-05-07]
2. Australian Psychological Society. Stress & Wellbeing: How Australians are coping with life. Heads Up. 2015. URL: https://www.headsup.org.au/docs/default-source/default-document-library/stress-and-wellbeing-in-australia-report.pdf [accessed 2020-05-05]
3. Health Promotion Agency. Wellbeing and mental distress in Aotearoa New Zealand: Snapshot 2016. 2018. URL: https://www.hpa.org.nz/sites/default/files/Wellbeing-And-Mental-Distress-Snapshot-2016-Final-FEB2018.PDF [accessed 2020-05-05]
4. Walsh R. Lifestyle and mental health. Am Psychol 2011 Oct;66(7):579-592. [doi: 10.1037/a0021769] [Medline: 21244124]
5. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a Mobile Phone App to Support Self-Monitoring of Emotional Well-Being: A Mental Health Digital Innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
6. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

7. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
8. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
9. Lehr D, Geraedts A, Asplund R, Khadjesari Z, Heber E, de Bloom J. Occupational e-mental health: Current approaches and promising perspectives for promoting mental health in workers. In: Healthy at work. New York: Springer; 2016:257-281.
10. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
11. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: Systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
12. Ryan C, Bergin M, Wells JS. Theoretical Perspectives of Adherence to Web-Based Interventions: a Scoping Review. Int J Behav Med 2018 Feb;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
13. Beatty L, Binnion C. A Systematic Review of Predictors of, and Reasons for, Adherence to Online Psychological Interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
14. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to Internet-Based Mobile-Supported Stress Management: A Pooled Analysis of Individual Participant Data From Three Randomized Controlled Trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
15. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. CAIS 2009;24:485-500. [doi: 10.17705/1CAIS.02428]
16. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive System Design Principles and Behavior Change Techniques to Stimulate Motivation and Adherence in Electronic Health Interventions to Support Weight Loss Maintenance: Scoping Review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
17. Short CE, Vandelanotte C, Dixon MW, Rosenkranz R, Caperchione C, Hooker C, et al. Examining participant engagement in an information technology-based physical activity and nutrition intervention for men: the manup randomized controlled trial. JMIR Res Protoc 2014 Jan 03;3(1):e2 [FREE Full text] [doi: 10.2196/resprot.2776] [Medline: 24389361]
18. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
19. Akhu-Zaheya LM, Shiyab WY. The effect of short message system (SMS) reminder on adherence to a healthy diet, medication, and cessation of smoking among adult patients with cardiovascular diseases. Int J Med Inform 2017 Feb;98:65-75. [doi: 10.1016/j.ijmedinf.2016.12.003] [Medline: 28034414]
20. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and Adherence to Web-Based Mental Health Interventions: A Systematic Review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
21. Leung AWY, Chan RSM, Sea MMM, Woo J. An Overview of Factors Associated with Adherence to Lifestyle Modification Programs for Weight Management in Adults. Int J Environ Res Public Health 2017 Aug 16;14(8):922 [FREE Full text] [doi: 10.3390/ijerph14080922] [Medline: 28813030]
22. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
23. Johansson O, Michel T, Andersson G, Paxling B. Experiences of non-adherence to Internet-delivered cognitive behavior therapy: A qualitative study. Internet Interventions 2015 May;2(2):137-142 [FREE Full text] [doi: 10.1016/j.invent.2015.02.006]
24. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
25. Lyubomirsky S, Dickerhoof R, Boehm JK, Sheldon KM. Becoming happier takes both a will and a proper way: an experimental longitudinal intervention to boost well-being. Emotion 2011 Apr;11(2):391-402 [FREE Full text] [doi: 10.1037/a0022575] [Medline: 21500907]
26. Zarski A, Berking M, Reis D, Lehr D, Buntrock C, Schwarzer R, et al. Turning Good Intentions Into Actions by Using the Health Action Process Approach to Predict Adherence to Internet-Based Depression Prevention: Secondary Analysis of a Randomized Controlled Trial. J Med Internet Res 2018 Jan 11;20(1):e9 [FREE Full text] [doi: 10.2196/jmir.8814] [Medline: 29326097]
27. Alfonsson S, Olsson E, Hursti T. Motivation and Treatment Credibility Predicts Dropout, Treatment Adherence, and Clinical Outcomes in an Internet-Based Cognitive Behavioral Relaxation Program: A Randomized Controlled Trial. J Med Internet Res 2016 Mar 08;18(3):e52 [FREE Full text] [doi: 10.2196/jmir.5352] [Medline: 26957354]

28. Lindhiem O, Bennett CB, Trentacosta CJ, McLear C. Client preferences affect treatment satisfaction, completion, and clinical outcome: a meta-analysis. Clin Psychol Rev 2014 Aug;34(6):506-517 [FREE Full text] [doi: 10.1016/j.cpr.2014.06.002] [Medline: 25189522]
29. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
30. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013 Aug;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
31. Shim M, Mahaffey B, Bleidistel M, Gonzalez A. A scoping review of human-support factors in the context of Internet-based psychological interventions (IPIs) for depression and anxiety disorders. Clin Psychol Rev 2017 Nov;57:129-140. [doi: 10.1016/j.cpr.2017.09.003] [Medline: 28934623]
32. Dear BF, Staples LG, Terides MD, Fogliati VJ, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for Social Anxiety Disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]
33. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided v. self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]
34. Dear BF, Fogliati VJ, Fogliati R, Johnson B, Boyle O, Karin E, et al. Treating anxiety and depression in young adults: A randomised controlled trial comparing clinician-guided versus self-guided Internet-delivered cognitive behavioural therapy. Aust N Z J Psychiatry 2018 Jul;52(7):668-679. [doi: 10.1177/0004867417738055] [Medline: 29064283]
35. Alfonsson S, Olsson E, Hursti T. The effects of therapist support and treatment presentation on the clinical outcomes of an Internet based applied relaxation program. Internet Interventions 2015 Sep;2(3):289-296 [FREE Full text] [doi: 10.1016/j.invent.2015.07.005]
36. Kobak KA, Greist R, Jacobi DM, Levy-Mack H, Greist JH. Computer-assisted cognitive behavior therapy for obsessive-compulsive disorder: a randomized trial on the impact of lay vs. professional coaching. Ann Gen Psychiatry 2015;14:10 [FREE Full text] [doi: 10.1186/s12991-015-0048-0] [Medline: 25722737]
37. Mak WW, Chio FH, Chan AT, Lui WW, Wu EK. The Efficacy of Internet-Based Mindfulness Training and Cognitive-Behavioral Training With Telephone Support in the Enhancement of Mental Health Among College Students and Young Working Adults: Randomized Controlled Trial. J Med Internet Res 2017 Dec 22;19(3):e84 [FREE Full text] [doi: 10.2196/jmir.6737] [Medline: 28330831]
38. Allexandre D, Bernstein AM, Walker E, Hunter J, Roizen MF, Morledge TJ. A Web-Based Mindfulness Stress Management Program in a Corporate Call Center: A Randomized Clinical Trial to Evaluate the Added Benefit of Onsite Group Support. J Occup Environ Med 2016 Mar;58(3):254-264 [FREE Full text] [doi: 10.1097/JOM.0000000000000680] [Medline: 26949875]
39. Renfrew ME, Morton DP, Morton JK, Hinze JS, Beamish PJ, Przybylko G, et al. A Web- and Mobile App-Based Mental Health Promotion Intervention Comparing Email, Short Message Service, and Videoconferencing Support for a Healthy Cohort: Randomized Comparative Study. J Med Internet Res 2020 Jan 06;22(1):e15592 [FREE Full text] [doi: 10.2196/15592] [Medline: 31904578]
40. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992 Jul 18;305(6846):160-164 [FREE Full text] [doi: 10.1136/bmj.305.6846.160] [Medline: 1285753]
41. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
42. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. Soc Indic Res 2009 May 28;97(2):143-156 [FREE Full text] [doi: 10.1007/s11205-009-9493-y]
43. Diener E, Emmons R, Larsen R, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
44. Pavot W, Diener E. The Satisfaction With Life Scale and the emerging construct of life satisfaction. The Journal of Positive Psychology 2008 Apr;3(2):137-152 [FREE Full text] [doi: 10.1080/17439760701756946]
45. Ebert DD, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Sep 01;42(5):382-394 [FREE Full text] [doi: 10.5271/sjweh.3573] [Medline: 27249161]
46. Ebert D, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. European Psychologist 2018 May;23(2):167-187 [FREE Full text] [doi: 10.1027/1016-9040/a000318]

47. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
48. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
49. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
50. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
51. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
52. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
53. Dear BF, Gandy M, Karin E, Fogliati R, Fogliati VJ, Staples LG, et al. The Pain Course: 12- and 24-Month Outcomes From a Randomized Controlled Trial of an Internet-Delivered Pain Management Program Provided With Different Levels of Clinician Support. J Pain 2018 Dec;19(12):1491-1503. [doi: 10.1016/j.jpain.2018.07.005] [Medline: 30099209]
54. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth Interventions Delivering Home-based Support Group Videoconferencing: Systematic Review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
55. Damianakis T, Tough A, Marziali E, Dawson DR. Therapy Online: A Web-Based Video Support Group for Family Caregivers of Survivors With Traumatic Brain Injury. J Head Trauma Rehabil 2016;31(4):E12-E20. [doi: 10.1097/HTR.0000000000000178] [Medline: 26291634]
56. Nyström K, Ohrling K. Parental support: mothers' experience of electronic encounters. J Telemed Telecare 2006;12(4):194-197. [doi: 10.1258/135763306777488726] [Medline: 16774701]
57. Wild S, Hanley J, Lewis S, McKnight J, McCloughan L, Padfield P, et al. The impact of supported telemetric monitoring in people with type 2 diabetes: study protocol for a randomised controlled trial. Trials 2013 Jul 06;14:198 [FREE Full text] [doi: 10.1186/1745-6215-14-198] [Medline: 23829417]
58. Marziali E, Donahue P. Caring for others: Internet video-conferencing group intervention for family caregivers of older adults with neurodegenerative disease. Gerontologist 2006 Jun;46(3):398-403. [doi: 10.1093/geront/46.3.398] [Medline: 16731880]
59. Paslakis G, Fischer-Jacobs J, Pape L, Schiffer M, Gertges R, Tegtbur U, et al. Assessment of Use and Preferences Regarding Internet-Based Health Care Delivery: Cross-Sectional Questionnaire Study. J Med Internet Res 2019 May 16;21(5):e12416 [FREE Full text] [doi: 10.2196/12416] [Medline: 31099338]
60. Williams R, Farquharson L, Palmer L, Bassett P, Clarke J, Clark DM, et al. Patient preference in psychological treatment and associations with self-reported outcome: national cross-sectional survey in England and Wales. BMC Psychiatry 2016 Jan 15;16:4 [FREE Full text] [doi: 10.1186/s12888-015-0702-8] [Medline: 26768890]
61. Dunlop BW, Kelley ME, Aponte-Rivera V, Mletzko-Crowe T, Kinkead B, Ritchie JC, PReDICT Team. Effects of Patient Preferences on Outcomes in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) Study. Am J Psychiatry 2017 Dec 01;174(6):546-556. [doi: 10.1176/appi.ajp.2016.16050517] [Medline: 28335624]
62. Beasley MJ, Ferguson-Jones EA, Macfarlane GJ. Treatment expectations but not preference affect outcome in a trial of CBT and exercise for pain. Can J Pain 2017;1(1):161-170 [FREE Full text] [doi: 10.1080/24740527.2017.1384297] [Medline: 29521378]
63. Hadjistavropoulos HD, Schneider LH, Mehta S, Karin E, Dear BF, Titov N. Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. J Anxiety Disord 2019 Apr;63:51-60 [FREE Full text] [doi: 10.1016/j.janxdis.2019.02.002] [Medline: 30844613]
64. Ajzen I. From intentions to actions: A theory of planned behavior. In: Action control. New York: Springer; 1985:11-39.
65. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Computers in Human Behavior 2009 May;25(3):749-760 [FREE Full text] [doi: 10.1016/j.chb.2009.02.003]
66. Waters LA, Galichet B, Owen N, Eakin E. Who participates in physical activity intervention trials? J Phys Act Health 2011 Jan;8(1):85-103. [doi: 10.1123/jpah.8.1.85] [Medline: 21297189]

Abbreviations

ANZCTR: Australian New Zealand Clinical Trials Registry

DASS: Depression, Anxiety and Stress Scales
eLMS: electronic learning management system
MHPI: mental health promotion intervention
RCT: randomized controlled trial
S: standard (automated emails only)
S+VCS: standard plus videoconferencing support
S+pSMS: standard plus personalized SMS
SF-36: Short Form Health Survey
SWLS: Satisfaction With Life Scale
VCS: videoconference support session

Edited by G Eysenbach; submitted 07.05.20; peer-reviewed by I Choi, JR Bautista; comments to author 15.06.20; revised version received 22.07.20; accepted 03.08.20; published 29.09.20.

Please cite as:
Renfrew ME, Morton DP, Morton JK, Hinze JS, Przybylko G, Craig BA
The Influence of Three Modes of Human Support on Attrition and Adherence to a Web- and Mobile App–Based Mental Health Promotion Intervention in a Nonclinical Cohort: Randomized Comparative Study
J Med Internet Res 2020;22(9):e19945
URL: http://www.jmir.org/2020/9/e19945/
doi:10.2196/19945
PMID:32990633

Original Paper

Web-Based Technology for Remote Viewing of Radiological Images: App Validation

Qiusha Min[1*], PhD; Xin Wang[1*], BSc; Bo Huang[2*], MBBS; Liangzhou Xu[2], MBBS

[1]School of Educational Information Technology, Central China Normal University, Wuhan, Hubei, China

[2]Department of Radiology, Wuhan Hospital of Traditional Chinese Medicine, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Qiusha Min, PhD
School of Educational Information Technology
Central China Normal University
No.152 Luoyu Road
Wuhan, Hubei, 430079
China
Phone: 86 27 67867597
Email: qiusham@mail.ccnu.edu.cn

Abstract

Background: Internet technologies can create advanced and rich web-based apps that allow radiologists to easily access teleradiology systems and remotely view medical images. However, each technology has its own drawbacks. It is difficult to balance the advantages and disadvantages of these internet technologies and identify an optimal solution for the development of medical imaging apps.

Objective: This study aimed to compare different internet platform technologies for remotely viewing radiological images and analyze their advantages and disadvantages.

Methods: Oracle Java, Adobe Flash, and HTML5 were each used to develop a comprehensive web-based medical imaging app that connected to a medical image server and provided several required functions for radiological interpretation (eg, navigation, magnification, windowing, and fly-through). Java-, Flash-, and HTML5-based medical imaging apps were tested on different operating systems over a local area network and a wide area network. Three computed tomography colonography data sets and 2 ordinary personal computers were used in the experiment.

Results: The experimental results demonstrated that Java-, Flash-, and HTML5-based apps had the ability to provide real-time 2D functions. However, for 3D, performances differed between the 3 apps. The Java-based app had the highest frame rate of volume rendering. However, it required the longest time for surface rendering and failed to run surface rendering in macOS. The HTML5-based app had the fastest surface rendering and the highest speed for fly-through without platform dependence. Volume rendering, surface rendering, and fly-through performances of the Flash-based app were significantly worse than those of the other 2 apps.

Conclusions: Oracle Java, Adobe Flash, and HTML5 have individual strengths in the development of remote access medical imaging apps. However, HTML5 is a promising technology for remote viewing of radiological images and can provide excellent performance without requiring any plug-ins.

(J Med Internet Res 2020;22(9):e16224) doi:10.2196/16224

KEYWORDS

internet access; medical informatics applications; computer-assisted image analyses; computer-assisted three-dimensional imaging; medical imaging; radiology; application

Introduction

Recently, modern technology has made it possible to generate digital images using medical equipment. Compared with traditional film-based images, these types of images have several advantages (eg, they are easy to share, transmit, and process) [1]. These advantages promote the popularity of the digital imaging systems in hospitals all over the world, and offer the

possibility for remote viewing and processing. However, the successful implementation of a teleradiology system requires a fast network and easy access [2]. If the system does not meet these requirements, radiologists may be reluctant to use the teleradiology system.

Internet technologies can create advanced and rich web-based apps that allow radiologists to easily access teleradiology systems and remotely view medical images. Compared with picture archiving and communication systems or other imaging workstations which require dedicated hardware and software, a web-based app is easy to set up and has a low cost [3]. These apps can be run on almost all personal computers without the need for powerful equipment on the client side. There are 3 major internet technologies, Oracle Java [4], Adobe Flash [5], and HTML5 [6], to create these apps. In the past few decades, these 3 technologies have been used in the field of medical imaging [7-16]; however, each technology has its own drawbacks. For example, plug-ins are required by Java and Flash. Regarding HTML5, the level of support and expected performance vary depending on the browser. Thus, it is difficult to balance the advantages and disadvantages of these internet technologies and identify an optimal solution for the development of medical imaging apps. Owing to the significant growth of teleradiology and web-based radiology subspecialty training, there is a need for quantitative and qualitative evaluations of different internet technologies in the field of medical imaging [17,18].

In this study, we used different technologies—Java (version 8; Oracle Corporation), Adobe Flash (version 32; Adobe Inc), and HTML5 (version 5.3; World Wide Web Consortium)—to develop web-based medical imaging apps. Subsequently, experiments were conducted to demonstrate the performance of these apps. Accordingly, the primary aim of this study was the evaluation of the performance of medical imaging apps developed with Oracle Java, Adobe Flash, and HTML5 in various scenarios. We also aimed to analyze the advantages and disadvantages of these technologies in the field of medical imaging. We believe these performance comparisons can guide developers in their efforts to identify suitable technologies to create web-based medical imaging apps, thus allowing radiologists to visualize and interpret images remotely, quickly, and effortlessly.

Methods

App Design

Medical imaging apps have several basic functions. First, the app needs to interact directly with the local file system to avoid network latency. The user can then use various 2D image processing tools, such as zooming and windowing, to identify useful information contained in the 2D image. In addition, the interpretation may be supported by 3D functions so that the volumetric data set can provide additional information on the anatomy or pathology of the patient [19]. Consequently, comprehensive medical imaging apps should meet the following requirements: (1) interact with local file systems, (2) have basic 2D image processing functions, and (3) allow 3D visualization of selected regions of interest in the data sets.

In this study, 3 demo apps for computed tomography (CT) colonography (also known as virtual colonoscopy) were developed using Oracle Java, Adobe Flash, and HTML5. These apps were designed to satisfy the aforementioned criteria, and they were used to evaluate Oracle Java, Adobe Flash, and HTML5 as tools to determine the best architecture for the development of a medical imaging app.

These apps provide remote access in such a way that radiologists can view images from a downloaded data set and manipulate them using 2D or 3D functions. They can be placed as a client component in a large teleradiology system. This study focuses solely on the client app and presents a comparison of 2D and 3D performance of the Oracle Java, Adobe Flash, and HTML5 technologies for the development of such medical imaging apps.

Operational Flow of Radiological Interpretation Using Demo Apps

Step 1

Select a data set to be interpreted and click the *Download* button to activate the download process. The selected data set is now stored on the local computer.

Step 2

The first slice of the data set is automatically displayed on the screen in the Java-based app. In the cases of the Flash and HTML5 apps, users need to click the *Choose file* button and select in the dialog box the downloaded file in either of the 2 apps. Thereafter, the first slice of the data set is displayed.

Step 3

Navigate through the image data set using the *Previous* and *Next* buttons or mouse scrolling.

Step 4

Interpret the data set using 2D image processing tools, such as zoom in, zoom out, and windowing.

Step 5

Interpret the data set using 3D visualization tools, such as 3D rendering and fly-through.

App Implementation

Access to the Local File System

All 3 apps enabled the user to choose a CT colonography data set for study (Figure 1). The selected data set was then transmitted to the client and stored on the local computer using a custom data format. Currently, Oracle Java, Adobe Flash, and HTML5 use different local file reading and writing technologies. Oracle Java downloaded the file via HTTP using the Java HttpURLConnection class. This class was used to read and write the resources referenced by a URL (uniform resource locator). Once the download was completed, the RandomAccessFile class was used to read local files in the Java-based app. Adobe Flash used the FileReference class to provide a safe way to directly read and write data to the local system (provided that the action was sanctioned by the user). Using this class in the app, the study data set was stored on the local computer disk and could then be navigated easily and

efficiently. HTML5 had a new input type <input type = “file”> which provided a standard way to interact with local files. After reading the downloaded file, the first slice in the data set was automatically displayed on the screen.

Figure 1. Graphical user interface used to display the first slice in the data set: (left) Java, (center) Flash, and (right) HTML5.

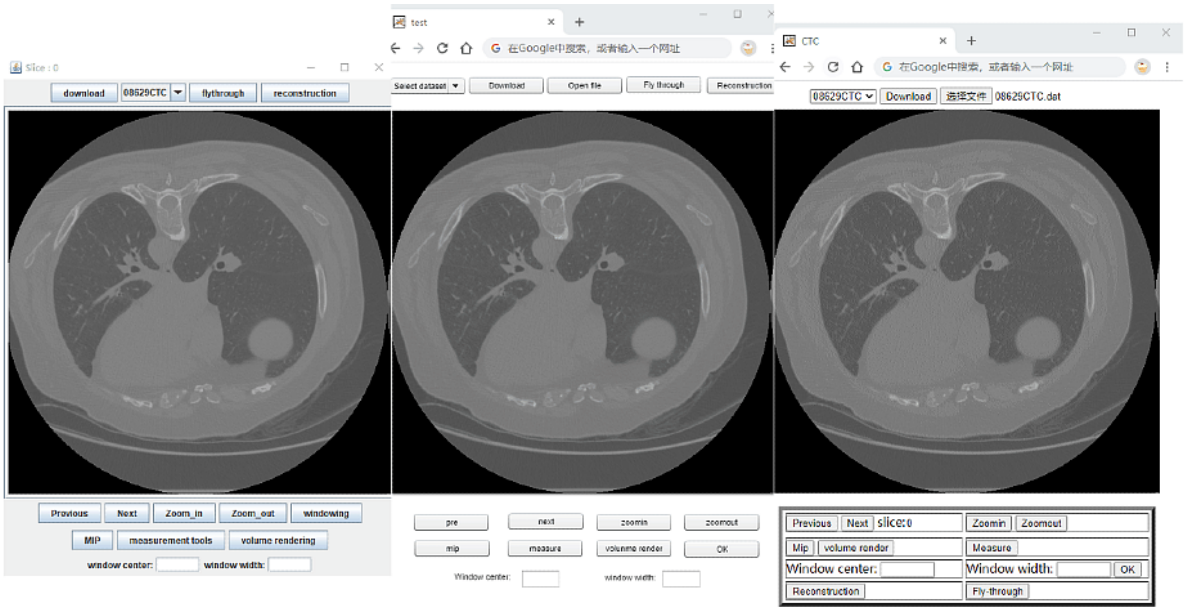


Image Processing

Another crucial requirement for these apps was pixel-level operation. For medical imaging apps, basic 2D image processing functions include magnification and windowing. For the magnification function, Java used *scaleX*() and *scaleY*() in the ScalePane class to zoom in and zoom out. Flash used the Zoom class to zoom in or out of the object. HTML5 used the canvas *drawImage*() method to zoom in and out. In terms of windowing, Java used *setRGB*() to set the pixels in BufferedImage to the specified RGB value. Flash was implemented using the BitmapData class. The *setPixel*() and *getPixel*() methods in the BitmapData class were used to change the value of each pixel in the image. For HTML5, a <canvas> element that has the ability to define the color of the pixels was used.

Figure 2 depicts a series of screenshots of the user interfaces of the windowing function in the 3 tested apps. The image is a windowed slice with the following parameters: the center of the window is 20 HU (where HU is Hounsfield units) and the width of the window is 200 HU. Figure 3 is a screenshot of the user interface of maximum intensity projection in Java-, Flash- and HTML5-based apps. Measurement and magnification functions can also be implemented by these 3 apps.

Figure 2. An anatomical axial slice at the thoracic level, wherein the center parameter is 20 HU, and the width is 200 HU. (left) Java, (center) Flash, and (right) HTML5. Screen resolution is 1920×1080.

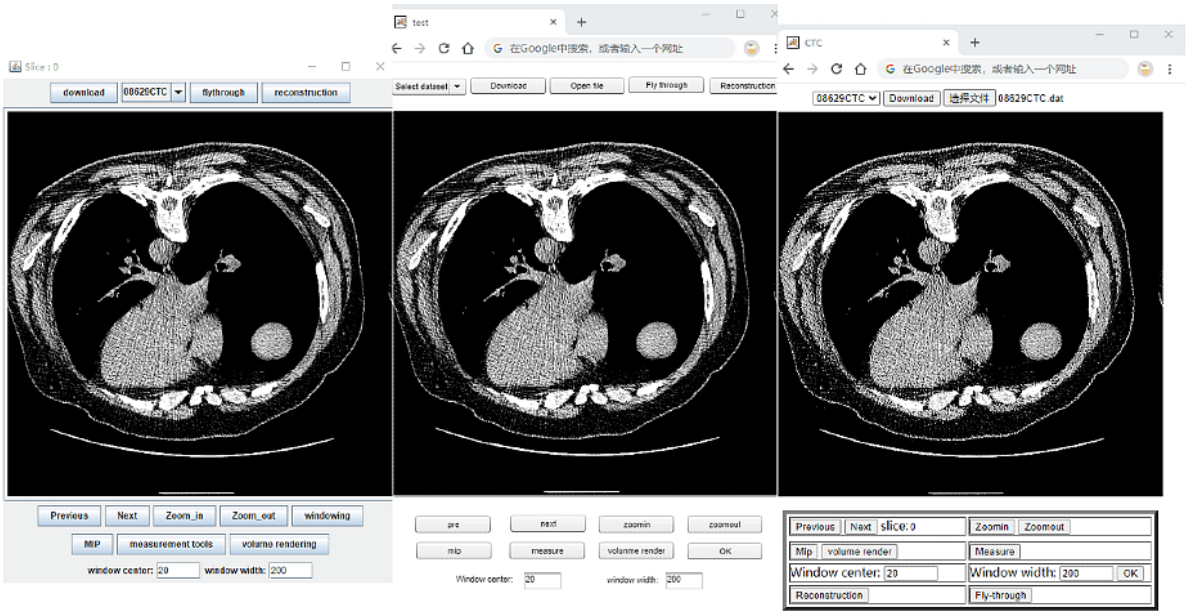
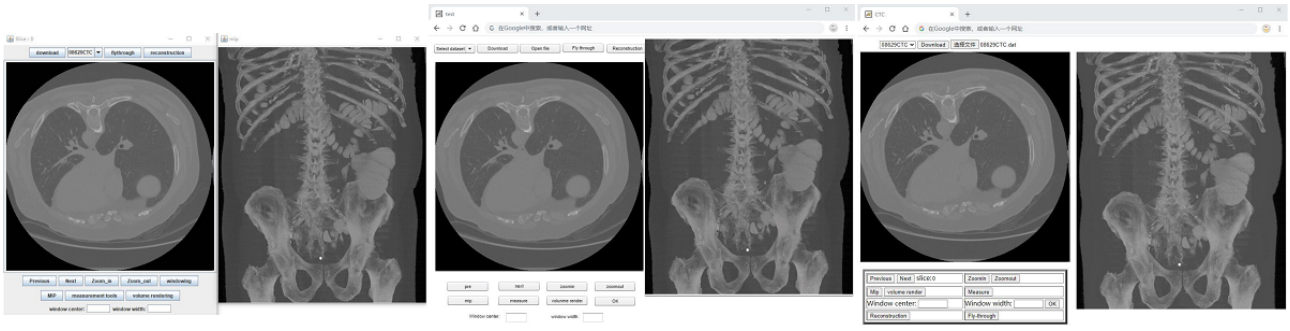


Figure 3. Graphical user interface and demonstration of maximum intensity projection. (left) Java, (center) Flash, and (right) HTML5.

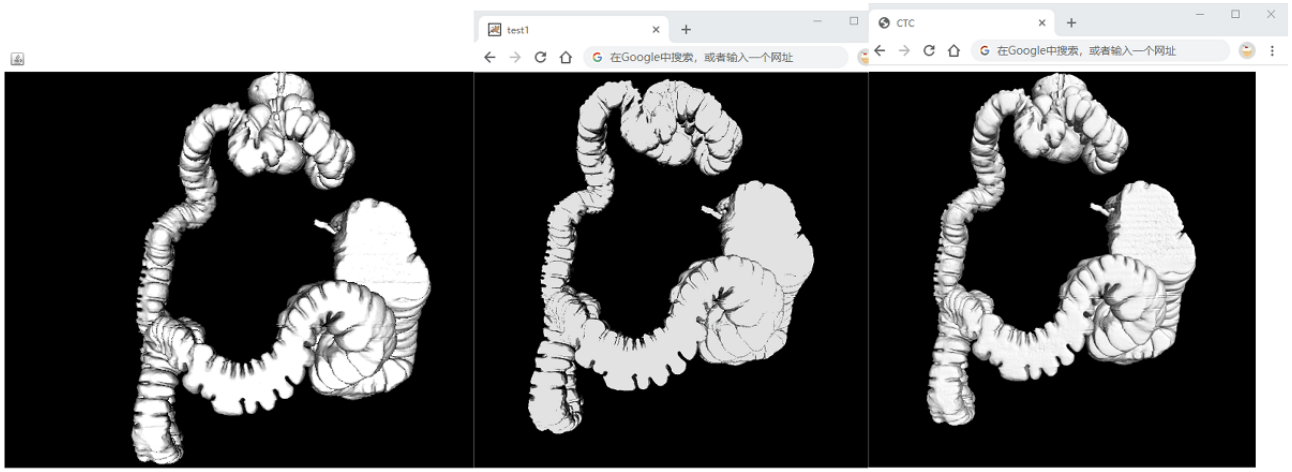


3D Visualization

It is well known that 3D visualization is a computationally intensive task. Hence, this work is typically implemented at a workstation equipped with a high-performance graphics processing unit (GPU). However, based on our previous research [11,13,15], it is feasible to implement 3D visualization on a personal computer using Oracle Java, Adobe Flash, or HTML5. In this implementation, 3D visualization was based on both surface rendering and volume rendering.

Surface rendering generally involved 2 stages: surface extraction and 3D rendering. The marching cubes algorithm was used to extract the isosurface from a volumetric data set [20]. The information of the extracted surface (ie, the vertex and the normal) were stored on a server. Once the user sent the request to view the 3D data, the corresponding vertex and normal files were sent to the client. Subsequently, the client side was responsible for rendering the 3D model surface. Currently, Oracle Java, Adobe Flash, and HTML5 enable the provision of hardware 3D rendering, which is a fast rendering mode when compared to that of software rendering. Java used Canvas3D to implement 3D rendering. For Adobe Flash, a Context3D object was used in its 3D app programming interface (API). Using the *createVertexBuffer*() method, Flash could send the vertices and normals to the GPU directly and perform a fast reconstruction. The combination of HTML5 with WebGL could also realize fast 3D rendering. Figure 4 presents 3D colon models rendered by Java-, Flash-, and HTML5-based apps. The user could also interact with this model and perform various operations, such as rotation and translation, using the mouse.

Figure 4. Screenshot of a 3D model of the entire colon in the browser. (left) Java, (center) Flash, and (right) HTML5.



Virtual fly-through navigation is a function used to manipulate the results of 3D reconstruction. It creates a virtual camera within the colon that moves along a planned path (commonly referred to as the colon centerline [21]); the radiologist can observe the interior of the colon using the continuous movement of the camera. This advanced imaging technique can help the radiologist make more accurate judgments about the lesion. Figure 5 presents the results of different techniques for implementing 3D fly-through within the colon, running in a browser.

Volume rendering is another type of 3D visualization that can represent the interior information of the 3D data set. Our implementation of volume rendering was based on a ray casting algorithm due to its ability to render high quality images [22]. This technique involved intensive computations resulting in low rendering speeds; however, it was feasible to define a subvolume to represent a region of interest. Volume rendering was then applied in this subvolume. In our apps, the size of the subvolume was 100×100×100 pixels and could be selected by the user. Figure 6 presents a region of interest rendered by the ray casting algorithm in Java-, Flash-, and HTML5-based apps.

Figure 5. Screenshot of the implementation of the fly-through in the browser. (left) Java, (center) Flash, and (right) HTML5.

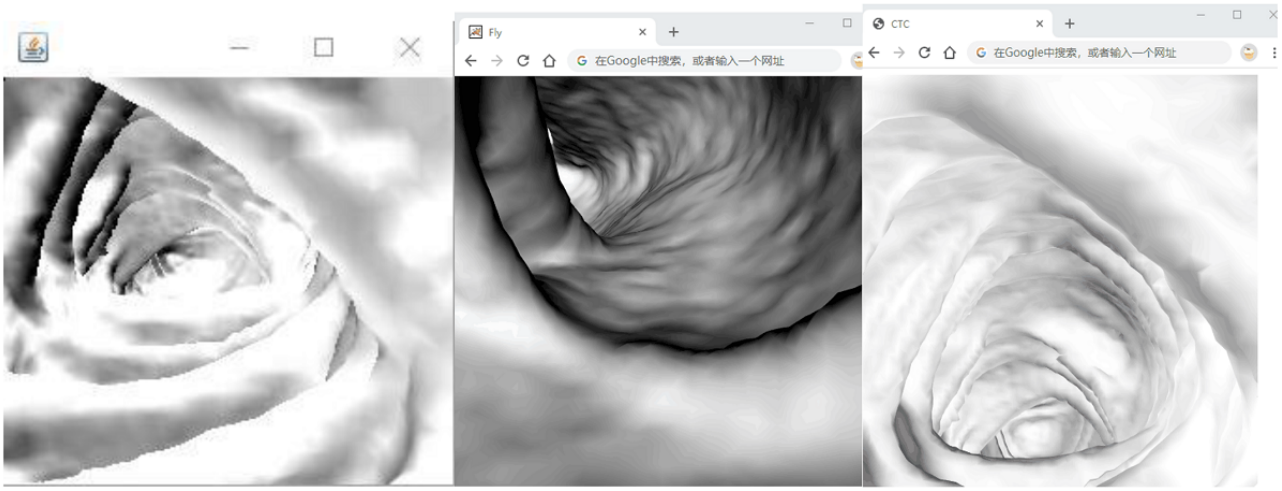
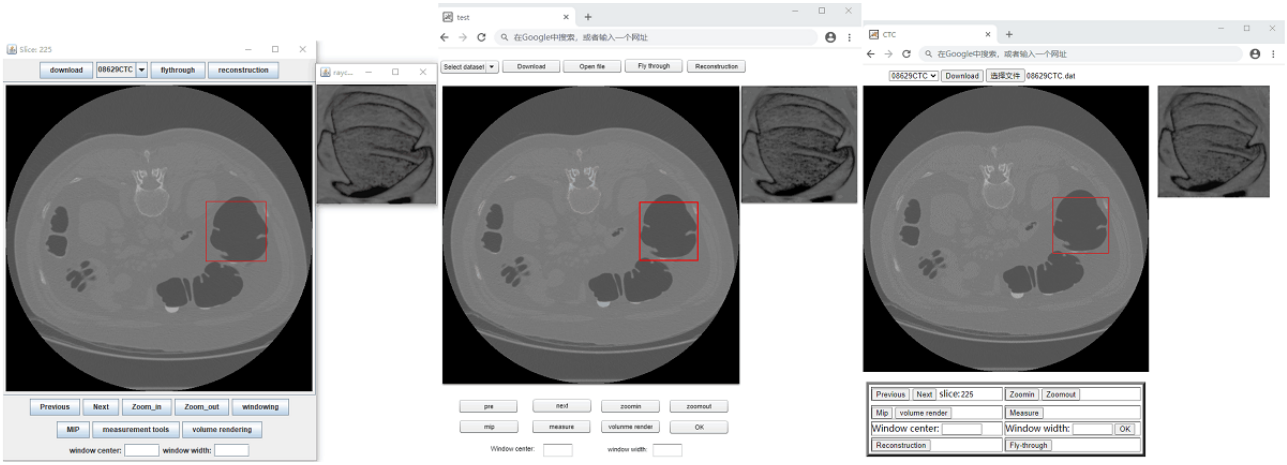


Figure 6. Screenshot of volume rendering in the browser. (left) Java, (center) Flash, and (right) HTML5.



Experiment Design

Overview

To compare the performances of the Java-, Flash-, and HTML5-based apps, 3 types of experiments were conducted: determining the performance of the apps running on the same platform (in Windows; Experiment 1); evaluating the performance of the apps on multiple platforms (Experiment 2); comparing the performances of the apps using a local area network (LAN) or a wide area network (WAN) (Experiment 3). CT colonography data sets (n=3), which were downloaded from The Cancer Imaging Archive [23], were used. Descriptions of the data sets are presented in Table 1, and the information about the computers used in the experiments is provided in Table 2. It is evident that the computers were ordinary personal computers for regular users.

Table 1. CT data sets used in the experiments.

Data set	Data set size Size		3D visualization		
	kB	pixel	Vertex file (kB)	Normal file (kB)	Number of faces
1	321,536	512×512×628	57,747	57,747	1,642,580
2	312,320	512×512×610	59,545	59,545	1,693,700
3	256,000	512×512×500	60,236	60,236	1,713,372

It should be noted that while the performances of Java- and Flash-based apps are browser independent, they are dependent on Java Virtual Machine and Flash Player, respectively. However, HTML5 is solely dependent on the browser, and our previous research [15] has demonstrated that Google Chrome can provide stable and excellent HTML5 performance. Thus, Google Chrome was used to run the HTML5-based app as well as to launch the Flash- and Java-based apps in this experiment. The details of the testing metrics in this study are presented in Table 3.

Table 2. Computers used in the experiments.

Computer	Type	Operating system	CPU[a]	Memory	GPU[b]
1	Desktop	Windows 10, 64 bits	Intel Core i5-8400 @2.80 GHz	16.00 GB	NVIDIA GeForce GTX 1050Ti
2	Desktop	Ubuntu 20.04	Intel Core i5-8400 @2.80 GHz	16.00 GB	NVIDIA GeForce GTX 1050Ti
3	Laptop (MacBook Pro)	MacOS Sierra 10.12.5	Intel Core Intel i5 @ 2.30 GHz	8.00 GB	Intel Iris Plus Graphics 640

[a]CPU: central processing unit.

[b]GPU: graphics processing unit.

Table 3. Details of testing metrics of this study.

Function	Label	Description	Measurement		
			Java	Adobe Flash	HTML5
Data access					
	M1	Execution time for downloading a medical image data set	Java	ActionScript 3.0	Manual
2D image processing					
	M2	Execution time for viewing a slice in a medical image data set	Java	ActionScript 3.0	JavaScript
	M3	Execution time for implementing windowing per slice	Java	ActionScript 3.0	JavaScript
	M4	Execution time for implementing magnification per slice	Java	ActionScript 3.0	JavaScript
	M5	Execution time for implementing mouse wheel per slice	Java	ActionScript 3.0	JavaScript
	M6	Execution time for implementing measure tool per slice	Java	ActionScript 3.0	JavaScript
	M7	Execution time for implementing maximum intensity projection	Java	ActionScript 3.0	JavaScript
3D visualization					
	M8	Execution time for downloading the vertex and normal files of a medical image data set	Java	ActionScript 3.0	Manual
	M9	Execution time for rendering a 3D model based on the downloaded vertex and normal files	Java	ActionScript 3.0	JavaScript
	M10	Frame rate of fly-through	Java	ActionScript 3.0	JavaScript
	M11	Frame rate of software-based volume rendering	Java	ActionScript 3.0	JavaScript

Experiment 1

The first experiment was carried out with 3 data sets using LAN to compare the performances of different apps running on the same platform. Computer 1 was chosen to test the Java-, Flash-, and HTML5-based apps on a Windows operating system. Each function was implemented 20 times in each app.

Experiment 2

The second experiment was used to determine the performance consistency of different internet technologies among multiple platforms. Computers 1, 2, and 3 were used in this experiment. Therefore, Java-, Flash-, and HTML5-based apps were run on Windows, macOS, and Linux platforms, respectively. Each function in the 3 apps was implemented 20 times on these platforms using data set 1 over a LAN.

Experiment 3

In the third experiment, computer 1 was used to evaluate the performances of Java-, Flash-, and HTML5-based apps over the LAN and WAN. This computer was equipped with Windows, and hence all the 3 apps were tested on the same platform. All 3 data sets were used in the experiment. Java-, Flash-, and HTML5-based apps were tested based on data sets 1, 2, and 3, to determine the performance differences when they ran over a LAN and WAN. In WAN, the 3 apps accessed the medical data set and vertex and normal files on the remote server. The bandwidth of the connecting network was 50 Mbps, and it had a download speed of approximately 5.1 MB/s. The download sizes for the medical image data set, vertex file, and normal file are listed in Table 2.

Each function in the app was implemented 20 times by Java Virtual Machine (version 1.8.0), Flash Player (version 32.0.0.433), and Chrome (version 83.0.4103.97), either over the LAN or WAN.

Radiologist Feedback

In order to collect feedback on the apps, we conducted a pilot trial. Radiologists (n=5) at Wuhan Hospital of Traditional Chinese Medicine participated in this trial. They first received a brief introduction of the project and our medical imaging apps.

After that, each of them was required to interpret 10 CT examinations using the Java-based app, 10 CT examinations using the Flash-based app, and 10 CT examinations using the HTML5-based apps on a Windows personal computer on the LAN.

After the trial, they filled out questionnaires (Multimedia Appendix 1). A 5-point Likert scale was used to represent the radiologists' opinions on a particular question or statement: strongly disagree, disagree, unsure, agree, and strongly agree. Radiologists' responses were recorded on a 1-5 scale, with higher numbers representing stronger agreement.

Results

Experiment 1: General Performance

The average performances for each function are presented in Table 4 (Experiment 1). The comparison revealed that each technology had its own advantages. Java was associated with the shortest downloading time and highest frame rate for software-based volume rendering. However, it performed poorly at surface rendering. HTML5 surface rendering performed best. In terms of 2D functions, such as zooming and windowing, all 3 apps performed similarly. Overall, HTML5 outperformed the other 2 technologies, with the exception of downloading and software-based volume rendering.

Table 4. Comparison of the performances of the 3 apps in various scenarios.

Technology	Data access	2D image processing						3D visualization			
	M1 (s[a])	M2 (s)	M3 (s)	M4 (s)	M5 (s)	M6 (s)	M7 (s)	M8 (s)	M9 (s)	M10 (fps[b])	M11 (fps)
Experiment 1											
Data set 1											
Java (Windows)	29.87	0.0073	0.0142	0.0015	0.0077	0.0010	0.3446	10.76	271.21	32.25	1.92
Flash (Windows)	27.76	0.0144	0.0193	0.0004	0.0146	0.0001	26.7071	10.24	16.20	10.21	0.05
HTML5 (Windows)	28.58	0.0009	0.0133	0.0006	0.0008	0.0002	2.4036	13.46	1.09	59.30	0.56
Data set 2											
Java (Windows)	29.16	0.0066	0.0094	0.0016	0.0058	0.0010	0.3701	10.29	280.54	34.23	1.97
Flash (Windows)	26.97	0.0148	0.0189	0.0007	0.0149	0.0002	27.5514	10.35	17.59	9.91	0.05
HTML5 (Windows)	27.98	0.0008	0.0154	0.0004	0.0009	0.0003	2.4560	12.38	1.25	60.05	0.61
Data set 3											
Java (Windows)	23.96	0.0046	0.0098	0.0018	0.0049	0.0008	0.2923	11.13	283.79	36.25	2.24
Flash (Windows)	22.11	0.0142	0.0186	0.0005	0.0144	0.0002	21.1775	10.42	18.72	10.13	0.06
HTML5 (Windows)	23.19	0.0008	0.0151	0.0005	0.0008	0.0002	1.9772	12.37	1.43	60.10	0.71
Experiment 2											
Data set 1											
Java (Windows)	29.87	0.0073	0.0142	0.0015	0.0077	0.0010	0.3446	10.76	271.21	32.25	1.92
Flash (Windows)	27.76	0.0144	0.0193	0.0004	0.0146	0.0001	26.7071	10.24	16.20	10.21	0.05
HTML5 (Windows)	28.58	0.0009	0.0133	0.0006	0.0008	0.0002	2.4036	13.46	1.09	59.30	0.56
Data set 1											
Java (macOS)	28.93	0.0132	0.0189	0.0028	0.0133	0.0002	0.3483	10.27	__[c]	—	1.42
Flash (macOS)	27.76	0.0207	0.0232	0.0002	0.0191	0.0006	21.4324	9.98	13.59	10.07	0.07
HTML5 (macOS)	29.04	0.0011	0.0194	0.0009	0.0012	0.0002	2.6352	12.31	0.9285	60.20	0.54
Data set 1											
Java (Linux)	19.61	00113	0.0177	0.0052	0.0104	0.0002	0.4398	8.49	77.84	19.33	1.68
Flash (Linux)	27.75	0.0291	0.0513	0.0003	0.0325	0.0004	16.3715	9.98	15.12	9.26	0.11
HTML5 (Linux)	33.22	0.0021	0.0414	0.0008	0.0019	0.0004	3.7369	13.95	1.39	58.7	0.37
Experiment 3											
Data set 1											

Technology	Data access	2D image processing					3D visualization				
	M1 (s[a])	M2 (s)	M3 (s)	M4 (s)	M5 (s)	M6 (s)	M7 (s)	M8 (s)	M9 (s)	M10 (fps[b])	M11 (fps)
Java (LAN)	29.87	0.0073	0.0142	0.0015	0.0077	0.0010	0.3446	10.76	271.21	32.25	1.92
Java (WAN)	311.47	0.0071	0.0153	0.0011	0.0072	0.0010	0.3417	169.89	268.65	32.23	1.89
Data set 2											
Flash (LAN)	26.97	0.0148	0.0189	0.0007	0.0149	0.0002	27.5514	10.35	17.59	9.91	0.05
Flash (WAN)	464.05	0.0145	0.0191	0.0006	0.0148	0.0003	27.4085	176.42	69.18	9.17	0.05
Data set 3											
HTML5 (LAN)	23.19	0.0008	0.0151	0.0005	0.0008	0.0002	1.9772	12.37	1.43	60.10	0.71
HTML5 (WAN)	377.51	0.0009	0.0162	0.0008	0.0008	0.0006	2.0147	186.39	1.31	60.02	0.72

[a]s: seconds.

[b]fps: frames per second.

[c]Not tested in macOS.

Experiment 2: Performance on Multiple Platforms

The average performances of all the functions are presented in Table 4 (Experiment 2). It can be observed from this table that although the 2D performances of Java-, Flash-, and HTML5-based apps running on multiple platforms (Windows, macOS, and Linux) were almost the same, there are obvious differences in 3D performance. Owing to the facts that Java3D is obsolete and the configuration in macOS was much more complicated than expected, surface rendering by Java was not tested in macOS but only tested in Windows and Linux. In terms of the 3D performance on different platforms, Java-based apps on Windows achieved better performance than on Linux and macOS. However, Flash- and HTML5-based apps demonstrated consistent performance across different platforms.

Experiment 3: Performance based on LAN and WAN

The average performances for each function are presented in Table 4 (Experiment 3). The results of the performances of Java-, Flash-, and HTML5-based apps over LAN and WAN revealed that there was little difference between the LAN and WAN, except for downloading. Given that the data transmission speed over the WAN was lower than that over the LAN, the downloading time was different, as expected. After downloading data to the client computer, the app performance over the WAN was the same as that over the LAN.

Summary

The experimental results demonstrated that Java-, Flash-, and HTML5-based apps have the ability to yield real-time performances for all 2D functions. However, the 3D performances differed between the 3 apps. In terms of software-based volume rendering, the Java-based app had the highest frame rate; however, it required the longest amount of time for surface rendering and failed to run surface rendering in macOS. In terms of surface rendering, the HTML5-based app had the fastest rendering and the highest speed for fly-through without platform dependence. However, the frame rate of software-based volume rendering by HTML5 was slightly lower than that by Java. The 3D performances of the Flash-based app were worse than both of the other apps.

Pilot Trial and Feedback From Radiologists

The results of radiologists' responses are presented in Table 5.

Most radiologists were satisfied with the functions that we provided. However, they were not satisfied with 3D functions in Java and Flash. Three radiologists reported that Java took a long time for surface rendering and Flash provided a significantly low frame rate for volume rendering. For question 8, every radiologist chose HTML5, which means that HTML5 obtained the highest satisfaction among these 3 technologies.

Table 5. Radiologists' responses to the questionnaires.

Radiologist	Question score[a]							
	1	2	3	4	5	6	7	8[b]
Radiologist 1	5	5	5	5	5	5	5	3
Radiologist 2	4	5	5	5	3	3	4	3
Radiologist 3	5	5	5	5	4	4	5	3
Radiologist 4	5	5	5	5	4	4	5	3
Radiologist 5	5	5	5	5	5	5	5	3
Total score	24	25	25	25	21	21	24	N/A[c]

[a]From 1 (strongly disagree) to 5 (strongly agree).

[b]Options: 1 (Java), 2 (Flash), 3 (HTML5).

[c]N/A: not applicable.

Discussion

Principal Findings

Currently, there are 3 main technologies for the development of web-based medical imaging apps, namely, Oracle Java, Adobe Flash, and HTML5. Around the 2000s, Oracle Java was a popular internet technology in the field of medical imaging [7-12,14,24]. Since 2010, Flash-based imaging apps have appeared, owing to the ubiquity and small size of the Flash Player [13,16]. Since the release of the World Wide Web Consortium HTML5 recommendation in 2014, there has been a growing trend toward the utilization of HTML5 in the development of medical imaging apps. McLaughlin et al [25] developed a digital training platform for interpreting radiographic images based on HTML5. Their platform had 2 tools, a search strategy training tool and an eye tracking tool, which were used to clarify the image interpretation process [25]. Gorgbjerg [26] presented an HTML5-based web app that could be manipulated as in a picture archiving and communication systems. Zhang [27] created a network-based medical data rendering and sharing system with a client app that was developed by HTML5. This client app has the ability to deliver real-time visualization on the web. Additionally, our previous study [15] provided an evaluation of HTML5 for medical imaging apps and demonstrated that HTML5 can provide an excellent remote access medical imaging experience.

In this study, 3 technologies were used to develop a comprehensive medical imaging app and to evaluate the performances of these technologies in the field of radiology. In terms of accessibility, both Java- and Flash-based apps require a browser plug-in. Despite the fact that the Flash Player has long been one of the most popular browser plug-ins, Apple decided to stop bundling Flash Player in macOS in 2010. Thus, for this group of users, to be able to run Flash-based apps, they must initially install Flash Player. Similarly, to be able to run Java, Java Virtual Machine must be installed. HTML5 does not suffer from this problem because it is the native language used in all browsers. Therefore, HTML5 requires no preinstallation and is a platform-independent technology that provides a high level of accessibility. However, the advantages associated with HTML5 exist only in the latest version of browsers. Older browser versions, such as Microsoft Internet Explorer 8 (or older versions), Mozilla Firefox 3.5 (or older versions), and Google Chrome 10 (or older versions), are not compatible with HTML5. In these cases, users would be required to update their browsers, otherwise, HTML5-based apps could not be launched in their browsers. Furthermore, browsers vary in their level of support for the HTML5 standard, and thus, this leads to inconsistent user experiences. For example, the implementation of the mouse wheel event is different among Internet Explorer, Chrome, Safari, and Firefox. In the case of Chrome, when the mouse wheel is rolled up, the value increases however, in Firefox, the value decreases.

In terms of functionality, all 3 technologies can realize the necessary functions for remote viewing of radiological images. Image processing, such as zooming and windowing, can be provided by all 3 technologies on all platforms. However, implementation of 3D visualization is more complicated than the implementation of image processing, especially for Oracle Java. Oracle Java realizes 3D surface rendering by depending on Java3D API. However, this API has not been updated since 2008. Hence, some problems emerge in recent implementations of Java3D (eg, the Java3D app failed to launch in macOS). 3D visualization by Adobe Flash and HTML5 can be successfully implemented, regardless of the platform. However, it should be noted that Adobe will terminate its support for Flash at the end of 2020. Thereafter, Flash 3D API and Flash Player will not be updated. In this case, only HTML5 has an advanced API and hence can provide a higher level of functionality (compared to Oracle Java and Adobe Flash).

In terms of usability, the experimental results reveal that all 3 technologies can provide 2D image processing on all platforms. However, the 3D performances of these technologies are different. Among these technologies, HTML5 presents the best surface rendering performances in terms of rendering time and frame rate. In terms of volume rendering, HTML5 is not good at software-based volume rendering. However, when integrated with a GPU, HTML5 can provide fast hardware-based volume rendering [28,29].

In terms of interoperability, Oracle Java, Adobe Flash, and HTML5 are designed for developing rich web apps. Therefore, all 3 apps can be connected to a large teleradiology system and

placed as client components. Moreover, the source code of HTML5 is exposed online, and therefore, the locations of image data sets can be easily assessed. The source codes of Java and Flash are hidden (inside .JAR and .SWF files, respectively), which prevents unauthorized access to image data sets. Thus, Java and Flash outperform HTML5 in data privacy.

Recently, cloud computing has been used in the field of medical imaging for high-capacity storage, sharing, and intensive computational tasks [30,31]. In this infrastructure, the image data and complex processing tasks are moved from user computers to the cloud. The users then launch an app to access the cloud. In this case, a radiologist can implement the cloud-based medical image analysis using a personal computer from any location. Furthermore, web technology supports the development of the client app in the cloud-based system. With its development, the client app can become more powerful than before. Among these web technologies, HTML5 can develop a zero-footprint web viewer, which requires zero plug-ins, zero latency, and zero maintenance. Therefore, most commercially web-based DICOM (Digital Imaging and Communications in Medicine standard) viewers, such as Ambra [32], medDream [33], and boxDicom [34], switched to an HTML5-based solution recently. All of them can be integrated into any picture archiving and communication systems system. Additionally, medDream provides 3D features, such as maximum intensity projection and 3D rendering, in a browser. We confirmed that their HTML5 solutions can implement necessary interpretation tools, such as 2D image processing and 3D visualization, inside the client browser with satisfactory performance.

Although web technology enables remote viewing of radiological images easily and efficiently, there are still 4 issues affecting current web-based medical imaging apps. First, when data are transmitted over the internet, security is the biggest challenge, and this has encouraged many researches to find ways to keep medical images safe and confidential [35,36]. Second, remote viewing of radiological images is heavily dependent on the network. When internet connections are slow or unavailable, our web apps cannot work properly; hence, network condition is an important factor in teleradiology settings. Third, the specifications of a personal computer are usually inferior to those of dedicated workstations, and therefore, intensive computational tasks, such as volume rendering, cannot be implemented on a personal computer. In our implementation of volume rendering, the rendering region was reduced in order to provide fast volume rendering. Therefore, some interpretation tools need to be customized and simplified in web-based apps. Furthermore, typical medical image sizes range from 512×512×8 bits up to 1024×1024×12 bits. For some imaging apps, the resolution is even higher. Therefore, the client's screen should support higher resolutions, otherwise the medical images cannot be properly displayed.

Limitations

Our study and the web-based apps that were developed also have some limitations. First, we were able to obtain feedback from 5 radiologists to conduct a pilot testing, but we were not able to conduct a large and comprehensive investigation on users' opinions. Therefore, the feedback of users may contain deviations due to a small sample size. Second, only 2 personal computers, one on which Windows and Linux were installed, and another on which macOS was installed, were used in our experiments. The performances of apps may be affected by the hardware specifications. In future, upgraded computer hardware could enhance the performance of our apps.

Conclusion

Based on the review of existing literature, it is apparent that there is a lack of studies on the evaluation of different internet technologies for remote viewing of radiological images. In this study, 3 main internet technologies (ie, Oracle Java, Adobe Flash, and HTML5) were used to develop comprehensive web-based medical imaging apps. Experiments were conducted to compare these technologies in terms of accessibility, functionality, and usability. Moreover, advantages and disadvantages were discussed. Our research demonstrated that HTML5 is a promising technology for remote viewing of radiological images and can provide excellent performance without requiring any plug-ins. Therefore, our research provides an important reference for future development of web apps in the field of medical imaging, and it could help to identify an optimal solution for remote viewing of radiological images.

Acknowledgments

This work was supported by the self-determined research funds of Central China Normal University from the colleges' basic research and operation of Ministry of Education (No. CCNU20QN019).

Authors' Contributions

All authors contributed to the design of the study. BH and LX designed the medical imaging apps; QM and XW collected the data sets and developed the Java-, Flash-, and HTML5-based apps; all authors performed the experiments. QM wrote the main body of the manuscript, and all authors reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.

[DOCX File , 34 KB - jmir_v22i9e16224_app1.docx]

References

1. James J, Davies A, Cowen A, O'Connor P. Developments in digital radiography: an equipment update. Eur Radiol 2001 Mar 13;11(12):2616-2626. [doi: 10.1007/s003300100828] [Medline: 11734969]
2. Bowers GH, Steiner E, Kalman M. Implementing teleradiology in a private radiology practice: lessons learned. J Digit Imaging 1998 Aug;11(3 Suppl 1):96-98 [FREE Full text] [doi: 10.1007/BF03168271] [Medline: 9735443]
3. Fraternali P, Rossi G, Sánchez-Figueroa F. Rich internet application. IEEE Internet Comput 2010 May;14(3):9-12. [doi: 10.1109/mic.2010.76]
4. JDK 12 documentation. Oracle. URL: https://docs.oracle.com/en/java/javase/12/index.html [accessed 2019-07-05]
5. ActionScript 3.0. Reference for the Adobe Flash Platform. 2019. URL: https://help.adobe.com/en_US/FlashPlatform/reference/actionscript/3/ [accessed 2019-07-05]
6. Living standard. HTML. URL: http://w3c.github.io/html/ [accessed 2019-07-10]
7. Slomka PJ, Elliott E, Driedger AA. Java-based remote viewing and processing of nuclear medicine images: toward "the imaging department without walls". J Nucl Med 2000 Jan;41(1):111-118. [Medline: 10647613]
8. Knoll P, Höll K, Mirzaei S, Koriska K, Köhn H. Distributed nuclear medicine applications using World Wide Web and Java technology. Eur Radiol 2000 Aug 18;10(9):1483-1486. [doi: 10.1007/s003300000432] [Medline: 10997440]
9. Choi H, Park S, Kang J, Kim S, Choi H. Tele-medical imaging conference system based on the Web. Comput Methods Programs Biomed 2002 Jun;68(3):223-231. [doi: 10.1016/s0169-2607(01)00174-2] [Medline: 12074849]
10. Kamauu AWC, DuVall SL, Wiggins RH, Avrin DE. Using applet-servlet communication for optimizing window, level and crop for DICOM to JPEG conversion. J Digit Imaging 2008 Sep;21(3):348-354 [FREE Full text] [doi: 10.1007/s10278-007-9038-3] [Medline: 17534682]
11. Min Q, Sadleir R. Distributed medical imaging apps using Java technology. 2015 Dec 16 Presented at: International Congress on Image and Signal Processing BioMedical Engineering and Informatics (CISP-BMEI); 2015; Hong Kong, China p. 16-19. [doi: 10.1109/APSIPA.2015.7415343]
12. Looney PT, Young KC, Halling-Brown MD. MedXViewer: providing a web-enabled workstation environment for collaborative and remote medical imaging viewing, perception studies and reader training. Radiat Prot Dosimetry 2016 Jun;169(1-4):32-37. [doi: 10.1093/rpd/ncv482] [Medline: 26628613]
13. Min Q, Wang Z, Wu M. An evaluation of adobe flash for remote access medical imaging applications. 2016 Oct 15 Presented at: IEEE 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); 2016; Datong, China p. 15-17. [doi: 10.1109/cisp-bmei.2016.7852789]
14. Doel T, Shakir DI, Pratt R, Aertsen M, Moggridge J, Bellon E, et al. GIFT-Cloud: A data sharing and collaboration platform for medical imaging research. Comput Methods Programs Biomed 2017 Feb;139:181-190 [FREE Full text] [doi: 10.1016/j.cmpb.2016.11.004] [Medline: 28187889]
15. Min Q, Wang Z, Liu N. An evaluation of HTML5 and WebGL for medical imaging applications. J Healthc Eng 2018 Aug 29;2018:1592821 [FREE Full text] [doi: 10.1155/2018/1592821] [Medline: 30245782]
16. Arguinarena EJC, Macchi JE, Escobar P, Del Fresno M, Massa J, Santiago M. Dcm-ar: a fast flash-based Web-PACS viewer for displaying large DICOM images. In: Conf Proc IEEE Eng Med Biol Soc. 2010 Presented at: International Conference of the IEEE Engineering in Medicine and Biology; Aug 31-Sept 4; Buenos Aires, Argentina p. 3463-3466. [doi: 10.1109/IEMBS.2010.5627827]
17. Rosenkrantz AB, Hanna TN, Steenburg SD, Tarrant MJ, Pyatt RS, Friedberg EB. The current state of teleradiology across the United States: a national survey of radiologists' habits, attitudes, and perceptions on teleradiology practice. J Am Coll Radiol 2019 Dec;16(12):1677-1687. [doi: 10.1016/j.jacr.2019.05.053] [Medline: 31271736]
18. Haj-Mirzaian A, Sethi N, de Francesca B, Sahni S, Zaheer A. Web-based radiology subspecialty training program: pilot feasibility and effectiveness analysis on Ethiopian radiologists. Acad Radiol 2020 Feb;27(2):293-299. [doi: 10.1016/j.acra.2019.02.025] [Medline: 31060981]
19. Pickhardt PJ, Lee AD, Taylor AJ, Michel SJ, Winter TC, Shadid A, et al. Primary 2D versus primary 3D polyp detection at screening CT colonography. AJR Am J Roentgenol 2007 Dec;189(6):1451-1456. [doi: 10.2214/AJR.07.2291] [Medline: 18029884]
20. Lorensen W, Cline H. Marching cubes: a high resolution 3D surface construction algorithm. 1987 Presented at: SIGGRAPH '87: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques; August 1987; New York, NY, USA. [doi: 10.1145/37401.37422]
21. Sadleir RJ, Whelan PF. Fast colon centreline calculation using optimised 3D topological thinning. Comput Med Imaging Graph 2005 Jun;29(4):251-258. [doi: 10.1016/j.compmedimag.2004.10.002] [Medline: 15890252]
22. Levoy M. Display of surfaces from volume data. IEEE Comput Grap Appl 1988 May;8(3):29-37. [doi: 10.1109/38.511]
23. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013 Dec 25;26(6):1045-1057 [FREE Full text] [doi: 10.1007/s10278-013-9622-7] [Medline: 23884657]

24. Poliakov AV, Albright E, Hinshaw KP, Corina DP, Ojemann G, Martin RF, et al. Server-based approach to web visualization of integrated three-dimensional brain imaging data. J Am Med Inform Assoc 2005;12(2):140-151 [FREE Full text] [doi: 10.1197/jamia.M1671] [Medline: 15561787]
25. McLaughlin L, Woznitza N, Cairns A, McFadden S, Bond R, Hughes C, et al. Digital training platform for interpreting radiographic images of the chest. Radiography (Lond) 2018 May;24(2):159-164. [doi: 10.1016/j.radi.2017.12.010] [Medline: 29605114]
26. Borgbjerg J. MULRECON: a web-based imaging viewer for visualization of volumetric images. Curr Probl Diagn Radiol 2019;48(6):531-534. [doi: 10.1067/j.cpradiol.2018.09.001] [Medline: 30340913]
27. Zhang Q. Web-based medical data visualization and information sharing towards application in distributed diagnosis. Inform Med Unlocked 2019;14:69-81. [doi: 10.1016/j.imu.2018.10.010]
28. Arbelaiz A. Volume visualization tools for medical applications in ubiquitous platforms. In: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.: Springer, Cham; 2017 Presented at: International Summit on eHealth; June 14-16, 2016; Budapest, Hungary p. 443-450. [doi: 10.1007/978-3-319-49655-9_54]
29. Lesar Ž, Bohak C, Marolt M. Real-time interactive platform-agnostic volumetric path tracing in webGL 2.0. 2018 Jun 20 Presented at: Web3D '18: Proceedings of the 23rd International ACM Conference on 3D Web Technology; 2018; Poznań, Poland p. 1-7. [doi: 10.1145/3208806.3208814]
30. Kagadis GC, Kloukinas C, Moore K, Philbin J, Papadimitroulas P, Alexakos C, et al. Cloud computing in medical imaging. Med Phys 2013 Jul 21;40(7):070901. [doi: 10.1118/1.4811272] [Medline: 23822402]
31. Liu L, Chen W, Nie M, Zhang F, Wang Y, He A, et al. iMAGE cloud: medical image processing as a service for regional healthcare in a hybrid cloud environment. Environ Health Prev Med 2016 Nov 25;21(6):563-571 [FREE Full text] [doi: 10.1007/s12199-016-0582-7] [Medline: 27783315]
32. DICOM web viewer: vew from anywhere, anytime, across all imaging types. Ambra Health. URL: https://ambrahealth.com/products-and-services/dicom-web-viewer/ [accessed 2020-09-01]
33. HTML 5 zero-footprint DICOM viewer. MedDream. URL: https://www.softneta.com/products/meddream-dicom-viewer/ [accessed 2020-09-01]
34. Medical images for the cloud. BoxDicom. URL: https://boxdicom.com/ [accessed 2020-09-01]
35. Rodrigues JJPC, de la Torre I, Fernández G, López-Coronado M. Analysis of the security and privacy requirements of cloud-based electronic health records systems. J Med Internet Res 2013 Aug 21;15(8):e186 [FREE Full text] [doi: 10.2196/jmir.2494] [Medline: 23965254]
36. Blanquer I, Brasileiro F, Brito A, Calatrava A, Carvalho A, Fetzer C, et al. Federated and secure cloud services for building medical image classifiers on an intercontinental infrastructure. Future Generation Computer Systems 2020 Sep;110:119-134. [doi: 10.1016/j.future.2020.04.012]

Abbreviations

CT: computed tomography
DICOM: Digital Imaging and Communications in Medicine standard
GPU: graphics processing unit
HTML: hypertext markup language
LAN: local area network
WAN: wide area network

Edited by G Eysenbach; submitted 18.09.19; peer-reviewed by J Borgbjerg, A Korchi, P Lei, J Lee; comments to author 11.03.20; revised version received 21.07.20; accepted 11.08.20; published 25.09.20.

Please cite as:
Min Q, Wang X, Huang B, Xu L
Web-Based Technology for Remote Viewing of Radiological Images: App Validation
J Med Internet Res 2020;22(9):e16224
URL: http://www.jmir.org/2020/9/e16224/
doi:10.2196/16224
PMID:32975520

Review

Data Quality Issues With Physician-Rating Websites: Systematic Review

Pavankumar Mulgund[1], PhD; Raj Sharman[1], PhD; Priya Anand[2*], MS; Shashank Shekhar[1*], MS; Priya Karadi[2*], MS

[1]School of Management, State University of New York Buffalo, Buffalo, NY, United States

[2]Institute of Computational and Data Sciences, State University of New York Buffalo, Buffalo, NY, United States

[*]these authors contributed equally

Corresponding Author:
Pavankumar Mulgund, PhD
School of Management
State University of New York Buffalo
160 Jacobs Management Center
Buffalo, NY
United States
Phone: 1 7167481552
Email: pmulgund@buffalo.edu

Abstract

Background: In recent years, online physician-rating websites have become prominent and exert considerable influence on patients' decisions. However, the quality of these decisions depends on the quality of data that these systems collect. Thus, there is a need to examine the various data quality issues with physician-rating websites.

Objective: This study's objective was to identify and categorize the data quality issues afflicting physician-rating websites by reviewing the literature on online patient-reported physician ratings and reviews.

Methods: We performed a systematic literature search in ACM Digital Library, EBSCO, Springer, PubMed, and Google Scholar. The search was limited to quantitative, qualitative, and mixed-method papers published in the English language from 2001 to 2020.

Results: A total of 423 articles were screened. From these, 49 papers describing 18 unique data quality issues afflicting physician-rating websites were included. Using a data quality framework, we classified these issues into the following four categories: intrinsic, contextual, representational, and accessible. Among the papers, 53% (26/49) reported intrinsic data quality errors, 61% (30/49) highlighted contextual data quality issues, 8% (4/49) discussed representational data quality issues, and 27% (13/49) emphasized accessibility data quality. More than half the papers discussed multiple categories of data quality issues.

Conclusions: The results from this review demonstrate the presence of a range of data quality issues. While intrinsic and contextual factors have been well-researched, accessibility and representational issues warrant more attention from researchers, as well as practitioners. In particular, representational factors, such as the impact of inline advertisements and the positioning of positive reviews on the first few pages, are usually deliberate and result from the business model of physician-rating websites. The impact of these factors on data quality has not been addressed adequately and requires further investigation.

(J Med Internet Res 2020;22(9):e15916) doi:10.2196/15916

KEYWORDS

physician-rating websites; data quality issues; doctor ratings; reviews; data quality framework

Introduction

Background

With the proliferation of mobile devices and instantaneous access to data, electronic word of mouth (e-WOM) has become a force to be reckoned with, affecting many aspects of our lives, including the things we buy, the shows we watch, and the places where we stay, directly or indirectly. Such dependence on e-WOM is especially true in the context of choosing a physician, as consumers historically have relied on word of mouth, including personal recommendations [1]. A simple check on Google Trends showed that the phrase "doctors near me" is now searched almost nine times more than it was 5 years ago; therefore, it is not surprising to see a rise in the number and

scope of physician-rating websites (PRWs), which are peer-to-peer information-sharing platforms that patients use to share reviews and ratings of their health care providers. National survey data indicated that one in six Americans consult online ratings [2]. More than 30% of consumers compare physicians online before choosing a provider [3]. Emphasizing the impact of PRWs, one study [4] noted that 35% of patients selected physicians based on good ratings, while 37% avoided physicians with bad ratings. Another study found that patients consult PRWs as their first step in choosing providers [5] and that 80% of users trust online physician ratings as much as personal recommendations. Millennials, who account for more than half of PRW consumers, were found to exhibit a different behavior when they were unhappy with their health care services [6]. People aged 65 years or above were more likely to complain to doctors directly, while people aged 18 to 24 years were more likely to tell their friends. This emphasizes the evolution of PRWs into a platform for open and honest communication.

Although PRWs are less popular compared with rating websites in other domains, such as fast-moving consumer goods and e-commerce, they have high potential for growth. However, PRWs historically have lagged behind user expectations [7,8], and one of the contributing factors is end users' lack of confidence in the data quality of PRWs. Furthermore, such data quality issues assume high importance, as poor data quality could affect consumers' care choices adversely. Previous research has discussed individual issues in specific contexts [9-57]; however, there is a need for a study that presents a holistic perspective by investigating a comprehensive set of data quality issues found in PRWs. This study fills that literature gap by gleaning data quality issues from several previous studies and classifying them based on the data quality framework.

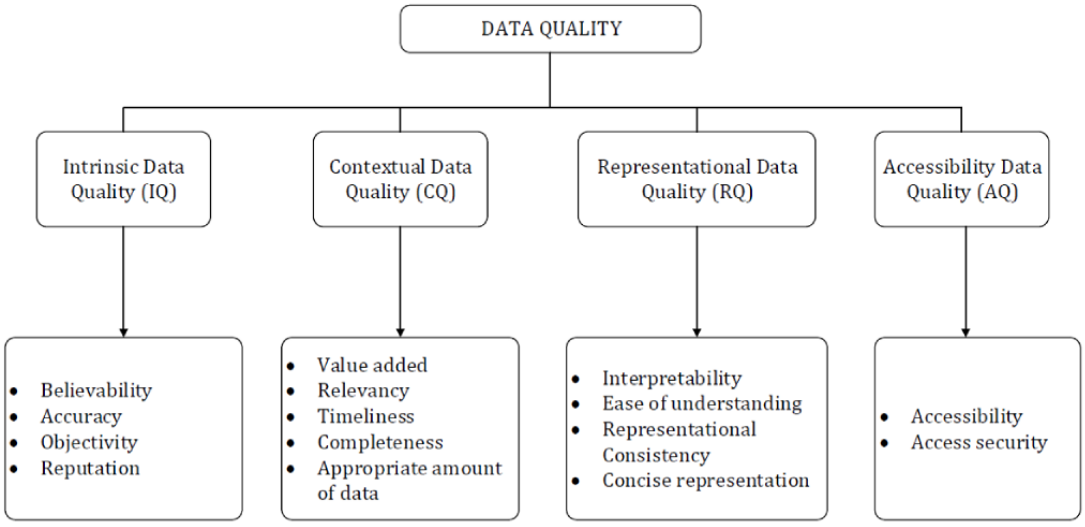
Data Quality Framework

We used a data quality framework developed by Wang and Strong [58] to classify data quality issues in PRWs. It was created by considering consumers' perspectives on data quality, which accommodates a broader definition of data quality. Consequently, Wang and Strong defined *data quality* as "data that are fit for use by data consumers." Furthermore, they empirically collected data quality attributes from consumers instead of determining these attributes theoretically or basing them on expert opinions to identify attributes that emerge from real consumers. They then used two-stage surveys and a two-phase sorting study to develop a hierarchical framework. They captured various data quality attributes, consolidated these attributes into dimensions, and distributed the dimensions across the following categories (Figure 1): intrinsic data quality (IQ), contextual data quality (CQ), representational data quality (RQ), and accessibility data quality (AQ).

IQ entails dimensions that are inherent to the nature of data, including accuracy, objectivity, reputation, and believability. While information system professionals typically have interpreted IQ to mean accuracy alone, consumers assess IQ broadly by considering other elements, such as reputation, objectivity, and believability of the source. CQ is a measure of data quality within the context of the task at hand. It includes dimensions, such as relevance, value addition, timeliness, and completeness, which are specific to a given situation. Together, RQ and AQ emphasize the role of systems that store the data. RQ underscores the importance of developing interfaces that concisely present data so that they are easy to understand and interpret. AQ focuses on making systems secure to ensure that data are safe and available only to relevant users. This framework also makes pragmatic sense, as consumers' view of "fit for use data" would include data that are accurate, objective, believable, obtained from a reputable source, relevant to a specific task at hand, easy to understand, and accessible to them.

We use this framework because, unlike other frameworks [59], it accommodates a much broader definition of data quality. In addition, researchers have used it to evaluate the data quality of several customer-centric products, such as online bookstores and auction sites [60]. It has also been used to study critical factors affecting consumer-purchase behaviors in shopping contexts [61]. In recent years, as e-WOM has gained considerable prominence, some studies have used this framework to examine the impact of data quality on e-WOM [62]. Previous studies have used it to categorize data quality issues with electronic health records [63]. Furthermore, it has been used to study information quality issues on websites in which users, not experts, generate content [64]. Such an end-user point of view is especially relevant to our research, as most shortlisted studies discussed data quality issues in PRWs from patients' perspective.

Figure 1. Data quality framework adapted from Wang and Strong [58].



Methods

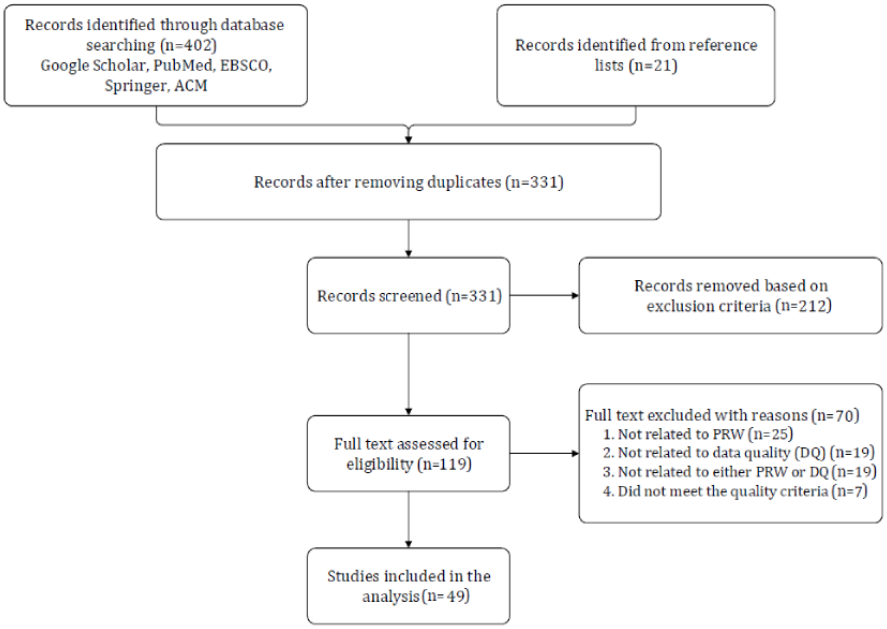
Overview

The study aimed to collect, analyze, and discuss data quality issues in PRWs based on the data quality framework of Wang and Strong. To accomplish this goal, we developed the following research questions:

- RQ1: What data quality issues exist in PRWs?
- RQ2: How are these data quality issues classified according to the Wang and Strong framework?
- RQ3: Which data quality issues have been addressed, and which ones warrant attention from researchers and practitioners?

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [65,66], we performed a systematic literature review (Figure 2).

Figure 2. Literature search following the PRISMA guidelines. PRW: physician-rating website.



Search Strategy

We systematically searched for literature published in the past 20 years (between January 1, 2000, and January 1, 2020) using the following databases: ACM Digital Library, EBSCO, Springer, PubMed, and Google Scholar. The searches were performed using the following search terms: (“physician” OR “doctor” OR “provider”) AND (“review” OR “rating”) AND (“online” OR “internet”) AND (“data” OR “quality”). Initially, title, abstract, and index terms were used to screen for published journal articles, conference papers, proceedings, case studies, and book chapters. Two reviewers performed the screening independently. The reviewers met on a regular basis to discuss the inclusion of studies. A third reviewer was consulted when there was disagreement between the reviewers. Furthermore, the reviewers performed hierarchical searches by identifying literature sources through references cited in the shortlisted papers selected from the keyword searches to find additional relevant articles.

Inclusion and Exclusion Criteria

Physician quality is an elusive concept to measure, as it means different things to different stakeholders. Health care professionals and policymakers have developed a plethora of clinical and process-quality measures to address the challenge of evaluating physician quality. Some well-known examples of such quality indicators include the risk-adjusted mortality rate [67], 30-day readmission rate [68], and percentage of patients receiving recommended preventive care. Although such clinical measures of quality are critical to improving the quality of care, they emphasize the process of care, not individual physicians' quality. Furthermore, these measures are neither easy to access nor simple to understand. They also do not place high emphasis on patients' perceptions of care quality. Owing to this lack of agreed-upon, meaningful, and readily available objective data on individual clinician performance, patients end up relying on other patients for recommendations.

PRWs fill this void by providing a platform on which patients can evaluate physicians based on their experiences; however, these ratings and reviews are individual patients' subjective opinions and may not be indicative of physicians' clinical quality. Thus, it is possible that a poorly rated physician has provided the correct or best treatment. Some studies argue that patients are not well-suited to evaluate physician quality because of the information asymmetry between patients and care providers [69]. In addition, several studies emphasize that ratings and reviews of patients are not correlated to clinical measures of physician quality [70,71]. Despite these shortcomings, PRWs have surged in popularity and have become instrumental in shaping prospective patients' opinions. Individuals also use them to make crucial decisions, such as selecting a provider, because just as in any other consumer service business, customers' perceptions impact revenue.

Therefore, in this paper, we focused on physician quality from patients' perspectives, as captured by PRWs. Prior studies that examined data quality of patient-reported reviews and ratings were included in this literature review. Studies were excluded from this review if they (1) focused primarily on clinical quality measures that health care providers or public health agencies reported; (2) examined data quality issues that are not related to public PRWs (eg, papers that catered to paid websites, such as Castle Connolly, were excluded from this review); (3) were not available as full text in the final search; (4) were not written in English; and (5) were white papers, reports, abstracts only, letters, or commentaries.

Data Extraction, Synthesis, and Evaluation

A Google document was created for data extraction. For each chosen study, the data collected included the title, author, year, country, abstract, study type, and number of participants. We assessed the selected studies' quality based on the criteria listed in Table 1. The quality criteria were developed using guidelines specified by *the Cochrane Handbook for Systematic Reviews* [72] and the report "Guidelines for performing systematic literature reviews in software engineering" [73]. Two reviewers independently assessed every included study by assigning "Yes," "No," or "Cannot tell" scores to each criterion. Only studies that received a "Yes" on all criteria were included in this review. A senior researcher was consulted for a resolution if there was disagreement between the reviewers.

Table 1. Quality criteria for the included studies.

Identifier	Issue
C1	Does the article clearly show the purpose of the research?
C2	Does the article adequately provide the literature review, background, or context?
C3	Does the article present the related work with regard to the main contribution?
C4	Does the article have a clear description of the research methodology?
C5	Does the article include research results?
C6	Does the article present a conclusion related to the research objectives?
C7	Does the article recommend future research directions or improvements?

Results

Characteristics of Reviewed Studies

We included 49 papers published between 2009 and 2019. Among these, 28 articles were identified through the initial search and additional 21 articles meeting the study criteria were identified by reviewing the reference lists of those articles. The list of included papers is presented in Multimedia Appendix 1. We identified 18 unique data quality issues that afflict PRWs and classified these issues into four categories based on the data quality framework. In recent years, PRWs have captured the research community's attention as evidenced by a surge in publications in the past 5 years. More than 71% (35/49) of the papers used quantitative methods to test their research hypotheses, 22% (11/49) adopted qualitative methods, and 6% (3/49) leveraged mixed methods. Healthgrades (n=12), RateMDs (n=8), Vitals (n=7), and Jameda (n=4) were the most targeted PRWs, while other sites, such as Zocdoc, Press Ganey, and Healthcare Reviews, were not examined as much. Three studies compared PRWs with business directory and review sites such as Yelp. In addition, several studies collected data from multiple PRWs to compare their results across different rating websites. Altogether, 26 articles focused on issues relating to IQ, while 30 discussed CQ concerns. Four articles emphasized RQ errors, and 13 were related to AQ challenges. Around half (26/49) the included studies focused on more than one type of data quality issue.

Discussion

Principal Findings

Consistent with this study's goals, we discuss different data quality issues in a narrative format based on the four types of issues specified in the data quality framework.

IQ Issues

As presented in Table 2, we discuss intrinsic data quality issues based on the following dimensions: accuracy, objectivity, believability, and reputation. The main hurdle affecting the accuracy of reviews and rating data was the glaring absence of negative ratings. A prior study highlighted the absence of negative ratings by empirically showing that physicians with low patient-perceived quality were less likely to be rated. Although a positive correlation between online ratings and physician quality was found, the association was the strongest for the medium segment. While the ratings were not sensitive for high-quality physicians, there were fewer ratings for physicians at the lower end of the quality distribution [9].

Another study showed the presence of ubiquitous high ratings for interventional radiologists, with mean ratings ranging from 4.3 to 4.5 on a five-point scale [10]. This lack of negative ratings was not only limited to interventional radiologists, but also spanned other medical specialties. Several studies corroborated this finding by noting that the average physician rating across all specialties was consistently very positive [11-16]. Sparse negative ratings could also be attributed to legal restrictions on entering negative feedback in some countries, such as Switzerland. However, this dearth of negative ratings adversely affects overall opinions about data quality greatly.

Table 2. Intrinsic data quality issues.

Issues	Dimension	Citations
Ratings were either positive or extremely positive, with a notable absence of negative ratings.	Accuracy and objectivity	[9-16]
A significant number of ratings contained extreme values, typically in the form of a dichotomous distribution of the minimum and maximum values.	Objectivity	[9,17-19]
A significant number of reviews contained emotionally charged comments, implying a lack of objectivity in the reviews.	Objectivity	[20-22]
Online ratings were less sensitive to physician quality at the high end of quality distribution, implying the presence of the halo effect.	Objectivity	[9,23,24]
Some physician-rating websites did not ensure ratings' accuracy by allowing anonymous ratings that were not entirely believable.	Believability	[18,20,25,26]
Some sites allowed premium-paying physicians to hide up to three negative comments.	Believability and reputation	[18,27,28]
Physicians were more likely to trust patient-experience surveys that health systems issued, whereas patients were more likely to trust ratings found on independent websites.	The data source's reputation	[29]

Several research studies [9,17-19] questioned the objectivity of review and rating data by uncovering the high presence of extreme ratings, such as one- or five-star ratings. Extreme ratings do not represent a balanced view and are usually an impulsive response to an emotional trigger. Other studies [20-22] further corroborated these findings by revealing the presence of emotionally charged review comments. On one hand, physicians with low perceived quality were hardly rated; thus, a single negative review had a disproportionate impact on the overall rating. On the other hand, physicians with several reviews incurred no relevant impact from a negative rating on overall numbers owing to relatively few negative ratings. Two studies also highlighted the presence of the halo effect [23,24]. One found that higher ratings were associated with marketing strategies that the physicians employed. It also discovered that physicians' online presence greatly impacted their ratings. This phenomenon demonstrates the susceptibility of reviews and ratings to external factors, such as marketing and promotion, casting serious doubt on the credibility of the data.

Other studies [18,20,25] examined data quality challenges that emerge when users can rate physicians anonymously. Anonymity exposes information to manipulation from sources such as competition, slanderers, and biased friends. The effect of anonymous ratings cascades into a relevant issue when the number of genuine negative ratings is small, as is the case with PRWs. An excellent example of such abuse of PRWs is found in how antiabortion groups deliberately target physicians working in abortion clinics with libelous comments and negative ratings under the veneer of anonymity [26].

Several business models of PRWs also skew the believability of ratings and review information by providing physicians with premium subscriptions having an option to hide up to three negative comments [27,28]. The hidden negative reviews may mislead consumers who are usually unaware of the business models of PRWs. Several researchers also questioned the ethics of hiding up to three negative ratings when, on average, there were less than three negative physician ratings [18].

RQ Issues

As presented in Table 3, RQ emphasizes clear representation and includes dimensions such as interpretability, ease of understanding, representational consistency, and conciseness. Of all data quality issues, the ones related to RQ are the most insidious, as even an accurate data set can lead to misleading conclusions if representation is not appropriate. One study observed how users can be influenced by the mere repositioning of positive reviews to the first few pages, as most users read initial reviews more than subsequent ones. The same study

examined the negative impact of placing poor ratings at the top of the profile page [25]. Furthermore, researchers have argued that the five-point scale that PRWs use is an imperfect proxy for physician quality [30,31] as the difference between a rating of 4.8 and 4.9 might be too small to be meaningful from an end user's perspective. These RQ issues seem to be more deliberate than the other data issues and result from the business or revenue model of PRWs. Thus, they may be more challenging to overcome.

Table 3. Representational data quality issues.

Issues	Dimension	Citations
The five-point scale used for measuring physician quality did not have the finer granularity needed to highlight the minor differences in physician quality.	Interpretability	[30,31]
The positioning of positive reviews and rating data on the first few pages greatly impacted patient perceptions.	Representational consistency	[25]
Every physician-rating website used different underlying scales to measure the effectiveness of the physicians. Therefore, interpreting results across different physicians can be difficult.	Interpretability	[32]

CQ Issues

As presented in Table 4, CQ examines issues in the context of the task at hand. For this paper, the task is assumed to be patient decision making in terms of choosing a provider by analyzing ratings and reviews. Several data quality issues plagued this dimension. The most fundamental concern stemmed from the need for an appropriate amount of data to make meaningful decisions. One critical issue in this segment was the nonexistence of ratings and reviews for most of the physicians. Several papers [22,23,33-37] found that more than half of the physicians had no ratings or reviews. They also argued that no meaningful decisions could be made with the unavailability of data for such a considerable volume of physicians. Another related threat to CQ was the low volume of reviews and ratings. One study [23] noted that 57% of the doctors received only one to three ratings. Such low volumes cast doubts on the ratings' credibility, especially when sites allow anonymous ratings. Previous research showed that users did not trust the rating and review data, and sought information from alternative sources until a minimum number of ratings was available. Thus, it should come as no surprise that multiple prior studies on e-WOM in other domains, such as e-commerce, found that high volumes lend more credence to rating and review data. Furthermore, data analysis showed that early negative reviews beget more negative reviews [25]. At the same time, doctors with great initial reviews might continue to benefit from these reviews, even if their clinical quality has declined over the years.

A nuanced challenge to CQ emerged from the underlying factors used to compute ratings and reviews. One study [51] compared the factors between two sites, one from the United States and another from Germany. They found that German PRWs focused on parameters that measure physician characteristics, while American sites focused on the entire clinical process, including registration, clinical pathways, and staff behaviors. Typically, most PRWs include wait times, staff behaviors, follow-ups, and ease of making appointments, some of which are not under physicians' direct control. In addition, these factors may not be truly representative of physician quality. Some studies suggested taking reviews and ratings with a "grain of salt," as the ratings reflected patients' perceptions and did not objectively measure physician quality.

Furthermore, patients might not be able to assess a wide range of physician attributes owing to information asymmetry between physicians and patients. Some studies have discussed the possibilities of bringing both patient perspectives and clinical quality measures together to enhance CQ [38]. The ease of decision-making from a user's perspective defines the essence of CQ. Such a user perspective was affected adversely when a low degree of correlation existed among different physician review websites [44,48-51], as users may not know which websites to trust.

Table 4. Contextual data quality issues.

Issues	Dimension	Citations
There was a low volume of reviews and ratings, with more than half the physicians having less than one to three ratings.	Appropriate amount of data	[22,23,33-37]
Physician-rating websites captured patient perceptions of physician quality; they did not capture and present objective measures of quality, such as Physician Quality Reporting System (PQRS) ratings for physicians or risk-adjusted mortality rate.	Objectivity completeness	[29,31,38,39]
Positive ratings were based on factors, such as ease of getting an appointment, short wait times, and staff behaviors, that did not directly represent physician characteristics.	Relevance	[21-23,29,41-46]
Higher ratings were associated with marketing strategies that physicians employed, such as significant online presence and promotion of satisfied patients' reviews.	Objectivity relevance	[40,43,47]
There was a low degree of correlation among online websites on surgeon ratings.	Value addition	[44,48-51]

AQ Issues

As presented in Table 5, AQ focuses on the dimensions ease of access and security of access. While PRWs are afflicted by only a limited set of accessibility challenges, some issues warrant further discussion. First, while the internet may be accessible universally, we must consider socioeconomic and psychographic barriers to the accessibility of PRWs [53]. Typically, tech-savvy people with reasonable income and education use PRWs. Several studies noted that PRWs did not represent the opinions of elderly people, who comprise the largest consumer segment for health care services. Second, the number of ratings and reviews that physicians received depended on their specialty. One study noted that physicians in specialties that warranted high interaction with patients, such as obstetrics and gynecology, were more likely to be rated, while other specialties, such as pathology, were less likely to be rated [54].

Several studies found that the maturity level of PRWs was not uniform across countries. While PRWs have been adopted widely in the United States, the United Kingdom, and Canada, they were at an early stage of adoption in other countries, such as Australia [34], Switzerland [55], Lithuania [56], and Germany [57]. Data quality issues with PRWs in these countries were more pronounced when compared with other nations. Furthermore, accessibility issues emerged owing to legal and regulatory challenges. For instance, Switzerland provides physicians with a legal option to have negative reviews deleted.

Table 5. Accessibility data quality issues.

Issues	Dimension	Citations
The frequency and volume of ratings varied greatly based on physician specialty; therefore, some specialists' ratings might not have been easily accessible.	Ease of accessibility	[21,52]
Even though the internet was widely accessible, financial and social access barriers had to be considered. Such barriers include income, culture, gender, and age. The effective use of physician-rating websites remained primarily dependent on users' cognitive and intellectual capabilities.	Accessibility	[37,51,53,54]
The maturity of physician-rating websites was inconsistent across countries. Physician-rating websites were in the early stage of adoption, with very few ratings in many countries, such as Lithuania and Australia.	Appropriate amount of data completeness	[34,49,53-57]

Opportunities for Future Research

In this review, the key observation was the lack of emphasis on RQ and AQ issues in prior research. Most studies highlighted IQ and CQ issues, which are foundational to achieving other types of data quality. Although the number of papers published on RQ [30-32] and AQ [55-57] issues has increased in recent years, more research is warranted on these issues.

Specifically, the misleading impact of inline advertisements and the effect from framing reviews differently warrant further investigation. Furthermore, while the application of data analytic methods has surged considerably, there is a conspicuous absence of studies that have examined the impact of different kinds of visualizations of rating and review data on patient decision-making. Similarly, machine learning and predictive analytic methods could well be used to forecast future physician ratings. Such studies could educate physicians further on strategies to improve their future ratings. Another observation is the need to examine the impact of physician reviews and ratings on providers' revenue or insurance payments. Understanding this impact can help explain physicians' behaviors and motivations.

Relatively few studies have discussed the use of data analytics, machine learning, and other statistical methods to identify and distinguish fake reviews from genuine ones. The development of such mechanisms can substantially enhance the quality of review and rating data. Further, data privacy laws, such as General Data Protection Regulation (GDPR), can potentially impact the quality of review and rating data adversely. More research is warranted to examine the impact of these regulations on accessibility data quality.

Limitations

This systematic review has several limitations. The first is associated with the methodology used to select studies. Relatively few (n=49) studies were included, based on the selection criteria that the authors set. Five databases were searched to select studies. Other databases potentially could have yielded additional studies. Furthermore, while the authors performed due diligence in execution, the search was limited to the set of selected keywords. It is possible that some relevant studies were not identified owing to keyword mismatches or title differences and, therefore, were not included in the review.

Second, we only included articles written in the English language, and business models can vary greatly among geographical locations, depending on existing practices, cultures, and regulations. We tried to find literature from diverse geographical areas, but few studies were accessible to us owing to limitations caused by language barriers and the extent of research in other geographical areas. Furthermore, we did not include non–peer-reviewed sources, such as white papers and dissertations. It is possible that relevant information from such sources could have influenced our findings.

Conclusion

This paper contributes to a better understanding of data quality issues in PRWs by highlighting several vital challenges that these issues pose. The paper acknowledges the tremendous potential that PRWs have in transforming health care by being the voice of consumers and increasing the transparency of health care processes. However, this study showed that data quality challenges present relevant hurdles to the realization of these benefits. The impact of these data quality issues will only surge as millennials base their decisions on PRW data.

Historically, IQ and CQ factors have been principal sources of data quality issues, and many researchers have studied them extensively. However, RQ and AQ factors warrant more research. In particular, RQ factors, such as the impact of inline advertisements and the positioning of positive reviews on the first few pages, are usually deliberate and result from the business or revenue model of PRWs. In addition, data privacy regulations, such as GDPR in the European Union and California Consumer Privacy Act (CCPA) in California, may greatly impact PRWs. More research is needed to understand their implications. Furthermore, the effect of cultural factors (eg, in some cultures, speaking negatively about authority figures is viewed as inappropriate), though relevant, was not considered, as it is under-addressed in the literature. Future innovations and research are needed to address these emerging data quality issues. We hope that this study's results inspire professionals and researchers to develop PRWs that are more robust and do not have many data quality issues.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of all included studies.
[XLSX File (Microsoft Excel File), 33 KB - jmir_v22i9e15916_app1.xlsx]

References

1. Tu H, Lauer J. Word of mouth and physician referrals still drive health care provider choice. Res Brief 2008 Dec(9):1-8. [Medline: 19054900]
2. Fox S. Vagal afferent controls of feeding: a possible role for gastrointestinal BDNF. Clin Auton Res 2013 Feb;23(1):15-31. [doi: 10.1007/s10286-012-0170-x] [Medline: 22717678]
3. Fox S, Purcell K. Chronic Disease and the Internet. Pew Research Center. 2010. URL: https://www.pewresearch.org/internet/2010/03/24/chronic-disease-and-the-internet/ [accessed 2020-09-21]
4. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
5. Findlay SD. Consumers' Interest In Provider Ratings Grows, And Improved Report Cards And Other Steps Could Accelerate Their Use. Health Aff (Millwood) 2016 Apr;35(4):688-696. [doi: 10.1377/hlthaff.2015.1654] [Medline: 27044970]
6. Ellimoottil C, Leichtle S, Wright C, Fakhro A, Arrington A, Chirichella T, et al. Online physician reviews: the good, the bad and the ugly. Bull Am Coll Surg 2013 Sep;98(9):34-39. [Medline: 24455818]
7. Bidmon S, Elshiewy O, Terlutter R, Boztug Y. What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data. J Med Internet Res 2020 Feb 03;22(2):e13830 [FREE Full text] [doi: 10.2196/13830] [Medline: 32012063]
8. Murphy GP, Radadia KD, Breyer BN. Online physician reviews: is there a place for them? RMHP 2019 May;Volume 12:85-89. [doi: 10.2147/rmhp.s170381]
9. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
10. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
11. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
12. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
13. Black E, Thompson L, Saliba H, Dawson K, Black NM. An analysis of healthcare providers' online ratings. Inform Prim Care 2009 Dec 01;17(4):249-253 [FREE Full text] [doi: 10.14236/jhi.v17i4.744] [Medline: 20359403]
14. Orhurhu MS, Chu R, Claus L, Roberts J, Salisu B, Urits I, et al. Neuropathic Pain and Sickle Cell Disease: a Review of Pharmacologic Management. Curr Pain Headache Rep 2020 Jul 24;24(9):52-86. [doi: 10.1007/s11916-020-00885-5] [Medline: 32705357]
15. Kirkpatrick W, Abboudi J, Kim N, Medina J, Maltenfort M, Seigerman D, et al. An Assessment of Online Reviews of Hand Surgeons. Arch Bone Jt Surg 2017 May;5(3):139-144 [FREE Full text] [Medline: 28656160]
16. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun 4;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
17. Pike CW, Zillioux J, Rapp D. Online Ratings of Urologists: Comprehensive Analysis. J Med Internet Res 2019 Jul 02;21(7):e12436 [FREE Full text] [doi: 10.2196/12436] [Medline: 31267982]

18. Samora JB, Lifchez SD, Blazar PE, American Society for Surgery of the Hand EthicsProfessionalism Committee. Physician-Rating Web Sites: Ethical Implications. J Hand Surg Am 2016 Jan;41(1):104-10.e1. [doi: 10.1016/j.jhsa.2015.05.034] [Medline: 26304734]
19. Mostaghimi A, Crotty BH, Landon BE. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156 [FREE Full text] [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]
20. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
21. Carbonell G, Brand M. Choosing a Physician on Social Media: Comments and Ratings of Users are More Important than the Qualification of a Physician. International Journal of Human–Computer Interaction 2017 Aug 03;34(2):117-128. [doi: 10.1080/10447318.2017.1330803]
22. Kizawa Y. Leveraging online patient reviews to improve quality of care: art-specific insights for practice administrators and physicians. Fertility and Sterility 2017 Sep;108(3):e112. [doi: 10.1016/j.fertnstert.2017.07.343]
23. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: International comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
24. Obele CC, Duszak R, Hawkins CM, Rosenkrantz AB. What Patients Think About Their Interventional Radiologists: Assessment Using a Leading Physician Ratings Website. J Am Coll Radiol 2017 May;14(5):609-614. [doi: 10.1016/j.jacr.2016.10.013] [Medline: 28017529]
25. Li S, Feng B, Chen M, Bell RA. Physician review websites: effects of the proportion and position of negative reviews on readers' willingness to choose the doctor. J Health Commun 2015 Apr;20(4):453-461. [doi: 10.1080/10810730.2014.977467] [Medline: 25749406]
26. Davis MM, Hanauer DA. Online Measures of Physician Performance: Moving Beyond Aggregation to Integration. JAMA 2017 Jun 13;317(22):2325-2326. [doi: 10.1001/jama.2017.3979] [Medline: 28609521]
27. Merrell JG, Levy BH, Johnson DA. Patient assessments and online ratings of quality care: a "wake-up call" for providers. Am J Gastroenterol 2013 Nov;108(11):1676-1685. [doi: 10.1038/ajg.2013.112] [Medline: 24192941]
28. Gordon HS. Consider Embracing the Reviews from Physician Rating Websites. J Gen Intern Med 2017 Jun;32(6):599-600 [FREE Full text] [doi: 10.1007/s11606-017-4036-8] [Medline: 28299603]
29. Haskins IN, Krpata DM, Rosen MJ, Perez AJ, Tastaldi L, Butler RS, et al. Online Surgeon Ratings and Outcomes in Hernia Surgery: An Americas Hernia Society Quality Collaborative Analysis. J Am Coll Surg 2017 Nov;225(5):582-589. [doi: 10.1016/j.jamcollsurg.2017.08.007] [Medline: 28838868]
30. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-Based Physician Ratings for California Physicians on Probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
31. Meszmer N, Jaegers L, Schöffski O, Emmert M. [Do online ratings reflect structural differences in healthcare? The example of German physician-rating websites]. Z Evid Fortbild Qual Gesundhwes 2018 Apr;131-132:73-80. [doi: 10.1016/j.zefq.2017.11.007] [Medline: 29331281]
32. Rothenfluh F, Schulz PJ. Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis. J Med Internet Res 2018 Jun 14;20(6):e212 [FREE Full text] [doi: 10.2196/jmir.9105] [Medline: 29903704]
33. Carbonell G, Meshi D, Brand M. The Use of Recommendations on Physician Rating Websites: The Number of Raters Makes the Difference When Adjusting Decisions. Health Commun 2019 Nov;34(13):1653-1662. [doi: 10.1080/10410236.2018.1517636] [Medline: 30222006]
34. Atkinson S. Current status of online rating of Australian doctors. Aust. J. Prim. Health 2014;20(3):222. [doi: 10.1071/py14056]
35. McLennan S, Strech D, Reimann S. Developments in the Frequency of Ratings and Evaluation Tendencies: A Review of German Physician Rating Websites. J Med Internet Res 2017 Aug 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
36. McLennan S, Strech D, Kahrass H. Why are so few patients rating their physicians on German physician rating websites? A qualitative study. BMC Health Serv Res 2018 Aug 29;18(1):670 [FREE Full text] [doi: 10.1186/s12913-018-3492-0] [Medline: 30157842]
37. Gilbert K, Hawkins CM, Hughes DR, Patel K, Gogia N, Sekhar A, et al. Physician rating websites: do radiologists have an online presence? J Am Coll Radiol 2015 Aug;12(8):867-871. [doi: 10.1016/j.jacr.2015.03.039] [Medline: 26250980]
38. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
39. Gray BM, Vandergrift JL, Gao G, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015 Feb;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
40. Trehan SK, DeFrancesco CJ, Nguyen JT, Charalel RA, Daluiski A. Online Patient Ratings of Hand Surgeons. J Hand Surg Am 2016 Jan;41(1):98-103. [doi: 10.1016/j.jhsa.2015.10.006] [Medline: 26710742]

41. Donnally CJ, Roth ES, Li DJ, Maguire JA, McCormick JR, Barker GP, et al. Analysis of Internet Review Site Comments for Spine Surgeons. SPINE 2018;43(24):1725-1730. [doi: 10.1097/brs.0000000000002740]
42. Saifee D, Bardhan I, Zheng Z. Do Online Reviews of Physicians Reflect Healthcare Outcomes? In: Chen H, Zeng D, Karahanna E, Bardhan I, editors. Smart Health. ICSH 2017. Lecture Notes in Computer Science. Cham: Springer; Jun 26, 2017:161-168.
43. Liu X, Guo X, Wu H, Vogel D. Doctor's Effort Influence on Online Reputation and Popularity. In: Zheng X, Zeng D, Chen H, Zhang Y, Xing C, Neill DB, editors. Smart Health. ICSH 2014. Lecture Notes in Computer Science. Cham: Springer; 2014:111-126.
44. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a Sports Medicine Surgeon's "Quality" in the Modern Era: an Analysis of Popular Physician Online Rating Websites. HSS J 2016 Oct;12(3):272-277 [FREE Full text] [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
45. Li S, Lee-Won RJ, McKnight J. Effects of Online Physician Reviews and Physician Gender on Perceptions of Physician Skills and Primary Care Physician (PCP) Selection. Health Commun 2019 Oct;34(11):1250-1258. [doi: 10.1080/10410236.2018.1475192] [Medline: 29792519]
46. Haug M, Gewald H. Are Friendly and Competent the Same?: The Role of the Doctor-Patient Relationship in Physician Ratings. In: SIGMIS-CPR'18: Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research. 2018 Jun Presented at: 2018 ACM SIGMIS Conference on Computers and People Research; June 18, 2018; Buffalo, New York p. 157. [doi: 10.1145/3209626.3209734]
47. Prabhu A, Randhawa S, Clump D, Heron D, Beriwal S. What Do Patients Think About Their Radiation Oncologists? An Assessment of Online Patient Reviews on Healthgrades. Cureus 2018 Feb 06;10(2):e2165 [FREE Full text] [doi: 10.7759/cureus.2165] [Medline: 29644154]
48. Calixto NE, Chiao W, Durr ML, Jiang N. Factors Impacting Online Ratings for Otolaryngologists. Ann Otol Rhinol Laryngol 2018 Aug;127(8):521-526. [doi: 10.1177/0003489418778062] [Medline: 29882425]
49. Patel S, Cain R, Neailey K, Hooberman L. Public Awareness, Usage, and Predictors for the Use of Doctor Rating Websites: Cross-Sectional Study in England. J Med Internet Res 2018 Jul 25;20(7):e243 [FREE Full text] [doi: 10.2196/jmir.9523] [Medline: 30045831]
50. McGrath RJ, Priestley JL, Zhou Y, Culligan PJ. The Validity of Online Patient Ratings of Physicians: Analysis of Physician Peer Reviews and Patient Ratings. Interact J Med Res 2018 Apr 09;7(1):e8 [FREE Full text] [doi: 10.2196/ijmr.9350] [Medline: 29631992]
51. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
52. Quigley DD, Elliott MN, Farley DO, Burkhart Q, Skootsky SA, Hays RD. Specialties differ in which aspects of doctor communication predict overall physician ratings. J Gen Intern Med 2014 Mar;29(3):447-454 [FREE Full text] [doi: 10.1007/s11606-013-2663-2] [Medline: 24163151]
53. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011 Dec 06;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
54. Liu JJ, Matelski JJ, Bell CM. Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study. J Med Internet Res 2018 Mar 07;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
55. McLennan S. Quantitative Ratings and Narrative Comments on Swiss Physician Rating Websites: Frequency Analysis. J Med Internet Res 2019 Jul 26;21(7):e13816 [FREE Full text] [doi: 10.2196/13816] [Medline: 31350838]
56. Bäumer F, Kersting J, Kur?elis V, Geierhos M. Rate Your Physician: Findings from a Lithuanian Physician Rating Website. In: Damaševičius R, Vasiljevienė G, editors. Information and Software Technologies. ICIST 2018. Communications in Computer and Information Science. Cham: Springer; Oct 2018:43-58.
57. McLennan S, Strech D, Meyer A, Kahrass H. Public Awareness and Use of German Physician Ratings Websites: Cross-Sectional Survey of Four North German Cities. J Med Internet Res 2017 Nov 09;19(11):e387 [FREE Full text] [doi: 10.2196/jmir.7581] [Medline: 29122739]
58. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 2015 Dec 11;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
59. Knight S, Burn J. Developing a Framework for Assessing Information Quality on the World Wide Web. Informing Sci J 2005;8:159-172. [doi: 10.28945/493]
60. Barnes S, Vidgen R. An integrative approach to the assessment of e-commerce quality. J Electron Commerce Res 2002;3(3):114-127.
61. Park C, Kim Y. Identifying key factors affecting consumer purchase behavior in an online shopping context. 2003 Jan;31(1):16-29. [doi: 10.1108/09590550310457818]
62. Filieri R, Alguezaui S, McLeay F. Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism Management 2015 Dec;51:174-185. [doi: 10.1016/j.tourman.2015.05.007]

63. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care 2012 Jul;50 Suppl:S21-S29 [FREE Full text] [doi: 10.1097/MLR.0b013e318257dd67] [Medline: 22692254]
64. Yaari E, Baruchson-Arbib S, Bar-Ilan J. Information quality assessment of community generated content: A user study of Wikipedia. Journal of Information Science 2011 Aug 15;37(5):487-498. [doi: 10.1177/0165551511416065]
65. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
66. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
67. Shine D. Risk-adjusted mortality: problems and possibilities. Comput Math Methods Med 2012;2012:829465 [FREE Full text] [doi: 10.1155/2012/829465] [Medline: 22474540]
68. Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, et al. An Administrative Claims Measure Suitable for Profiling Hospital Performance Based on 30-Day All-Cause Readmission Rates Among Patients With Acute Myocardial Infarction. Circ Cardiovasc Qual Outcomes 2011 Mar;4(2):243-252. [doi: 10.1161/circoutcomes.110.957498]
69. Daskivich T, Houman J, Fuller G, Black J, Kim H, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2018 Apr 01;25(4):401-407. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
70. Trehan SK, Nguyen JT, Marx R, Cross MB, Pan TJ, Daluiski A, et al. Online Patient Ratings Are Not Correlated with Total Knee Replacement Surgeon-Specific Outcomes. HSS J 2018 Jul;14(2):177-180 [FREE Full text] [doi: 10.1007/s11420-017-9600-6] [Medline: 29983660]
71. Emmert M, Adelhardt T, Sander U, Wambach V, Lindenthal J. A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites. BMC Health Serv Res 2015 Sep 24;15:414 [FREE Full text] [doi: 10.1186/s12913-015-1051-5] [Medline: 26404452]
72. Savović J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014 Apr 15;3:37 [FREE Full text] [doi: 10.1186/2046-4053-3-37] [Medline: 24731537]
73. Kitchenham B, Charters S. Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report. 2007. URL: https://userpages.uni-koblenz.de/~laemmel/esecourse/slides/slr.pdf [accessed 2020-09-19]

Abbreviations

AQ: accessibility data quality
CQ: contextual data quality
e-WOM: electronic word of mouth
GDPR: General Data Protection Regulation
IQ: intrinsic data quality
PRW: physician-rating website
RQ: representational data quality

Edited by T Kool; submitted 19.08.19; peer-reviewed by S Bidmon, R Terlutter, B Ranard; comments to author 08.10.19; revised version received 24.04.20; accepted 13.08.20; published 28.09.20.

Please cite as:
Mulgund P, Sharman R, Anand P, Shekhar S, Karadi P
Data Quality Issues With Physician-Rating Websites: Systematic Review
J Med Internet Res 2020;22(9):e15916
URL: http://www.jmir.org/2020/9/e15916/
doi:10.2196/15916
PMID:32986000

Original Paper

Clusters of Adolescent Physical Activity Tracker Patterns and Their Associations With Physical Activity Behaviors in Finland and Ireland: Cross-Sectional Study

Kwok Ng[1,2,3], PhD; Sami Kokko[4], PhD; Tuija Tammelin[5], PhD; Jouni Kallio[5], PhD; Sarahjane Belton[6], PhD; Wesley O'Brien[7], PhD; Marie Murphy[8], PhD; Cormac Powell[3,9], PhD; Catherine Woods[2,3], PhD

[1]School of Educational Sciences and Psychology, University of Eastern Finland, Joensuu, Finland
[2]Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland
[3]Physical Activity for Health Cluster, Health Research Institute, University of Limerick, Limerick, Ireland
[4]Faculty of Sport and Health Sciences, University of Jyvaskyla, Jyväskylä, Finland
[5]LIKES Research Centre for Physical Activity and Health, Jyväskylä, Finland
[6]School of Health and Human Performance, Faculty of Science and Health, Dublin City University, Dublin, Ireland
[7]School of Education, University College Cork, Cork, Ireland
[8]School of Sport, Ulster University, Belfast, United Kingdom
[9]Performance Department, Swim Ireland, Irish Sport HQ, Dublin, Ireland

Corresponding Author:
Kwok Ng, PhD
School of Educational Sciences and Psychology
University of Eastern Finland
PO Box 111
Joensuu, 80101
Finland
Phone: 358 504724051
Email: kwok.ng@hbsc.org

Abstract

Background: Physical activity trackers (PATs) such as apps and wearable devices (eg, sports watches, heart rate monitors) are increasingly being used by young adolescents. Despite the potential of PATs to help monitor and improve moderate-to-vigorous physical activity (MVPA) behaviors, there is a lack of research that confirms an association between PAT ownership or use and physical activity behaviors at the population level.

Objective: The purpose of this study was to examine the ownership and use of PATs in youth and their associations with physical activity behaviors, including daily MVPA, sports club membership, and active travel, in 2 nationally representative samples of young adolescent males and females in Finland and Ireland.

Methods: Comparable data were gathered in the 2018 Finnish School-aged Physical Activity (F-SPA 2018, n=3311) and the 2018 Irish Children's Sport Participation and Physical Activity (CSPPA 2018, n=4797) studies. A cluster analysis was performed to obtain the patterns of PAT ownership and usage by adolescents (age, 11-15 years). Four similar clusters were identified across Finnish and Irish adolescents: (1) no PATs, (2) PAT owners, (3) app users, and (4) wearable device users. Adjusted binary logistic regression analyses were used to evaluate how PAT clusters were associated with physical activity behaviors, including daily MVPA, membership of sports clubs, and active travel, after stratification by gender.

Results: The proportion of app ownership among Finnish adolescents (2038/3311, 61.6%) was almost double that of their Irish counterparts (1738/4797, 36.2%). Despite these differences, the clustering patterns of PATs were similar between the 2 countries. App users were more likely to take part in daily MVPA (males, odds ratio [OR] 1.27, 95% CI 1.04-1.55; females, OR 1.49, 95% CI 1.20-1.85) and be members of sports clubs (males, OR 1.37, 95% CI 1.15-1.62; females, OR 1.25, 95% CI 1.07-1.50) compared to the no PATs cluster, after adjusting for country, age, family affluence, and disabilities. These associations, after the same adjustments, were even stronger for wearable device users to participate in daily MVPA (males, OR 1.83, 95% CI 1.49-2.23; females, OR 2.25, 95% CI 1.80-2.82) and be members of sports clubs (males, OR 1.88, 95% CI 1.55-2.88; females, OR 2.07,

95% CI 1.71-2.52). Significant associations were observed between male users of wearable devices and taking part in active travel behavior (OR 1.39, 95% CI 1.04-1.86).

Conclusions: Although Finnish adolescents report more ownership of PATs than Irish adolescents, the patterns of use and ownership remain similar among the cohorts. The findings of our study show that physical activity behaviors were positively associated with wearable device users and app users. These findings were similar between males and females. Given the cross-sectional nature of this data, the relationship between using apps or wearable devices and enhancing physical activity behaviors requires further investigation.

(J Med Internet Res 2020;22(9):e18509) doi:10.2196/18509

KEYWORDS

wearables; children; activity trackers; active travel; organised sport; self-quantification

Introduction

Background

Physical inactivity is one of the leading causes of worldwide mortality. There is an urgent need to understand how to increase physical activity levels among young adolescents (typically aged between 11 years and 15 years). The habits developed during early adolescence will continue through adulthood [1], particularly for physical activity, in both the short-term [2,3] and long-term [4,5]. Although technologies have evolved in the last generation, less is known about the how the use of physical activity trackers (PATs) in young adolescents can be used to understand their physical activity behaviors. Lee et al [6] conducted a scoping review and included 14 studies that included intervention components such as websites, apps, and wearables devices. Apps and wearable devices include the interaction of sensors that have the capability to measure physical activity, and for the purpose of this study, they are referred throughout as PATs. Small to medium effect sizes in adult studies have been demonstrated, specifically when individuals used the PAT information to modify their behaviors and increase their physical activity levels [7]; however, intervention studies among adolescents are rare [6]. The lesser number of studies on adolescents may be attributed to the way apps and wearable devices have been designed—they are primarily designed with the adult population in mind, thereby leading to low levels of use among adolescents [8]. Even though there have been recent developments in companies to build PATs for youngsters, the feedback from children through their parents still indicate that design issues are present, which can cause a barrier for sustainable use [9]. Understanding the association between PAT ownership and use may help inform their use as an effective intervention tool in the young adolescent population.

Country-Specific Usage

The use of apps requires the use of smartphones and consistent internet connectivity. In Ireland, the prevalence of mobile phone use among 13-year-old adolescents has been reported to be 98% [10], with 54% reporting to use the internet with their phones [11]. In Finland, 93% of the adolescents aged 16 years use their mobile phones to access the internet and this has been facilitated by the way mobile phone subscription plans in Finland typically offer unlimited data [12]. According to Eurostat, 94% of the homes in Finland have access to the internet, whereas, in Ireland, internet household accessibility stands at 89%, which is equivalent to the average in European Union countries [13].

In Finland, wearable devices are used in 22% of the households [12], but the use of wearable devices in Ireland has not been reported. Younger generations often report greater use of the internet than the average population [14], and this information is frequently used to ensure that PAT functionality is optimal for age-appropriate usage.

Both countries perform highly in terms of progression in making societies mediated by digital technology. For examples, out of all the OECD (Organization for Economic Cooperation and Development) countries, Finns use the most amount of mobile data per subscriber (OECD, 2018), and Finland has been ranked the highest for digital services in all of Europe [15]. Ireland, on the other hand, is the second ranked country in Europe, outside of the Nordic countries, in the digital economy and society index [13]. One of the major differences between Finland and Ireland is the human capital. Ireland is ranked number 1 in terms of open data usage; however, it is placed 21st in the European Union for eHealth services, with just 1 in 10 people (11%) using it. However, in Finland, open data usage tops the eHealth services index, with 49% of the individuals using it [15].

PATs for Facilitation of Physical Activity Behaviors

Emerging evidence suggests that PATs have a positive effect on physical activity behaviors, particularly as facilitators, rather than as drivers of health behavior change [16]. From the end user's perspective, one significant limitation of commercial PAT products is the short life cycle [17,18]. It has been estimated that 10% of the global population use fitness apps, but only 2% are paying users [19]. These figures have increased in the last few years and are expected to increase by 4.5% every year at least until 2024 [19]. Despite the commercial growth in the market, the expected level of success of PAT ownership in terms of physical activity behavior change is yet to be fully interrogated or proven. Most of the functions within apps include between 5 and 8 behavioral change techniques, with common features including goals and planning, feedback and monitoring, social support and shaping behavior [20]. Several trials have investigated the efficacy of PATs in increasing physical activity levels, and often, although an initial increase in the physical activity level is documented, these changes are not sustained in the long term [18,21,22] or could not be replicated with younger populations [23]. High attrition levels are particularly common for PATs, where the novelty effect wears off and so does the

usage and effectiveness [24]. Designers of PATs could be partly responsible for this attrition, as the products made may not be sufficiently accurate enough to meet the demands of the end user [25,26] or the automated systems for providing reminders and feedback are deemed inadequate [27]. The majority of PAT features include functions that can be used for socializing between users, which is important for young adolescent populations [9]. Moreover, young adolescents see the benefits of multipurpose devices as this can help individuals track various aspects of their lives such as health [28]. Despite this, young adolescents have reported that PATs currently do not provide enough customization or personalization [8].

Differences have been observed in the way that males and females use the multiple functions of PATs. For example, male adolescents prefer to socialize with their PATs through "banter" and other friendly conversations [29]. Female PAT users report lower levels of aspirations to take part in competitive sports than male PAT users [30], which may translate into less use of the goal and planning functions [31]—the areas commonly used for optimal performance in sports [32]. Feedback from PATs is also central to the control theory [33], as it confirms user performance to prompt further behaviors. For the purpose of this study, we focused on the association between PAT ownership and use and 3 behaviors, namely, (1) overall physical activity behavior, (2) participation in sports clubs, and (3) active commuting. Research on PATs as an intervention for adolescents is limited [8], and to our knowledge, only a handful of observational studies have been published on PAT usage in young adolescents [34]. Therefore, the purposes of this study were to investigate the differences in PAT ownership and usage between Finnish and Irish adolescents, while investigating the association between PAT ownership and self-reported physical activity behaviors in both cohorts and considering the potential influence of gender.

Methods

Recruitment

Data for this cross-sectional study were collected in Finland and Ireland during the first half of 2018. Both the Finnish and Irish data were collected from national representative cross-sectional studies. In Finland, the Finnish School-aged Physical Activity (F-SPA) study [35] is the national physical activity monitoring study for children and adolescents (LIITU in Finnish) and the Irish equivalent is the Children's Sport Participation and Physical Activity (CSPPA) study [36].

The F-SPA 2018 was based on 2-level cluster analyses [35] and is an extension to the F-SPA 2014 and F-SPA 2016 studies by including students as young as 7 years of age. A parallel study was conducted for special education classes and schools; hence, sampling did not include special education schools. The probability proportion size was used to calculate the primary sampling unit to generate a nationally representative sample. This study consisted of a total of 311 schools (Finnish schools: 267/311, 85.8%; Swedish schools: 44/311, 14.1%) in Finland and 9940 students altogether.

The CSPPA 2018 was a follow-up and extension to the original 2010 study [37]. CSPPA 2018 included schools from both the Republic of Ireland and Northern Ireland. The sampling frame for the schools involved in CSPPA 2010 included all schools with students aged between 10 years and 18 years in the Republic of Ireland. A systematic one-stage cluster sampling method was used for CSPPA 2010 and schools were stratified by 4 criteria (school gender, school socioeconomic status, school location, and school size). The same schools in 2010 were invited again to take part in 2018, and a replacement list based on the creation of an equivalence sample was made to ensure that sufficient number of students were included to allow for study design effects. Schools from Northern Ireland were not part of CSPPA 2010 and all mainstream schools from Northern Ireland were included in the sampling frame for CSPPA 2018. Special schools, primary schools, and colleges of further education were excluded from the database. In total, in the Republic of Ireland and Northern Ireland, 115 schools were included in the overall study sample and 6651 students took part in CSPPA 2018.

All surveys were completed on either a tablet, laptop, or personal computer in the students' own classroom and under the supervision of teachers in F-SPA 2018 or specifically trained research assistants in CSPPA 2018. Students who were given permissions by their parents or guardians had the right to withdraw from the study at any time. The completion of the survey was done anonymously and voluntarily. The Finnish study was approved by the ethics committee of the University of Jyvaskyla, Finland, where no number was provided, and the Irish study was approved by the ethics committee of the University of Limerick, Ireland.

For the purpose of comparisons between the studies, only responses from young adolescents aged between 11 years and 15 years were included (Finland, n=3311; Ireland, n=4797) in the final data set. Variables for the country data files were relabeled to allow for merging in SPSS 25.0 (IBM Corp). The details of the measures in both surveys used for this study are reported in the table in Multimedia Appendix 1.

Measures

Both surveys collected demographic information on gender, age, disability status, and self-reported socioeconomic status via the Family Affluence Scale (FAS) [38]. The CSPPA 2018 study also included the option of "other" to respond for gender, whereas the F-SPA 2018 did not; therefore, respondents with "other" (69/6649, 1.0%) were removed from the CSPPA 2018 sample prior to cleaning the data file for analyses.

Items of PATs had slight variation (Multimedia Appendix 1). In Finland, a block question was designed to keep the survey length to the minimum. The opening question was, "How often in a week do you use the following physical activity tracking devices?" with the following options, "mobile apps," "activity meter or sports watch," and "heart rate monitor." This item was used in the previous edition of the Finnish monitoring study [34]. Although the response options were updated, based on the feedback from the test-retest reliability study [39], the previous 3 categories of "none," "own but do not use", and "own and use" were extended to a 6-item category frequency scale of (1)

I don't have, (2) Never, (3) At least once a week, (4) Once a day, (5) More than once a day, and (6) All the time. The variables were divided into 3 groups to match with the previous reporting, whereby respondents to 1 were grouped as "do not have," 2 were grouped as "own but do not use," and 3 to 6 were grouped as "own and use."

The Irish version had separate questions on ownership, use, and frequency of use for (1) physical activity apps, (2) smartwatches, (3) heart rate monitors, (4) pedometers, and (5) other devices. For comparison purposes, individuals who responded to only having a pedometer or other device (528/4797; 11.0%) were recoded as not having a PAT since this was not compatible between the 2 studies. There was a slight variation in the frequency of use of the PATs, because the question was, "How often do you use your physical activity tracking device during a typical week" with response options (1) Never, (2) Once, (3) Sometimes, (4) Almost every day, and (5) Every day. Responses of Never were grouped into "own but do not use" and 2-5 were grouped into "own and use." Null responses to the ownership were deemed as "do not have."

The other survey responses used included the self-reported number of days of at least 60 minutes of moderate-to-vigorous physical activity (MVPA) participation in both 2018 F-SPA and CSPPA studies. The CSPPA 2018 study included 2 items based on the past 7 days and the usual week. The 2 items were summed and divided by 2 and rounded up, whereas the F-SPA 2018 item included 1 item based on the past 7 days. Previous studies suggest that an average between the previous week and the usual week can provide a more accurate recall of physical activity behaviors [40]; however, in other international surveys, a single item was used to reduce the number of items in the survey [41]. Both surveys used the same explanation to define MVPA, as comparable to an international study protocol [42]. The physical activity survey item was then dichotomized into meeting the guidelines (specifically when the respondents reported a total of 7 days) and not meeting the guidelines (specifically when the respondents reported anything between 0 and 6 days). This physical activity survey item has been previously tested for validity use against accelerometers with young adolescents [40,43] and within test-retest environments [44,45].

Respondents provided details of their mode of transport to school with walking or cycling categorized as "active commuters." Motorized transport included options such as getting a lift by parents or taking the bus. The distance between the primary home and school was also asked. To ensure that the distances were plausible for active transport, the Finnish legislation for the provision of free transport costs were set at distances over 5 km as the cut-off point to differentiate between people who were close (within 5 km) and far (over 5 km). For the inferential statistics regarding active commuting, only respondents who lived within the close range (5 km) of the school were included. Therefore, living beyond 5 km was an exclusion criterion for the analyses in relation to active commuting.

Statistical Analysis

Descriptive statistics of the population characteristics were produced by chi-square tests of independence for gender, after stratifying for country. The test of independence between the countries was also tested through chi-square tests after considering the gender. A two-step approach was used to describe the phenomenon of PAT ownership and use among young adolescents in Finland and Ireland. The number of possible combinations of PAT habits was investigated using cluster analysis to the fewest number of clusters, yet attempting to retain a meaningful structure (ie, values of the average silhouette width defining the cluster quality as "good" [exceeding 0.5]) [46]. When the sample was pooled, 3 clusters were deemed to be sufficient; however, when the test of clusters was examined for each country, one of the clusters in each country had different features; therefore, an extra cluster was added. The characteristics of the 4 clusters were then tested in the pooled sample, and individually, each country achieved good cluster quality (Finland silhouette width=0.6, Ireland silhouette width=0.7) and led to 4 clusters being identified. The first cluster (and reference category for regression analyses) included individuals who reported no ownership of apps, smartwatches, or heart rate monitors. This category was labelled as "no PATs." The second cluster predominately included individuals who reported ownership but not usage of PATs and were labelled as "PAT owners." The third cluster was labelled as "app users" as the majority of the app users belonged to this cluster, with none in the cluster reporting the use of smartwatches or heart rate monitors. The fourth cluster included a mixture of individuals consistently using smartwatches and heart rate monitors and they were subsequently labelled as "wearable device users."

Chi-square tests were used to assess the statistical significance of gender, age groups, FAS, and disability for the clusters, and the Kruskal-Wallis test with pairwise comparisons was used to assess the statistical significance of the differences in the average number of days reporting 60 minutes of MVPA for each country.

The binary logistic associations of meeting the physical activity guidelines (7 days vs <7 days, reference category), being an active traveler (cyclist and walker vs motorized transport who live within 5 km of the school, reference category), and organized sport participant (sports club member vs not active in sports clubs, reference category) with no ownership of PATs as the reference category were investigated. The crude associations for each indicator (Model 1) were assessed before adjusting for age, gender, FAS, and disability (Model 2). All statistics were run using SPSS 25.0 for Windows (released 2017).

Results

User Statistics

The descriptive statistics are provided in Table 1 with comparisons between and within countries.

Table 1. Descriptive statistics of the samples by country and gender.

Characteristics	Finland (n=3311)			Ireland (n=4797)			Total[a] (N=8108)		
	Males, n= 1610, n (%)	Females, n=1701, n (%)	*P* value	Males, n=2370, n (%)	Females, n=2427, n (%)	*P* value	Males, n= 3980, n (%)	Females, n= 4128, n (%)	*P* value
Age (years)			.81			.003			.01
11	628 (39.0)	680 (39.9)		402 (16.9)	468 (19.3)		1030 (25.9)	1148 (27.8)	
13	452 (28.1)	477 (28.0)		1090 (45.9)	1168 (48.1)		1542 (38.7)	1645 (39.8)	
15	530 (32.9)	544 (31.9)		878 (37.0)	791 (32.6)		1408 (35.4)	1335 (32.3)	
FAS[b]			.81			.05			.09
Low	357 (24.8)	374 (24.4)		495 (20.9)	529 (21.8)		852 (22.4)	903 (22.8)	
Middle	827 (57.4)	874 (56.9)		1398 (58.9)	1351 (55.7)		2225 (58.4)	2225 (56.2)	
High	256 (17.8)	287 (18.7)		477 (20.1)	546 (22.5)		733 (19.2)	833 (20.2)	
Disability			.07			.66			.17
None	1439 (89.9)	1493 (87.8)		2035 (85.9)	2073 (85.4)		3474 (87.5)	3566 (86.5)	
Disabled	161 (10.06)	204 (11.9)		335 (14.1)	354 (14.6)		496 (12.5)	558 (13.5)	
Daily MVPA[c]			<.001			<.001			<.001
Inactive[d]	1041 (64.8)	1211 (71.3)		1955 (82.5)	2157 (88.9)		2996 (75.3)	3368 (81.6)	
Active[e]	566 (35.2)	488 (28.7)		415 (17.5)	270 (11.1)		981 (24.7)	758 (18.4)	
Transport			.15			<.001			.02
Motorized, Close	213 (13.5)	213 (12.7)		504 (28.7)	583 (33.1)		717 (21.5)	796 (23.1)	
Active, Close	916 (58.1)	1005 (59.8)		430 (24.5)	335 (19)		1346 (40.4)	1340 (39)	
Active, Far	80 (5.07)	60 (3.6)		44 (2.5)	30 (1.7)		124 (3.7)	90 (2.6)	
Motorized, Far	368 (23.3)	403 (24)		779 (44.3)	811 (46.1)		1147 (34.4)	1214 (35.3)	
Sports club			.64			.04			.06
Nonmember	629 (40.5)	687 (41.31)		852 (35.95)	941 (38.77)		1481 (38.8)	1628 (39.8)	
Member	924 (59.5)	976 (58.69)		1518 (64.05)	1486 (61.23)		2442 (61.2)	2462 (60.2)	
Apps			<.001			<.001			<.001
Not owned	624 (38.7)	649 (38.2)		1602 (67.6)	1457 (60)		2226 (55.9)	2106 (51.0)	
Do not use	342 (21.2)	244 (14.3)		274 (11.6)	347 (14.3)		616 (15.5)	591 (14.3)	
Use	644 (40)	808 (47.5)		494 (20.8)	623 (25.7)		1138 (28.6)	1431 (34.7)	
Sports watch			<.001			.56			<.001
Not owned	1080 (68.6)	1240 (74.1)		1836 (77.5)	1903 (78.4)		2916 (73.9)	3143 (76.6)	
Do not use	269 (17.1)	182 (10.9)		169 (7.1)	155 (6.4)		438 (11.0)	337 (8.2)	
Use	225 (14.3)	252 (15.1)		365 (15.4)	369 (15.2)		590 (15.0)	621 (15.1)	
Heart rate monitor			<.001			.19			<.001
Not owned	1119 (71.4)	1331 (79.6)		2146 (90.5)	2230 (91.9)		3265 (82.9)	3561 (86.9)	
Do not use	285 (18.2)	201 (12)		55 (2.3)	55 (2.3)		340 (8.6)	256 (6.2)	
Use	164 (10.5)	141 (8.4)		169 (7.1)	142 (5.9)		333 (8.5)	283 (6.9)	

[a]The percentages in this column are the actual percentages and not of the total population because some data on the variables of the total population were missing.

[b]FAS: Family Affluence Scale.

[c]MVPA: moderate-to-vigorous physical activity.

[d]0-6 days of MVPA.

[e]7 days of MVPA.

There were statistical differences between the characteristics of Finnish and Irish young adolescents between the 2 surveys. The CSPPA 2018 study had fewer 11-year-old adolescents than those in the F-SPA 2018, as the participants were more evenly distributed across the varying age groups in the F-SPA 2018 (chi-square *P*=.75). There were fewer Finnish (n=3311) than Irish (n=4797) respondents in the final sample (Table 1).

The estimates of ownership and usage of apps (*P*<.001), sports watches (*P*<.001), and heart rate monitors (*P*<.001) were significantly different between Finnish and Irish adolescents. In Finland, almost two-thirds (2038/3311, 61.6%) of the young adolescents reported owning or using apps to monitor physical activity, whereas in Ireland, the majority (3059/4797, 63.8%) did not own apps to monitor physical activity. Over three-quarter of the Irish adolescents do not own sports watches and this was a greater estimate than that among Finnish adolescents (2320/3311, 71.4%; *P*<.001). Moreover, a quarter (791/3311, 24.4%) of the Finnish adolescents reported owning or using heart rate monitor, which was much higher than that in Ireland (421/4797, 8.8%; *P*<.001).

Cluster Analyses

The 4 clusters were ”no PATs,” “PAT owners,” “app users,” and “wearable device users” (Table 2), with a silhouette of 0.7 for cluster quality exceeding the good threshold of 0.5 [46]. The crude percentage of the individuals who reported daily MVPA was almost double among wearable device users (498/1631, 30.6%) compared to those in no PATs (576/3523, 16.3%). The estimates of those involved in active transport from cluster 1 (1033/1732, 59.6%), cluster 2 (236/373, 63.3%), cluster 3 (804/1217, 66.1%), and cluster 4 (571/826, 69.1%) and of sports clubs members (cluster 1, 1957/3523, 56.2%; cluster 2, 399/677, 59.6%; cluster 3, 1333/2200, 61.4%; and cluster 4, 1178/1631, 73.1%) were high.

Table 2. Features of the four clusters from pooled data and crude estimates of the behaviors.

Features	Cluster 1 (No PATs), n=3523, n (%)	Cluster 2 (PAT owners), n=677, n (%)	Cluster 3 (app users), n=2200, n (%)	Cluster 4 (wearable device users), n=1631, n (%)
Apps				
None	3523 (100.0)	265 (39.1)	0 (0)	531 (32.6)
Own	0 (0)	412 (60.9)	701 (31.9)	85 (5.2)
Use	0 (0)	0 (0)	1499 (68.1)	1015 (62.2)
Smartwatches				
None	3523 (100.0)	72 (10.6)	2200 (100)	256 (15.7)
Own	0 (0)	605 (89.4)	0 (0)	167 (10.2)
Use	0 (0)	0 (0)	0 (0)	1208 (74.1)
Heart rate monitors				
None	3523 (100.0)	286 (42.2)	2200 (100)	814 (49.9)
Own	0 (0)	391 (57.8)	0 (0)	204 (12.5)
Use	0 (0)	0 (0)	0 (0)	613 (37.6)
Crude percentage				
Daily MVPA[a]	576 (16.3)	130 (19.2)	503 (22.9)	498 (30.6)
Active Transport[b]	1033[c] (59.6)	236[d] (63.3)	804[e] (66.1)	571[f] (69.1)
Sports club	1957 (56.2)	399 (59.6)	1333 (61.4)	1178 (73.1)

[a]MVPA: moderate-to-vigorous physical activity.

[b]Fewer people in this subcategory met the criteria of living within 5 km; therefore, the sample population for this row is different in each cluster as shown in the following footnotes.

[c]n=1732.

[d]n=373.

[e]n=1217.

[f]n=826.

Multivariate Analyses of the Males

In the unadjusted model (Model 1, Table 3), daily MVPA (odds ratio [OR] 1.56, 95% CI 1.29-1.87; OR 2.16, 95% CI 1.79-2.60), active transport (OR 1.41, 95% CI 1.13-1.77; OR 1.83, 95% CI 1.41-2.36), and being members of sports clubs (OR 1.32, 95% CI 1.12-1.56; OR 1.97, 95% CI 1.64-2.36) showed positive associations with male app users and wearable device users, respectively. Moreover, owners of PATs were positively associated with active travel (OR 1.40, 95% CI 1.02-1.91). After controlling for country, age, FAS, and disabilities, the associations had lower ORs and the association for active travel and app users was no longer statistically significant (Model 2, Table 3). Disabilities were also not associated with active travel,

whereas as adolescents became older, the less likely they would be involved in physical activity behaviors. Higher FAS was positively associated with daily MVPA and being a member of sports clubs, whereas it was negatively associated with active travel. There were more Irish adolescents who were members of sports clubs, but significantly fewer who took part in daily MVPA or active travel.

Table 3. Male-adjusted odds ratios and 95% confidence intervals without Model 1 and with Model 2 confounders for each cluster. Italics represents statistically significant associations.

Variables	Moderate-to-vigorous physical activity[a], OR[b] (95% CI)	Active travel[c], OR (95% CI)	Sports club[d], OR (95% CI)
Model 1			
No PATs	Reference (1.0)	Reference (1.0)	Reference (1.0)
PAT owners	1.17 (0.89-1.53)	*1.40 (1.02-1.91)*	1.03 (0.82-1.29)
App users	*1.56 (1.29-1.87)*	*1.41 (1.13-1.77)*	*1.32 (1.12-1.56)*
Wearable device users	*2.16 (1.79-2.60)*	*1.83 (1.41-2.36)*	*1.97 (1.64-2.36)*
Model 2			
No PATs	Reference (1.0)	Reference (1.0)	Reference (1.0)
PAT owners	1.07 (0.81-1.42)	1.19 (0.84-1.69)	1.08 (0.85-1.37)
App users	*1.27 (1.04-1.55)*	1.06 (0.82-1.36)	*1.37 (1.15-1.62)*
Wearable device users	*1.83 (1.49-2.23)*	*1.39 (1.04-1.86)*	*1.88 (1.55-2.28)*
Country			
Finland	Reference (1.0)	Reference (1.0)	Reference (1.0)
Ireland	*0.46 (0.39-0.54)*	*0.26 (0.21-0.32)*	*1.44 (1.25-1.67)*
Age			
Young	Reference (1.0)	Reference (1.0)	Reference (1.0)
Older	*0.69 (0.63-0.77)*	*0.59 (0.50-0.69)*	*0.68 (0.62-0.74)*
Family Affluence Scale			
Lower	Reference (1.0)	Reference (1.0)	Reference (1.0)
Higher	*1.22 (1.08-1.37)*	*0.78 (0.66-0.92)*	*1.39 (1.23-1.55)*
Disability			
Without	Reference (1.0)	Reference (1.0)	Reference (1.0)
With	*0.59 (0.45-0.77)*	1.00 (0.74-1.40)	*0.60 (0.49-0.73)*

[a]Reference=not daily, Nagelkerke R^2 (Model 1)=0.026, Nagelkerke R^2 (Model 2)=0.102.

[b]OR: odds ratio.

[c]Reference=motorized, Nagelkerke R^2 (Model 1)= 0.017, Nagelkerke R^2 (Model 2)=0.217.

[d]Reference=not member, Nagelkerke R^2 (Model 1)=0.020, Nagelkerke R^2 (Model 2)=0.075.

Multivariate Analyses of the Females

In the unadjusted model (Model 3, Table 4), wearable device users were twice as likely to report MVPA (OR 2.45, 95% CI 2.00-3.02) and be member of organized sports (OR 2.29, 95% CI 1.91-2.74). The associations were not as strong for app users, and for active travel, the association was similar between app users (OR 1.24, 95% CI=1.01-1.53) and wearable device users (OR 1.28, 95% CI 1.01-1.63). Owners of PATs were more likely to be members of sports clubs (OR 1.31, 95% CI 1.01-1.68) when compared to the no PATs cluster. After controlling for country, age, FAS, and disabilities (Model 4, Table 4), the ORs were lower, but the model strengths were stronger. There were no significant differences between no ownership or usage of PATs and any cluster of PATs. Females with and without disabilities were not different in terms of MVPA and active travel.

Table 4. Female-adjusted odds ratios and 95% confidence intervals without Model 3 and with Model 4 confounders for each cluster. Italics represents statistically significant associations.

Variables	Moderate-to-vigorous physical activity[a], OR[b] (95% CI)	Active travel[c], OR (95% CI)	Sports club[d], OR (95% CI)
Model 3			
None	Reference (1.0)	Reference (1.0)	Reference (1.0)
Owners	1.21 (0.86-1.72)	0.92 (0.65-1.30)	*1.31 (1.01-1.68)*
App user	*1.61 (1.32-1.96)*	*1.24 (1.01-1.53)*	*1.21(1.04-1.40)*
Wearable device user	*2.45 (2.00-3.02)*	*1.28 (1.01-1.63)*	*(2.29 (1.91-2.74)*
Model 4			
None	Reference (1.0)	Reference (1.0)	Reference (1.0)
Owners	1.27 (0.87-1.84)	1.14 (0.76-1.72)	1.27 (0.97-1.66)
App user	*1.49 (1.20-1.85)*	0.98 (0.76-1.25)	*1.25 (1.07-1.50)*
Wearable device user	*2.25 (1.80-2.82)*	0.91 (0,68-1.21)	*2.07 (1.71-2.52)*
Country			
Finland	Reference (1.0)	Reference (1.0)	Reference (1.0)
Ireland	*0.38 (0.31-0.45)*	*0.16 (0.13-0.20)*	*1.35 (1.17-1.55)*
Age			
Young	Reference (1.0)	Reference (1.0)	Reference (1.0)
Older	*0.55 (0.49-0.62)*	*0.50 (0.43-0.59)*	*0.62 (0.57-0.68)*
Family Affluence Scale			
Lower	Reference (1.0)	Reference (1.0)	Reference (1.0)
Higher	*1.19 (1.04-1.36)*	*0.85 (0.72-0.99)*	*1.57 (1.42-1.75)*
Disability			
Without	Reference (1.0)	Reference (1.0)	Reference (1.0)
With	1.00 (0.76-1.31)	1.19 (0.89-1.58)	*0.59 (0.48-0.71)*

[a]Reference=not daily, Nagelkerke R^2 (Model 3)=0.030, Nagelkerke R^2 (Model 4)=0.141.

[b]OR: odds ratio.

[c]Reference=motorized, Nagelkerke R^2 (Model 3)=0.005, Nagelkerke R^2 (Model 4)=0.320.

[d]Reference=not member, Nagelkerke R^2 (Model 3)=0.028, Nagelkerke R^2 (Model 4)=0.109.

Discussion

Principal Results

Apps were owned by approximately two-thirds of the Finnish adolescents and by one-third of the Irish adolescents, with more females in both countries owning apps than males. The estimates of sports watch ownership or use is 28.6% (928/3311) among young Finns and 22.1% (1058/4797) among young Irish adolescents. Approximately 9.2% (305/3311) of the Finnish adolescents and 6.5% (311/4797) of the Irish adolescents use heart rate monitors. Despite these differences, the clustering patterns of PATs were similar between both the countries.

Four cluster patterns for PATs were identified: (1) no PATs, (2) PAT owners, (3) app users, and (4) wearable device users. Compared to individuals in the no PATs cluster, wearable device users had stronger association with physical activity behaviors (daily MVPA, sports club member, active travel). The likelihood of taking part in daily MVPA, being a member of a sports club, or travelling to school by foot or bike among females was higher than that in males, thereby indicating strong positive associations between PAT usage and physical activity behaviors.

More males than females reported meeting the physical activity guidelines of daily MVPA for at least 60 minutes per day [47]. Moreover, approximately twice as many males and two-and-a-half times as many females in Finland reported meeting the daily MVPA guidelines (ie, ≥60 minutes) compared to Irish males and females. Finland tends to perform better than Ireland in studies that report international comparisons of MVPA [48], thereby indicating that the results of this study aligns with other international level findings. Over two and three times as many Finnish males and females, respectively, take part in active travel to school compared to Irish adolescents. The differences in the membership of sports clubs were not statistically significant between the adolescents in the 2 countries. According to the results of the 2016 Global Matrix on physical activity, there were similar differences in the grades between the Finland and Ireland [49].

The majority of the reports from the adult surveys suggest that there are similarities in the use of PATs between Finland and Ireland [50,51]. However, in this study, there were significantly more Finnish adolescents who owned or used PATs compared with their Irish counterparts. In 2016, approximately 53% of the Finnish adolescents aged between 11 years and 15 years reported to own apps [34]; however, this study (2018 data) shows that the rate of app ownership increased to approximately 62%, which is a 9% increase in a 2-year period. There was also a notable increase in the ownership of sports watches and heart rate monitors between 2016 and 2018, and this has followed the rate of market penetration in the last 2 years [19]. This is the first time that PAT figures from Irish adolescents have been reported, and data suggest that adolescents reported a higher rate of ownership of apps (36.2%) when compared to an Irish 18-24-year-old cohort (20%) [50]. Based on these growth figures, it may be viable to design interventions using physical activity apps in Finland currently and in the near future, whereas in Ireland, more adoption is needed for natural experiments or trials to take place; otherwise, interventions would need to control the novelty effect of introducing a new wearable device [24].

Comparison With Prior Work

PATs and Daily MVPA

Both male and female users of apps and wearable devices had positive associations with daily MVPA compared to adolescents with no PATs. However, the physical activity behaviors of young adolescents who merely own but reported to not use PATs were not statistically different from that of individuals in the no PATs cluster. Some of the underlying reasons for these results can be related to the ownership and usage of PATs as a proxy for readiness for the behavior [52]. The functions available in PATs can enable regular monitoring [53], self-comparison of previous performances, and setting targets for current performances [9], and these are all deemed to be effective behavioral change techniques, as reported in previous studies [20]. As such, the feedback from PATs to the individual can be of educational benefit to the user [54]. The details of the habits in using PATs need to be explored further to understand the mechanisms of these associations, particularly as individuals introduced to PATs in interventions experience attrition [17], thus limiting the sustainability of behavior change.

Use of PATs in Organized Sport

Similar to the users' associations with MVPA, app users and wearable device users were more likely to report memberships in sports clubs, when compared to individuals with no PATs. Depending on the features and functionality of the specific PATs that individuals use, young adolescents can share data with other members of the sports club. This may increase motivation among males, as males are known to boast about their achievements with their peers [29]. Moreover, PATs could be used to support coaching practices by providing individualized information on athletes' performance. Data have been used, wherein feedback on the physiological parameters such as heart rate can be informative to athletes and could reduce overtraining and thus the risk to injuries [55]. This is a promising area of development within wearable technology, wherein safety promotion and injury prevention features are built in. Seshadri and colleagues [56] further suggest that noninvasive sensors around the body such as earrings, headphones, rings, or articles made within textiles can act as a crucial pieces of technology to reduce the risk of injuries. Gabbett [57] argued that training harder can be protective of injuries if done smartly and if assisted through trackable data. This may be an important message for the general adolescent population, where health promoters aim to increase the levels of physical activity through careful planning to build up physical fitness. Not only do individuals have feedback on their own behavior, as postulated by the control theory, but the ideas and the information for taking part in more physical activities can be supported by the environment of sports clubs—typically the youth sports coach.

Active Travel and PATs

After adjusting for country, age, family affluence, and disabilities, the only significant association observed was between the male users of wearable devices and active travel. Although there are studies that suggest that PATs can help support more walking [21,27], previous studies have been based primarily on adults and the active travel behaviors of young adolescents are known to be heavily influenced by their parents [58]. Furthermore, distances between home and school that were considered as too far were over 4 km (2.5 miles) in Ireland [59] and 5 km or more away from the school in Finland, which resulted in a large reduction in the number of active commuters among young Finnish adolescents [60]. None of the adolescents in this study could legally use their own independent motorized transport by the age of 15 years, and yet, the active transport behavior was extremely different between Finland and Ireland. These national differences have been previously reported in the Global Matrix 2.0 physical activity report card, where Finland was graded "B" and Ireland as "D" [49].

Research on active travel is limited in terms of PATs; however, there have been some initiatives to promote active travel directly or indirectly through gamification [61,62]. These programs may start off well as the excitement of gamification kicks in, but later, the novelty can wear off, thereby reducing the potential to have sustained active travel [24]. Attrition may be avoided if designers of PATs use an established framework for functionalities in apps [63] and follow the principles around design and usage, as outlined by Attig and Franke [17].

Other initiatives for promoting active school travel and physical activity in general in schools may be created by using step challenges [54]. For young adolescents, such activities need to be considered with care. There could be negative effects [64] because some students have reported that they feel such challenges are impossible to win, given the head start others have on them if they started early in the morning. Alternatively, young adolescents have the feeling of guilt for not keeping up with the pace of their friends, as Goodyear and colleagues termed as "peer surveillance" [54].

Other innovative ways to increase active transport require the combination of technology with the Internet of Things, relying upon multiple sensors such as gyroscopes, GPS, and connectivity sensors so that students can interact more with each other [65]. In a previous study, children took part in a

design session and identified that a light backpack and track commuting with their friends provided more opportunities to socialize [66]; this is an example of going beyond the concept of PATs for the purpose of tracking physical activity, but these can be used as a tool to engage with peers. The concept of wearables on bags is not a new idea, as the concepts have been considered already in the early part of the millennium [67], but it seems that commercial companies have been slow to convert this into the market.

Theoretical Perspectives

Despite the differences in the levels of ownership and usage of PATs, this study found similarities in the clusters between Finnish and Irish adolescents. One of the limitations of the cluster analysis is the data-driven approach, which may lack representativeness outside of the population studied [68]. However, it is a recognized approach to investigate hierarchies and commonalities among groups [69]. It is likely that the difference between the 4 cluster groups is a combination of readiness for behavioral change [70] as well as a personal investment for self-quantification purposes [33,71]. The majority of the apps are free and can perform many of the same tasks, as what wearables can offer in terms of measuring the minimum level of physical activity for health, although it should be noted that the majority of the currently available PATs have been designed with the adult user in mind. Even with the latest models designed for children, the functions of PATs could be better improved to meet the needs of the young users [9]. Central to the sustainable use of PATs is the way in which feedback is given to the user. According to the control theory [33], a feedback loop is introduced between the motives and the behavior. As more wearable devices become available, this can form a part of the identity of an individual or be worn as a fashion item such as jewelry [65], which may be more appealing to females. Stronger beliefs can be seen through stronger commitment to a behavior [72]. Therefore, further research may be needed in the areas of clustering PAT ownership and usage with the role to maintain physical activity behaviors.

Covariates of Associations

In both Finland and Ireland, there is a clear association between affluence and frequency in taking part in organized sports [11,73], as demonstrated in the fully adjusted model 2 and 4 for males and females, respectively. Moreover, there was a decline in physical activity behaviors among the older adolescent cohort. In addition, disabilities were negatively associated with sports club membership and MVPA participation for males only. Similar findings have been reported in female populations across 15 European countries [74], and this could be related to the already existing low levels of physical activity participation among females. Nonetheless, several studies have been conducted on PATs to improve the lives of people with impairments [75-78]. Given these study findings, female user-friendly PATs may be a potentially worthwhile future area of research.

Strengths and Limitations

The data in this study were collected through self-report surveys, and reporting bias from this type of measurement tool is a common limitation in cross-sectional survey-based studies. The data in this study were collected from national representative samples, and such inconsistencies would be typically eradicated by using larger representative samples. Although we attempted to harmonize our data as much as possible, not all items were the same, specifically when translated into the English language. However, the cultural translation, rather than the literal translation, was used in the study to make comparisons possible. This process was carried out by a researcher (KN) with competences in both languages and cultures. Other study limitations are that some residual confounders may be more relevant in one country when compared to the other and therefore were not comparable although stratification by gender and controlling for country, age, family affluence, and disability were included in the adjusted models. Finally, the survey and data collection only gave the options for the respondents to report 3 main types of PATs, and as the market continues to grow, the researchers may have missed some information related to the behaviors from other types of PATs, and the time during which the individual has owned the PATs. The results of this study were cross-sectional, and the length of time that the individuals have been using PATs has not been reported. Increased understanding about the PAT use of young adolescents is needed to not only consider it as a useful tool for promoting physical activity during the adolescent years but also to use it as a part of the daily life at a later stage in adulthood.

Conclusions

The growing pervasiveness of PAT use across both Finland and Ireland is evident in our study, with similar clustering properties. The association between PAT usage and MVPA provides very useful information for both researchers and practitioners. Evidence from this study highlights the positive physical activity behaviors in adolescents who regularly use and wear PATs, particularly with regards to males. The emergence, pervasiveness, and reducing cost of wearable PATs presents opportunities for researchers to incorporate these into interventions to promote physical activity among young adolescents. Moreover, the application of evidence emerging from physical activity behavior change studies could inform the design and function of future PATs. National efforts in Finland and Ireland should consider using effective dissemination strategies seeking to increase the prevalence of youth gaining access to these wearable devices, while of course acknowledging the feasibility and cost constraints in existence. Advances in technology coupled with reductions in the cost of PATs offer researchers a more viable opportunity to target adolescent-specific physical activity interventions to increase the number of individuals meeting the physical activity guidelines.

Acknowledgments

We would like to acknowledge the F-SPA and CSPPA study research teams involved in the data collection. Data collection of F-SPA 2018 was funded by the Ministry of Education and Culture (grant number: OKM/23/626/2018) involving researchers from University of Jyvaskyla, LIKES research center for physical activity, and the UKK institute. The CSPPA 2018 study was funded by Department of Transport Tourism and Sport, Sport Ireland; Department of Health, Healthy Ireland; and Sport Northern Ireland involving researchers from the University of Limerick, University College Cork, Dublin City University, and Ulster University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the measures in both surveys.
[DOCX File , 14 KB - jmir_v22i9e18509_app1.docx]

References

1. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. The Lancet 2016 Jun;387(10036):2423-2478. [doi: 10.1016/s0140-6736(16)00579-1]
2. Hardie Murphy M, Rowe D, Woods C. Sports Participation in Youth as a Predictor of Physical Activity: A 5-Year Longitudinal Study. J Phys Act Health 2016 Jul;13(7):704-711. [doi: 10.1123/jpah.2015-0526] [Medline: 26800567]
3. Hardie Murphy M, Rowe DA, Woods CB. Impact of physical activity domains on subsequent physical activity in youth: a 5-year longitudinal study. J Sports Sci 2017 Feb;35(3):262-268. [doi: 10.1080/02640414.2016.1161219] [Medline: 27067829]
4. Telama R, Yang X, Leskinen E, Kankaanpaa A, Hirvensalo M, Tammelin T, et al. Tracking of Physical Activity from Early Childhood through Youth into Adulthood. Medicine and Science in Sports and Exercise 2014;46(5):955-962. [doi: 10.1249/mss.0000000000000181]
5. Hayes G, Dowd KP, MacDonncha C, Donnelly AE. Tracking of Physical Activity and Sedentary Behavior From Adolescence to Young Adulthood: A Systematic Literature Review. J Adolesc Health 2019 Oct;65(4):446-454. [doi: 10.1016/j.jadohealth.2019.03.013] [Medline: 31248803]
6. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and Effectiveness of Mobile Health Technologies for Facilitating Physical Activity in Adolescents: Scoping Review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847. [doi: 10.2196/11847]
7. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
8. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
9. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental Perspectives of a Wearable Activity Tracker for Children Younger Than 13 Years: Acceptability and Usability Study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
10. Dempsey S, Lyons S, McCoy S. Later is better: mobile phone ownership and child academic development, evidence from a longitudinal study. Economics of Innovation and New Technology 2018 Dec 20;28(8):798-815. [doi: 10.1080/10438599.2018.1559786]
11. Williams J, Thorton M, Morgan M, Quail A, Smyth E, Murphy D, et al. The Lives of 13 Year Olds: Growing up in Ireland, National Longitudinal Study of Children in Ireland. Dublin: Minister for Children and Youth Affairs; 2018. URL: https://www.growingup.ie/pubs/Lives-of-13-Year-olds-report.pdf [accessed 2019-08-09]
12. Statistics F. Science, Technology and Information Society. January 2020. URL: http://www.stat.fi/tup/suoluk/suoluk_tiede_en.html [accessed 2020-08-12]
13. European C. Digital economy and society statistics: households and individuals. June. 2019. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals [accessed 2020-08-12]
14. European C. Being young in Europe today: Digital World. December. 2017. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Being_young_in_Europe_today_-_digital_world [accessed 2020-08-12]
15. The Digital Economy and Society Index (DESI). European Commission. 2019. URL: https://ec.europa.eu/digital-single-market/en/desi [accessed 2020-08-12]

16. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 03;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
17. Attig C, Franke T. Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition. Computers in Human Behavior 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
18. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can Smartphone Apps Increase Physical Activity? Systematic Review and Meta-Analysis. J Med Internet Res 2019 Mar 19;21(3):e12053. [doi: 10.2196/12053] [Medline: 30888321]
19. Statista. eServices Report 2020 - Fitness. Statista Digital Market outlook - Segment Report. 2019. URL: https://www.statista.com/study/36674/fitness-report/ [accessed 2020-08-12]
20. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of Publicly Available Physical Activity Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
21. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
22. Vandelanotte C, Kolt GS, Caperchione CM, Savage TN, Rosenkranz RR, Maeder AJ, et al. Effectiveness of a Web 2.0 Intervention to Increase Physical Activity in Real-World Settings: Randomized Ecological Trial. J Med Internet Res 2017 Nov 13;19(11):e390 [FREE Full text] [doi: 10.2196/jmir.8484] [Medline: 29133282]
23. Cheatham S, Stull K, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
24. Shin G, Feng Y, Jarrahi M, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019 Apr;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]
25. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2017 May 08;52(13):844-850. [doi: 10.1136/bjsports-2016-096990]
26. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
27. Attig C, Franke T. I track, therefore I walk – Exploring the motivational costs of wearing activity trackers in actual users. International Journal of Human-Computer Studies 2019 Jul;127:211-224. [doi: 10.1016/j.ijhcs.2018.04.007]
28. Goodyear VA, Armour KM, Wood H. Young people learning about health: the role of apps and wearable devices. Learning, Media and Technology 2018 Nov 07;44(2):193-210. [doi: 10.1080/17439884.2019.1539011]
29. Goodyear V, Quennerstedt M. #Gymlad - young boys learning processes and health-related social media. Qualitative Research in Sport, Exercise and Health 2019 Oct 07;12(1):18-33. [doi: 10.1080/2159676x.2019.1673470]
30. Ng K, Ryba T. The Quantified Athlete: Associations of Wearables for High School Athletes. In: Advances in Human-Computer Interaction. The quantified athlete: Associations of wearables for high school athletes. Advances in Human Computer Interaction 2018; Oct 01, 2018:1-8.
31. Baretta D, Bondaronek P, Direito A, Steca P. Implementation of the goal-setting components in popular physical activity apps: Review and content analysis. In: Digit Health. Implementation of the goal-setting components in popular physical activity apps: Review and content analysis. Digital Health 2019 01/01; 2019:2055207619862706.
32. Weinberg R. Making Goals Effective: A Primer for Coaches. Journal of Sport Psychology in Action 2010 Oct 11;1(2):57-65. [doi: 10.1080/21520704.2010.513411]
33. Carver CS, Scheier MF. Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. Psychological Bulletin 1982;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111]
34. Ng K, Tynjälä J, Kokko S. Ownership and Use of Commercial Physical Activity Trackers Among Finnish Adolescents: Cross-Sectional Study. JMIR Mhealth Uhealth 2017 May 04;5(5):e61 [FREE Full text] [doi: 10.2196/mhealth.6940] [Medline: 28473304]
35. Kokko S, Martin L, Villberg J. Itsearvioitu liikunta-aktiivisuus, ruutuaika ja sosiaalinen media sekä liikkumisen seurantalaiteet ja -sovellukset (Self-reported physical activity, screen-time, social media physical activity tracking devices). In: Kokko S, editor. Lasten ja nuorten liikuntakäyttäytyminen Suomessa:LIITU-tutkimuksen tuloksia 2018 (Finnish children and adolescents physical activity behaviours: Results of the F-SPA study 2018). Helsinki: Valtion Liikuntaneuvosto Finnish Sports Council 2019; 2019:1-26.
36. Woods C, Powell C, Saunders J. The children's sport participation and physical activity study 2018 (CSPPA 2018). Limerick, Ireland; Dublin, Ireland; Belfast, Northern Ireland: Department of Physical Education and Sport Sciences, University of Limerick; Sport Ireland and Healthy Ireland; Sport Northern Ireland; 2018. URL: https://www.sportireland.ie/sites/default/files/2019-10/csppa-2018-final-report_1.pdf [accessed 2020-05-08]
37. Woods C, Tannehill D, Quinlan A, et al. The children's sport participation and physical activity study (CSPPA). Dublin, Ireland: Irish Sports Council; 2010.

38. Makransky G, Schnohr CW, Torsheim T, Currie C. Equating the HBSC Family Affluence Scale across survey years: a method to account for item parameter drift using the Rasch model. Qual Life Res 2014 Dec;23(10):2899-2907. [doi: 10.1007/s11136-014-0728-2] [Medline: 24902938]
39. Ng KW, Badura P, Dzielska A, Kokko S, Woods C, Hamrik Z. Test-retest reliability of survey items on ownership and use of physical activity trackers. Acta Gymnica 2019 Jun 30;49(2):67-74. [doi: 10.5507/ag.2019.001]
40. Prochaska JJ, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. Arch Pediatr Adolesc Med 2001 May;155(5):554-559. [doi: 10.1001/archpedi.155.5.554] [Medline: 11343497]
41. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. The Lancet Child & Adolescent Health 2020 Jan;4(1):23-35. [doi: 10.1016/s2352-4642(19)30323-2]
42. Biddle SJ, Gorely T, Pearson N, Bull FC. An assessment of self-reported physical activity instruments in young people for population surveillance: Project ALPHA. Int J Behav Nutr Phys Act 2011 Jan 02;8:1 [FREE Full text] [doi: 10.1186/1479-5868-8-1] [Medline: 21194492]
43. Hardie Murphy M, Rowe DA, Belton S, Woods CB. Validity of a two-item physical activity questionnaire for assessing attainment of physical activity guidelines in youth. BMC Public Health 2015 Oct 23;15:1080 [FREE Full text] [doi: 10.1186/s12889-015-2418-6] [Medline: 26498827]
44. Ng K, Hämylä R, Tynjälä J, Villberg J, Tammelin T, Kannas L, et al. Test-retest reliability of adolescents' self-reported physical activity item in two consecutive surveys. Arch Public Health 2019;77:9 [FREE Full text] [doi: 10.1186/s13690-019-0335-3] [Medline: 30891238]
45. Liu Y, Wang M, Tynjälä J, Lv Y, Villberg J, Zhang Z, et al. Test-retest reliability of selected items of Health Behaviour in School-aged Children (HBSC) survey questionnaire in Beijing, China. BMC Med Res Methodol 2010 Aug 10;10:73 [FREE Full text] [doi: 10.1186/1471-2288-10-73] [Medline: 20696078]
46. Kauffman J, Landrum T. Characteristics of emotional and behavioral disorders of children and youth. 9th edition. Boston: Pearson; 2009.
47. WHO. Global recommendations on physical activity for health. WHO. 2010. URL: https://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2020-08-12]
48. Kalman M, Inchley J, Sigmundova D, Iannotti RJ, Tynjälä JA, Hamrik Z, et al. Secular trends in moderate-to-vigorous physical activity in 32 countries from 2002 to 2010: a cross-national perspective. Eur J Public Health 2015 Apr;25 Suppl 2:37-40. [doi: 10.1093/eurpub/ckv024] [Medline: 25805785]
49. Tremblay M, Barnes J, González SA, Katzmarzyk PT, Onywera VO, Reilly JJ, Global Matrix 2.0 Research Team. Global Matrix 2.0: Report Card Grades on the Physical Activity of Children and Youth Comparing 38 Countries. J Phys Act Health 2016 Nov;13(11 Suppl 2):S343-S366. [doi: 10.1123/jpah.2016-0594] [Medline: 27848745]
50. Wearables - Ireland 2019. Statista. URL: https://www.statista.com/outlook/319/140/wearables/ireland [accessed 2020-08-12]
51. Wearables - Finland 2019. Statista. URL: https://www.statista.com/outlook/319/135/wearables/finland [accessed 2020-08-12]
52. Gaudet J, Gallant F, Bélanger M. A Bit of Fit: Minimalist Intervention in Adolescents Based on a Physical Activity Tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
53. Cullen KW, Thompson D, Boushey C, Konzelmann K, Chen T. Evaluation of a web-based program promoting healthy eating and physical activity for adolescents: teen choice: food and fitness. Health Educ Res 2013 Aug;28(4):704-714. [doi: 10.1093/her/cyt059] [Medline: 23748162]
54. Goodyear VA, Kerner C, Quennerstedt M. Young people's uses of wearable healthy lifestyle technologies; surveillance, self-surveillance and resistance. Sport, Education and Society 2017 Sep 22;24(3):212-225. [doi: 10.1080/13573322.2017.1375907]
55. Seshadri DR, Drummond C, Craker J, Rowbottom JR, Voos JE. Wearable Devices for Sports: New Integrated Technologies Allow Coaches, Physicians, and Trainers to Better Understand the Physical Demands of Athletes in Real time. IEEE Pulse 2017 Jan;8(1):38-43. [doi: 10.1109/mpul.2016.2627240]
56. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfes CM, Zorman CA, et al. Wearable sensors for monitoring the internal and external workload of the athlete. NPJ Digit. Med 2019 Jul 29;2(1):71. [doi: 10.1038/s41746-019-0149-2]
57. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? Br J Sports Med 2016 Mar;50(5):273-280 [FREE Full text] [doi: 10.1136/bjsports-2015-095788] [Medline: 26758673]
58. Murtagh EM, Dempster M, Murphy MH. Determinants of uptake and maintenance of active commuting to school. Health Place 2016 Jul;40:9-14. [doi: 10.1016/j.healthplace.2016.04.009] [Medline: 27160529]
59. Nelson NM, Foley E, O'Gorman DJ, Moyna NM, Woods CB. Active commuting to school: how far is too far? Int J Behav Nutr Phys Act 2008;5(1):1. [doi: 10.1186/1479-5868-5-1]
60. Kallio J, Turpeinen S, Hakonen H, Tammelin T. Active commuting to school in Finland, the potential for physical activity increase in different seasons. Int J Circumpolar Health 2016;75:33319 [FREE Full text] [doi: 10.3402/ijch.v75.33319] [Medline: 27924739]
61. Hunter RF, de Silva D, Reynolds V, Bird W, Fox KR. International inter-school competition to encourage children to walk to school: a mixed methods feasibility study. BMC Res Notes 2015 Jan 27;8:19 [FREE Full text] [doi: 10.1186/s13104-014-0959-x] [Medline: 25623073]

62. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on Physical Activity: Study and Implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
63. Mollee J, Middelweerd A, Kurvers R, Klein M. What technological features are used in smartphone apps that promote physical activity? A review and content analysis. Pers Ubiquit Comput 2017 Jul 5;21(4):633-643. [doi: 10.1007/s00779-017-1023-3]
64. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. A cluster-randomised controlled trial to promote physical activity in adolescents: the Raising Awareness of Physical Activity (RAW-PA) Study. BMC Public Health 2017 Jan 04;17(1):6 [FREE Full text] [doi: 10.1186/s12889-016-3945-5] [Medline: 28052773]
65. Ahde-Deal P, Mikkonen J. Hello, Bracelets communicating nearby presence of friends. 2008 Oct 4 Presented at: 10th Conference on Participatory Design; 2008-10-04; Bloomington, Indiana, USA URL: https://dl.acm.org/doi/abs/10.5555/1795234.1795326
66. Montero C, Kärnä E, Ng K. Digipack Pro: Revamping Social InteractionsSupporting Physical Activity. USA: Association for Computing Machinery; 2019 Presented at: Proceedings of the 18th ACM International Conference on Interaction Design Children; 2019-06; Boise, ID, USA p. A. [doi: 10.1145/3311927.3326595]
67. Hurford R. Chapter 2: Types of smart clothes and wearable technology. In: Smart clothes and wearable technology. United Kingdom: Woodhead Publishing; 2009:25-44.
68. Everitt B. Cluster analysis. Qual Quant 1980 Jan;14(1):75-100. [doi: 10.1007/bf00154794]
69. Kaufman L, Rousseeuw P. Finding groups in data: An introduction to cluster analysis. Hoboken: John Wiley & Sons Inc; 1990:A.
70. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: Toward an integrative model of change. Journal of Consulting and Clinical Psychology 1983;51(3):390-395. [doi: 10.1037/0022-006x.51.3.390]
71. Ruckenstein M, Pantzar M. Beyond the Quantified Self: Thematic exploration of a dataistic paradigm. New Media & Society 2016 Jul 09;19(3):401-418. [doi: 10.1177/1461444815609081]
72. Hagger MS, Chatzisarantis N, Biddle SJ. The influence of self-efficacy and past behaviour on the physical activity intentions of young people. J Sports Sci 2001 Sep;19(9):711-725. [doi: 10.1080/02640410152475847] [Medline: 11522147]
73. Kokko S. Sports clubs as settings for health promotion: fundamentals and an overview to research. Scand J Public Health 2014 Nov;42(15 Suppl):60-65. [doi: 10.1177/1403494814545105] [Medline: 25416575]
74. Ng K, Tynjälä J, Sigmundová D, Augustine L, Sentenac M, Rintala P, et al. Physical Activity Among Adolescents With Long-Term Illnesses or Disabilities in 15 European Countries. Adapt Phys Activ Q 2017 Oct 01;34(4):456-465. [doi: 10.1123/apaq.2016-0138] [Medline: 28985099]
75. Vooijs M, Alpay LL, Snoeck-Stroband JB, Beerthuizen T, Siemonsma PC, Abbink JJ, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. Interact J Med Res 2014 Oct 27;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]
76. Vaes AW, Cheung A, Atakhorrami M, Groenen MTJ, Amft O, Franssen FME, et al. Effect of 'activity monitor-based' counseling on physical activity and health-related outcomes in patients with chronic diseases: A systematic review and meta-analysis. Ann Med 2013 Sep;45(5-6):397-412. [doi: 10.3109/07853890.2013.810891] [Medline: 23952917]
77. Thilarajah S, Clark RA, Williams G. Wearable sensors and Mobile Health (mHealth) technologies to assess and promote physical activity in stroke: a narrative review. Brain Impairment 2016 Feb 15;17(1):34-42. [doi: 10.1017/brimp.2016.1]
78. Tang KT, Richardson AM, Maxwell D, Spence WD, Stansfield BW. Evaluation of an activity monitor for the objective measurement of free-living physical activity in children with cerebral palsy. Arch Phys Med Rehabil 2013 Dec;94(12):2549-2558. [doi: 10.1016/j.apmr.2013.07.019] [Medline: 23924440]

Abbreviations

CSPPA: Children's Sport Participation and Physical Activity
FAS: Family Affluence Scale
F-SPA: Finnish School-aged Physical Activity
MVPA: moderate-to-vigorous physical activity
OECD: Organization for Economic Cooperation and Development
OR: odds ratio
PAT: physical activity tracker

Edited by G Eysenbach; submitted 06.03.20; peer-reviewed by A Henriksen, M Nomali; comments to author 12.06.20; revised version received 14.06.20; accepted 14.06.20; published 01.09.20.

Please cite as:
Ng K, Kokko S, Tammelin T, Kallio J, Belton S, O'Brien W, Murphy M, Powell C, Woods C
Clusters of Adolescent Physical Activity Tracker Patterns and Their Associations With Physical Activity Behaviors in Finland and Ireland: Cross-Sectional Study
J Med Internet Res 2020;22(9):e18509
URL: https://www.jmir.org/2020/9/e18509
doi:10.2196/18509
PMID:32667894

Original Paper

Cost-Effectiveness of a Continuous Glucose Monitoring Mobile App for Patients With Type 2 Diabetes Mellitus: Analysis Simulation

Shintaro Tsuji[1*], PhD; Tomoki Ishikawa[1,2*], MSc, MBA, PhD; Yasuhiro Morii[1], MSc, PhD; Hongjian Zhang[1], MSc; Teppei Suzuki[1,3], MSc, MBA, PhD; Takumi Tanikawa[4], PhD; Jun Nakaya[1], MD, PhD; Katsuhiko Ogasawara[1], MBA, PhD

[1]Graduate School of Health Sciences, Hokkaido University, Sapporo, Japan

[2]Institute for Health Economics and Policy, Tokyo, Japan

[3]Art and Sports Business, Iwamizawa Campus, Hokkaido University of Education, Hokkaido, Japan

[4]Department of Radiological Technology, Hokkaido University of Science, Sapporo, Japan

[*]these authors contributed equally

Corresponding Author:
Katsuhiko Ogasawara, MBA, PhD
Graduate School of Health Sciences
Hokkaido University
N12-W5, Kita-ku
Sapporo, 060-0812
Japan
Phone: 81 117063409
Email: oga@hs.hokudai.ac.jp

Abstract

Background: Apps for real-time continuous glucose monitoring (CGM) on smartphones and other devices linked to CGM systems have recently been developed, and such CGM apps are also coming into use in Japan. In comparison with conventional retrospective CGM, the use of CGM apps improves patients' own blood glucose control, which is expected to help slow the progression of type 2 diabetes mellitus (DM) and prevent complications, but the effect of their introduction on medical costs remains unknown.

Objective: Our objective in this study was to perform an economic appraisal of CGM apps from the viewpoint of assessing public medical costs associated with type 2 DM, using the probability of developing type 2 DM–associated complications, and data on medical costs and utility value to carry out a medical cost simulation using a Markov model in order to ascertain the cost-effectiveness of the apps.

Methods: We developed a Markov model with the transition states of insulin therapy, nephrosis, dialysis, and cardiovascular disease, all of which have a major effect on medical costs, to identify changes in medical costs and utility values resulting from the introduction of a CGM app and calculated the incremental cost-effectiveness ratio (ICER).

Results: The ICER for CGM app use was US $33,039/quality-adjusted life year (QALY).

Conclusions: Sensitivity analyses showed that, with the exception of conditions where the transition probability of insulin therapy, utility value, or increased medical costs increases, the ICER for the introduction of CGM apps was below the threshold of US $43,478/QALY used by the Central Social Insurance Medical Council. Our results provide basic data on the cost-effectiveness of introducing CGM apps, which are currently starting to come into use.

(J Med Internet Res 2020;22(9):e16053) doi:10.2196/16053

KEYWORDS

Markov model; telehealth; continuous glucose monitoring (CGM); type 2 diabetes mellitus; cost-effectiveness; incremental cost and effective ratio (ICER)

Introduction

According to a 2016 World Health Organization (WHO) global report of patients with type 2 diabetes mellitus (DM), there are currently over 420 million people with this condition worldwide, a 3.9-fold increase compared with 1980 [1]. In Japan, the number of people with type 2 DM reached 3.29 million in 2017, representing an increase of over 120,000 from the previous survey in 2014 [2]. The increased medical costs of diabetes and related complications associated with the rising number of people with type 2 DM are becoming an issue [3]. The total medical cost of type 2 DM in Japan is US $10.6 billion, and the cost of related diseases is US $38.3 billion, accounting for around 14% of the country's total medical expenditure of US $354.8 billion [4]. The complications of type 2 DM include nephropathy, neuropathy, and retinopathy; nephropathy, in particular, is a driver of increasing medical costs because patients with this condition frequently transition to dialysis, which costs around 5 million yen per year. To prevent medical costs from ballooning further, it will be necessary to both prevent the onset of type 2 DM itself and control its complications, which requires control of elevated blood glucose levels with insulin administration or oral medication. Proactive treatments to lower blood glucose levels, however, are leading to increases in the occurrence of cardiovascular disease and coma as side effects of hypoglycemia [5,6]. In a recent large-scale retrospective study using a medical fees database, around 20,000 patients with type 2 DM were hospitalized annually for hypoglycemia, 3.8% of whom died [7]. In the same analysis, the estimated incidence of hypoglycemia in patients with type 2 DM was 2.1 episodes/1000 people/year and was associated with incidents such as traffic accidents and falls. Preventing the progression of type 2 DM and its associated complications thus requires treatments to control hyperglycemia while avoiding putting patients in a hypoglycemic state, and this necessitates the daily real-time assessment of blood glucose levels by individual patients. One current method of hypoglycemia management for patients with type 2 DM is retrospective continuous glucose monitoring (CGM), in which blood glucose measurements over a certain period are aggregated and the user examines them after the event to determine any trends [8-10], but this method entails measuring blood glucose levels at regular intervals, and checking blood glucose levels in real time and taking measures to avert hypoglycemia or other conditions are difficult tasks. To resolve this issue, apps for real-time CGM on smartphones and other devices linked to CGM systems have recently been developed, and such CGM apps are also coming into use in Japan. CGM apps enable medical staff in distant locations and patients' families to monitor and share blood glucose level data via smartphones or other mobile devices. Clinical trials are currently underway to investigate the clinical effectiveness of CGM apps in actual use, such as whether they reduce the incidence of complications or have any effect on HbA1c levels, and it is hoped that, in comparison with conventional retrospective CGM, they may help delay the progression of type 2 DM and control complications by improving patients' own blood glucose control, enabling real-time patient guidance and family members to monitor patients' blood glucose levels. In advance of their introduction, however, simulations to establish their effect on medical costs will also be required. Our objective in this study was to carry out an economic evaluation of CGM apps from the viewpoint of assessing public medical costs associated with type 2 DM, using a Markov model of the probability of transitioning to type 2 DM–associated complications and data on medical costs and utility value to carry out a medical cost simulation in order to ascertain the cost-effectiveness of the apps.

Methods

Analysis Design

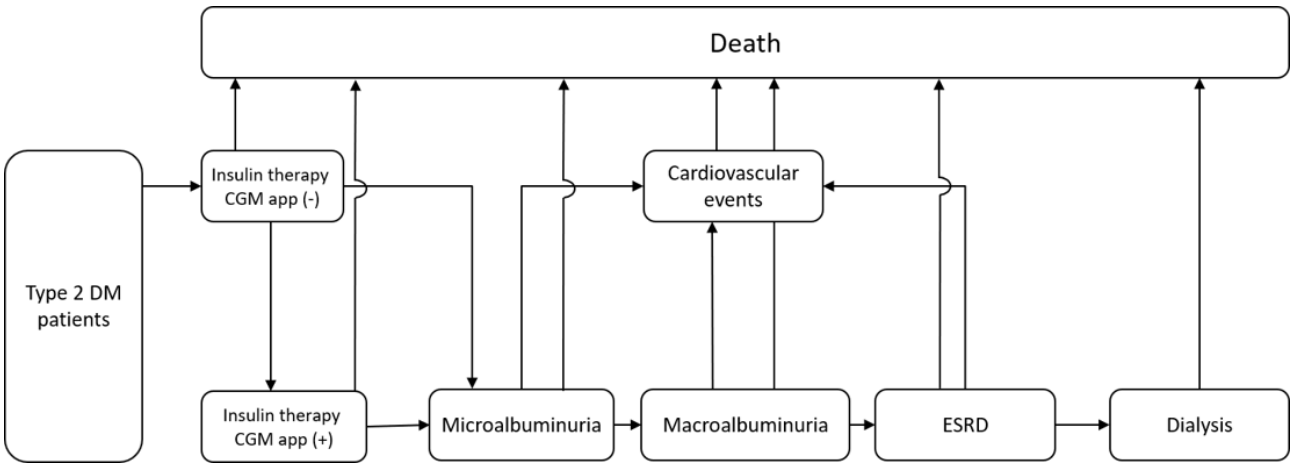
The subjects of this study were patients with type 2 DM receiving insulin therapy, and changes in their medical costs and utility values as the result of the introduction of a CGM app were estimated using a Markov model. In general, there are several modeling designs to measure or simulate cost-effectiveness, of which the most common are the Markov and decision tree models. The latter is used in situations where cost-effectiveness of multiple treatments are evaluated, when the patient condition changes irreversibly, or with a relatively short time period, while the Markov model is more suitable for diseases that span longer time periods, such as chronic diseases [11]. The duration of DM treatments is generally long, and the patient condition can change multiple times. For these reasons, the Markov model was more suitable than the decision tree model for this study.

The main patient states in the constructed model were diabetic nephropathy, dialysis, and cardiovascular disease, all of which have major effects on the medical costs of patients undergoing insulin treatment. To simulate this in a more detailed way, the patient's disease situation was classified as any of following; insulin therapy, microalbuminuria, macroalbuminuria, end-stage renal disease (ESRD) dialysis, cardiovascular events, or death. The assumption was that the first complication developed by patients receiving insulin therapy would be diabetic nephropathy, followed by either cardiovascular disease induced by type 2 DM or the introduction of dialysis. It was also assumed that diabetic nephropathy would progress to the point at which dialysis was required. The model also assumed that patients with diabetic nephropathy, with cardiovascular disease, or undergoing dialysis would eventually die. The Markov model was updated at yearly intervals for a 20-year analysis period (Figure 1).

The Markov model was adopted to define the patient's chronic condition at any point in time. The construction of the model is based on the Markov property that the state at any time point is dependent only on the state at the previous time point. If the initial number of patients used in the model is X, the model is updated at 1-year intervals, and the total time period is n years, the medical costs of state k ($k = j_1, j_2, j_3, \ldots j_n$) in year t ($===t = 1, 2, 3, \ldots n$) can be expressed by Equation (1), where the transition probability is $p_{(jt,jt\text{-}1)} = P(k_t = j_t \mid k_{t-1} = j_{t-1})$ and the discount rate on yearly medical costs is 4%, with $C_{kt} = C_{t-1}/(1+0.04)^t$ [8]. Similarly, the total medical costs in the model

can be expressed as the cumulative sum for the patient states over the number of years concerned.

Figure 1. Overview of the Markov model. CGM: continuous glucose monitoring; DM: diabetes mellitus; ESRD: end-stage renal disease.



Initial Sample Size

We initially included 1.2 million patients identified as undergoing insulin therapy from open data held in the NDB (the Japanese national database of health insurance claims and specific health checkup data) published by the Ministry of Health, Labour and Welfare (MHLW) [12]; after the number of patients was adjusted for the number of times a claim was made for a "blood glucose self-measuring instrument," the final sample size included 686,000 patients judged to be capable of using a CGM, who had not developed nephrosis [13], who were able to use a smartphone [14], and who had not had any skin problems caused by sensor-induced itching [15]. Surveys have indicated that the rate of use of health care apps is approximately 10% [16], and it was assumed that this will tend to increase in the future, so the transition probability that patients will start using the app was set to correspond to 15% of the patients in the study over the 20-year period assumed by the model, equivalent to 100,000 users.

Assignment of Transition Probabilities in the CGM App Non-Use Model

The transition probability that patients receiving insulin therapy would develop diabetic nephropathy was assigned with reference to the results of the Kumamoto Study [13] of type 2 DM patients in Japan. The transition probability that patients with diabetic nephropathy would develop cardiovascular disease induced by type 2 DM was defined using results of the United Kingdom Prospective Diabetes Study [17]. The ultimate transition probability from diabetic nephropathy to death was defined using the Japan Society for Dialysis Therapy Annual Report [18], and the transition probability from cardiovascular disease induced by type 2 DM to death was defined with reference to the numbers of deaths from different diseases published by the MHLW [19]. This shows the transition probabilities between the different states for patients not using a CGM app (upper part of Figure 1). We treated the transition probabilities in the models as constant values during the time horizon.

Assignment of Transition Probabilities in the CGM App Use Model

The probability that patients undergoing insulin therapy while using a CGM app was defined as 2.6% of the initial patient population described previously. The transition probability that patients receiving insulin therapy while using any CGM app would develop diabetic nephropathy was obtained from the rate of onset of patients in the Kumamoto Study [13], which is a randomized prospective 6-year study of complications of type 2 DM. The other transition probabilities for patients using a CGM app were assumed to be equivalent to those of patients not using such an app. Sensitivity analyses were conducted to investigate the effects of the probabilities so as to exclude arbitrariness of value selection at this point.

Validity of the Models

Finally, to verify the validity of the Markov model for CGM app non-use, clinical data not used in the development of the model were used to verify the estimated medical costs. This shows the transition probabilities between the different states for patients using a CGM app (lower part of Figure 1).

Assignment of Quality-Adjusted Life Years (QALY)

The index of effectiveness used in the cost-effectiveness analysis performed in this study was quality adjusted life years (QALY). QALY is an indicator of survival that takes quality of life into account. The utility value was set at 1 for good health and 0 for death, with states under illness expressed as numbers between 0 and 1. Utility values can be measured by means such as quality of life surveys.

Table 1. Probability value of each patient condition.

Transition (to each condition)	Probability value, %	References
Insulin therapy	2.60	National Health and Nutrition Survey [20]
Microalbuminuria		
App use group	4.8	Statistics of Medical Care Activities in Public Health Insurance [21]
App non-use group	1.6	Statistics of Medical Care Activities in Public Health Insurance [21]
Macroalbuminuria	2.80	United Kingdom Prospective Diabetes Study 64 [17]
Dialysis	2.30	Patient surveys [22]
Cardiovascular events	10.0	Viana et al [23]
Death	12.3	Patient surveys [22]

In this study, we used the utility value measured for type 2 DM patients in Japan. When they could not be obtained from the literature, the utility values for patients with complications were analyzed as the range of variation in the sensitivity analysis ±10% so as to grasp the range of uncertainty, which at the same time excludes arbitrariness in the setting of each parameters. For example, the utility values for the different stages of diabetic nephropathy, one of the conditions addressed in this study, comprising microalbuminuria, macroalbuminuria, and ESRD, were assumed to be values for diabetic nephropathy during insulin therapy without taking the stage classification by Sakamaki et al [24] into account. The utility values of diabetic nephropathy in patients undergoing insulin treatment in combination with either a cardiovascular event or dialysis were taken from the literature, as shown in Table 1. A 4% discount rate was applied to utility values.

Assignment of Medical Costs

To start with, values for the medical costs of insulin therapy, microalbuminuria, macroalbuminuria, ESRD, cardiovascular disease (induced by type 2 DM), and dialysis were taken from those used in previous studies [25-30] (Table 2). The medical cost of a CGM app was calculated as the cost of a CGM sensor (US $57.39 [6,600 yen] × 3) and transmitter (US $761.74 [87,600 yen]), with the cost of installing and using the app on a mobile device assumed to be free. The medical costs of the various complications, such as developing ESRD while undergoing insulin therapy, were calculated by combining the costs of the different states, as shown in Table 3. The medical costs used in this study were based on outpatient treatment.

To calculate the changes in medical costs for each disease state, the yearly medical costs and the yearly transition probabilities were multiplied by the initial numbers of patients, and the values were aggregated over a 20-year period. Medical expenses estimated from the Markov model for *n* years were calculated using Equation (2). In this equation, all indices are the same as in Equation (1).

Table 2. Utility value of each patient condition.

Patient condition	Utility value	References
Insulin therapy	0.83 (0.79-0.88)	Sakamaki et al [24]
Insulin therapy + diabetic nephropathy	0.81 (0.72-0.90)	Sakamaki et al [24]
Microalbuminuria	0.81	Sakamaki et al [24], Okubo et al [31]
Macroalbuminuria	0.81	Sakamaki et al [24], Okubo et al [31]
End-stage renal disease	0.81	Sakamaki et al [24], Okubo et al [31]
Cardiovascular events	0.71	Hara et al [25]
Diabetic nephropathy	0.68	Takura et al [26]

Table 3. Medical fee for each patient status.

Patient condition	Annual medical expenses, US $ (yen)	References
Insulin therapy	4,891.76 (562,552)	Ministry of Internal Affairs and Communication [14], Wong et al [15], Dentsu Digital [16]
Microalbuminuria	1892.61 (217,650)	Ministry of Health, Labour and Welfare [19]
Macroalbuminuria	3256.41 (374,487)	Ministry of Health, Labour and Welfare [19]
ESRD[a]	6564.92 (754,966)	Ministry of Health, Labour and Welfare [19]
Cardiovascular events from diabetes	3587.30 (412,540)	Dentsu Digital [16]
Dialysis	41,739.13 (4,800,000)	Ministry of Health, Labour and Welfare [4]
CGM[b] app	2,827.83 (3,25,200)	Ministry of Health, Labour and Welfare [19], Sakamaki et al [24]

[a]ESRD: end-stage renal disease.

[b]CGM: continuous glucose monitoring.

Sensitivity Analysis

To confirm the robustness of our model, we carried out sensitivity analyses of stochastic transitions, utility values, and medical costs. The values for the sensitivity analysis of stochastic transitions were taken from the literature, if standard deviations were available, and were assigned from 0.5-fold to 2.0-fold if not. The values in the sensitivity analyses of both medical costs and utility values were a uniform ±10%.

Calculation of the Incremental Cost-Effectiveness Ratio

We investigated the effectiveness of the use of a CGM app by calculating its incremental cost-effectiveness ratio (ICER). The ICER is obtained by dividing the difference between the medical cost of the medical technology under evaluation and that of the conventional technology (Δ Cost) by the difference between the utility value of the medical technology under evaluation and that of the conventional technology (Δ Effect).

Investigation of the Validity of the Markov Model

We investigated the validity of our Markov model from two aspects. The first was its internal validity, in which we considered it from the perspectives of whether it fully expressed the natural progression of the states concerned and whether the parameters used were appropriate. The second was its external validity, which we investigated by replacing the clinical data used in the model with other data and examining whether the estimated values thus obtained were appropriate. We picked up complications associated with DM, particularly the medical expenditure of a complication. In the process of model construction and selection of the parameters, the model overview and each parameter were reviewed by a cardiology specialist. At the same time, we adopted the values of the parameters from previous studies conducted in Japan. These operations guaranteed the validity of our Markov model, and we judged it could concisely express the progression of the states and the difference of medical expenditure derived from type 2 DM.

Results

Medical Costs

The total medical costs for patients undergoing insulin therapy were US $50,417,581,024 over 20 years with the use of a CGM app and US $47,817,427,894 over 20 years without app use, an increase of US $2,600,153,130 over 20 years. The increases in utility value, ICER, and QALY over 20 years were 78,699, US $33,039/QALY, and 0.11 QALY per person, respectively.

In terms of the medical costs for each patient state, CGM app use increased the costs (over 20 years) of insulin therapy (US $3,503,535,847), macroalbuminuria (US $20,981,856), and dialysis (US $1,465,525), whereas it decreased the cost (over 20 years) of microalbuminuria (US $842,698,713), ESRD (US $4,280,039), and cardiovascular disease (US $78,851,346). In terms of changes in patient numbers, the number of patients undergoing insulin therapy increased by 21,649 people over 20 years, and the number with macroalbuminuria increased by 830 people over 20 years, whereas the number of patients with microalbuminuria decreased by 13,547 people over 20 years, with ESRD by 22 people over 20 years, undergoing dialysis by 23 people over 20 years, and with cardiovascular disease by 3357 people over 20 years. There were 5529 fewer deaths over 20 years.

Sensitivity Analyses

The sensitivity analysis of transition probabilities showed that the minimum and maximum ICERs (US $26,989/QALY and US $54,650/QALY) were obtained when the transition probability from insulin therapy to macroalbuminuria was 0.5 times or 2 times the standard value (1.24% and 5.35%, respectively; Figure 2). The sensitivity analysis of utility values showed that a variation of ±5% in the utility value of insulin therapy caused the ICER to vary from US $21,516/QALY to US $71,142/QALY (Figure 3). Finally, the sensitivity analysis of medical costs showed that the maximum variation occurred when the medical cost of insulin therapy was increased by 10% (US $42,436/QALY) and the medical cost of macroalbuminuria was decreased by 10% (US $31,732/QALY; Figure 4) We carried out sensitivity analyses to investigate the effect of variations in the CGM app use model parameters (transition

probabilities, utility values, and medical costs) on the ICER. First, the results of the sensitivity analysis of transition probabilities showed that the ICER for transitioning from insulin therapy to macroalbuminuria was US $54,650/QALY. Although this is below the ICER used by the WHO [29], it greatly exceeds the value of US $43,478 (5 million yen) per QALY used by the Central Social Insurance Medical Council. The same trend was also evident for the second and third highest ICERs, which were seen when the transition probabilities from insulin therapy to microalbuminuria and from microalbuminuria to macroalbuminuria, respectively, were increased. From the ICER perspective, the introduction of CGM apps for patients who develop diabetic nephropathy and whose condition progresses rapidly should be viewed with caution. Next, the results of the sensitivity analysis of utility values showed that a reduction in the utility value of insulin therapy dramatically increased the ICER (to US $71,142/QALY). However, as the utility value for insulin therapy used in this study for patients using a CGM app was similar to that for patients not using such an app, the introduction of CGM apps is unlikely to cause the utility value to decline, and it is more likely that a rise in the utility value would decrease the ICER. Increasing the utility value by 10% may also reduce the ICER to US $21,516/QALY. With respect to the effect of a decrease in the utility value of microalbuminuria, which exhibited a high ICER, the value of US $35,517/QALY was well below the value of US $43,478/QALY used by the Central Social Insurance Medical Council, suggesting that it did not correspond to a factor causing major variation in the ICER obtained in this study. Finally, the sensitivity analysis of medical costs showed that an increase of 10% in the cost of insulin therapy increased the ICER from US $33,039/QALY to US $42,436/QALY (increase of US $9397/QALY), but this was very close to the value used by the Central Social Insurance Medical Council (US $43,478/QALY), suggesting that the introduction of CGM apps should be considered even if the cost of insulin therapy and other medical costs were to increase.

The WHO suggests that the ICER for the introduction of a new medical technology should be no more than three times GDP per capita [32]. Applied to the GDP of Japan, this figure would be US $116,649/QALY (as of 2016). The ICER used by the Central Social Insurance Medical Council is US $43,478/QALY (5,000,000 yen/QALY), less than one-quarter of the value proposed by the WHO. The ICER obtained in this study (US $33,039/QALY) is well below either of these two thresholds. In comparison, cost-effectiveness analyses of retrospective CGM (in type 1 DM patients) have reported ICERs between US $41,000/QALY and US $99,000/QALY [33-35]. For example, a cost-effectiveness analysis of self-injection compared with CGM plus intensified insulin therapy found that, over a period of 33 years, the ICER for CGM plus intensified insulin therapy was US US $45,033/QALY [32]. The ICER for the introduction of CGM apps that we found in our study (US $33,039/QALY) was thus both lower than the ICER for the conventional method of retrospective CGM and the ICERs suggested by the WHO and the Central Social Insurance Medical Council (US $116,649/QALY and US $43,478/QALY, respectively), indicating that the introduction of this medical technology should be considered in Japan.

Figure 2. Sensitivity analysis of the incremental cost-effectiveness ratio (ICER) using transition probabilities. CVD: cardiovascular disease; ESRD: end-stage renal disease; QALY: quality-adjusted life year.

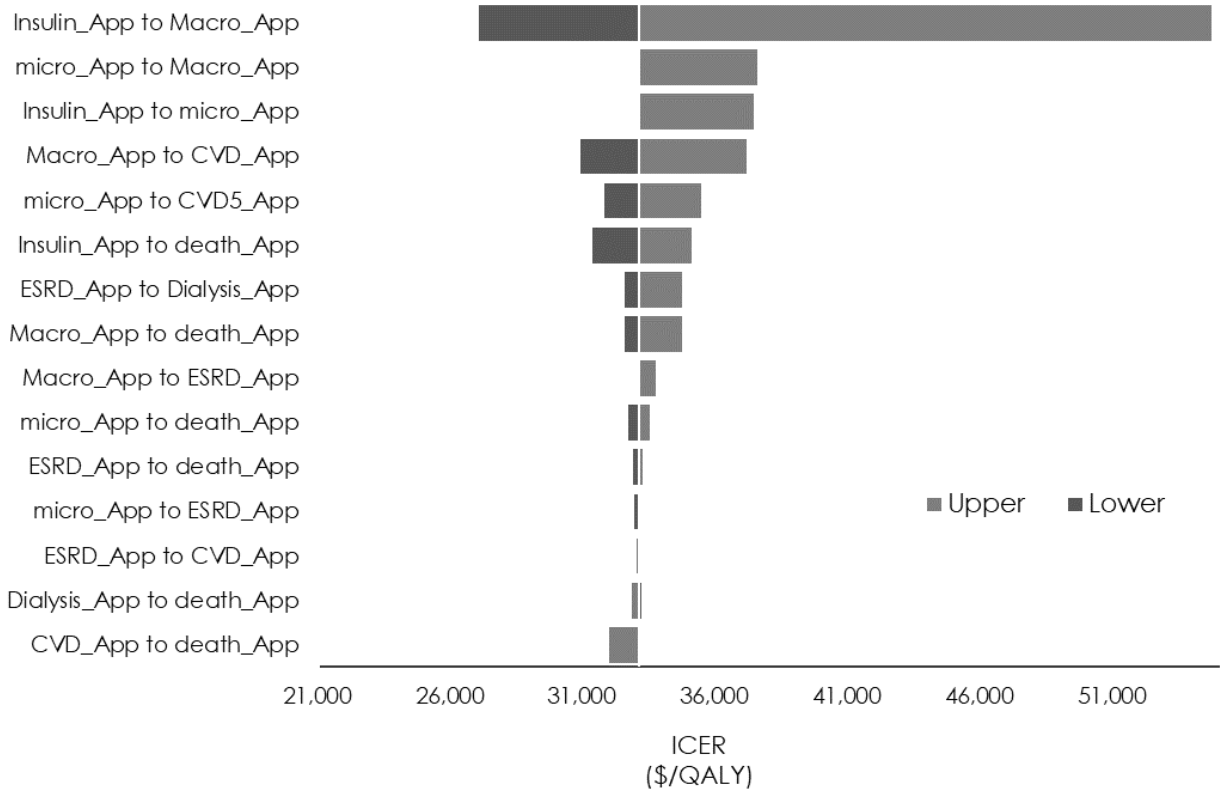


Figure 3. Sensitivity analysis of the incremental cost-effectiveness ratio (ICER) using utility values. CVD: cardiovascular disease; ESRD: end-stage renal disease; QALY: quality-adjusted life year.

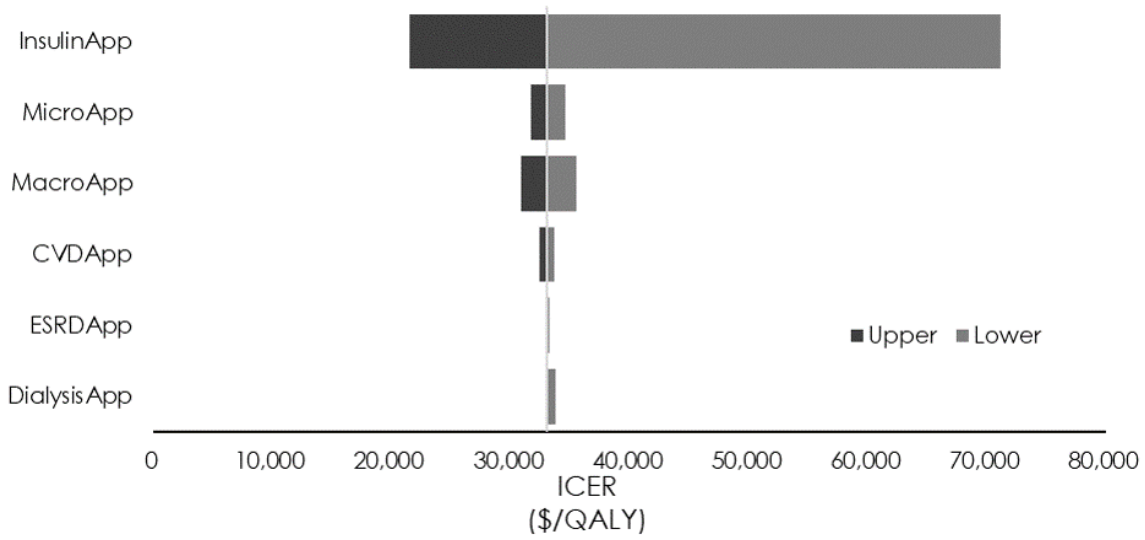
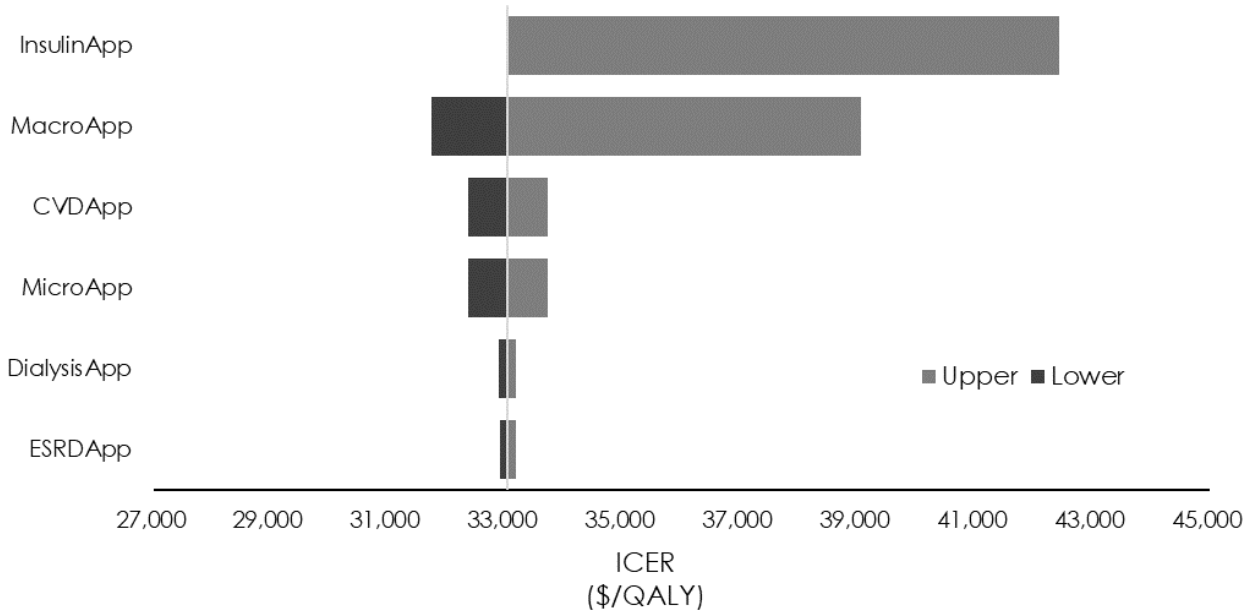


Figure 4. Sensitivity analysis of the incremental cost-effectiveness ratio (ICER) using medical fees. CVD: cardiovascular disease; ESRD: end-stage renal disease; QALY: quality-adjusted life year.



Discussion

In this study, we used a Markov model to calculate the ICER of the use of a CGM app to investigate the cost-effectiveness of such apps for patients receiving insulin treatment and calculated that the ICER was US $33,039/QALY.

Internal and External Validity of the Model

The model used in this study was developed around the medical costs for patients undergoing insulin treatment with a focus on diabetic nephropathy, dialysis, and cardiovascular disease, all of which entail high medical costs. As far as possible, the values assigned to the parameters were obtained from publications describing patients of similar ages and backgrounds. The transition probability values and utility values for patients using a CGM app were basically taken from clinical data on intensified insulin therapy with the aim of avoiding any overestimation of the effect achieved by the introduction of CGM apps, and although the range of variation in the parameters as a result of the introduction of CGM apps is currently unclear, this was assured by sensitivity analysis. For our model in this study, we chose diseases with a major effect on medical costs and for which data on each of the parameters were available; for this reason, the model did not include neuropathy and retinopathy, two of the three major complications of type 2 DM. Because we were unable to obtain sufficient data on the proportion of patients whose diseases improved, this was not reflected in our model. On these two points, the model must be revised and revalidated as and when usable clinical data are published in the future.

In terms of the external validity of the model, a comparison of the transition probability for diabetic nephropathy used in this study and the prevalence reported by Yokoyama *et al [36]* showed that the prevalence of nephropathy in our study was higher until year 18 of type 2 DM, but that subsequently it was lower. One reason may have been that the prevalence of diabetic nephropathy is known to increase in patients who have had type 2 DM for longer [36], and the patients in the study by Yokoyama *et al* [36] were aged ≤30 years, whereas the data used in our study came from patients aged approximately 50 years. The difference in the transition probability caused by this age difference meant that the patients in our study may have transitioned to the state of death earlier than those studied by Yokoyama *et al* [36]. Although the clinical data were obtained from patients with diverse attributes, including some who had already developed complications, for the patients simulated in this study, the analysis started from the state of receiving insulin therapy alone, and this may have been another reason for the differences in prevalence.

The reported incidences of cardiovascular events induced by type 2 DM include those obtained from the Hisayama Study (5/1000 people/year) [37] and the Japan Diabetes Complications Study (14.7-17.4/1000 people/year) [38], and despite the differences in patient attributes, we obtained a similar rate in our study (11.0/1000 people/year over 20 years). Evidence from cohort studies will be required for further internal validation of the state of cardiovascular events. For dialysis, according to materials published by the Japan Society for Dialysis Therapy [18], diabetic nephropathy was the reason for the introduction of dialysis in 43.2% of patients who started dialysis in 2015. A calculation of the proportion of the 3,166,000 type 2 DM patients who were undergoing insulin therapy in 2015 and started on dialysis for diabetic nephropathy who were aged 70-75 years found that they constituted 24.18% of such patients, and in our model, the 1.20 million patients undergoing insulin treatment who were aged 70 years (after 20 years) also accounted for 25.16% of patients, a very similar figure. Other clinical data tend to compare data from patients with different attributes, and their evaluation is dependent on the data provided. Considered from the viewpoint of medical costs, the costs of treating type 2 DM, including insulin therapy, published by the MHLW in its 2016 summary of national medical costs (for patients aged 45-65 years, $2.7 billion yen for type 2 DM and $1.4 billion yen for cardiovascular disease) were higher than the medical costs for patients receiving insulin therapy estimated in this study (for patients age 60 years, approximately 2.5 billion yen), but the two were broadly consistent. To validate our model, comparisons with long-term cohort follow-up data are required, but the only method available is to compare the model's estimated values with the small number of cases reported in Japan; therefore, there is room for further analysis of the validation results.

Limitations and Topics for Further Investigation

This study has several limitations. One of the limitations is the problem of assigning transition probabilities, as shown in the results of the investigation of the model's external validity. The patients in this study were 50 years old, and over a 20-year analysis period, the probability of developing complications would change rapidly as they aged, potentially generating errors compared with the actual numbers of patients over longer analysis periods. To bridge this gap with reality, it will be necessary to divide the analysis period into shorter periods and assign different transition probabilities for the different periods.

Second, we could not consider unmeasured values because this analysis was based on a simulation. There is no evidence that the assumed value should be due to a lack of data, such as the proportion of patients whose diseases improved after using the CGM app.

This study assumed that CGM apps offer the advantages of enabling patients to control their own blood glucose levels by monitoring them in real time and share this information with medical institutions and family members. Although the economic evaluation itself is likely not new, this simulation method has not been applied to CGM apps yet, as far as we were able to determine. Analysis of the cost-effectiveness of the second of these advantages, that of sharing blood glucose level information, is a topic for further investigation.

The present study is limited by a lack of considering the effect by population aging, because we employed transition probabilities as a constant in the construction of the Markov model. For instance, it is assumed that the per capita medical cost increases as the population ages, but it was not possible to consider the decrease in the number of patients due to the rise in mortality. Although this could be overcome in future studies, the present study was unable to incorporate this consideration.

Finally, we assumed that the effect of CGM app use on type 2 DM is the same as intensified insulin therapy. At this point, there is no evidence to support this assumption. Therefore, we could only perform a simulation analysis based on some assumptions. To measure cost-effectiveness with greater accuracy, further research or accumulation of real-world data about app use is required.

Conclusion

We calculated that the ICER of the introduction of CGM apps for type 2 DM patients was $33,039/QALY, and a comparison with ICERs reported in previous studies and by independent organizations indicated that this is a medical technology worthy of consideration.

Conflicts of Interest

ST and KO received donation from Medtronic Foundation Japan, and this study was performed with a part of the support. .

References

1. World Health Organization. Global report on diabetes. 2016. URL: http://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;jsessionid=9D54951ECD3E4E76D766FC26F53ADDB4?sequence=1 [accessed 2018-10-04]
2. Ministry of Health, Labour and Welfare. 2014 patient survey. 2014. URL: https://www.mhlw.go.jp/toukei/saikin/hw/kanja/14/ [accessed 2018-10-04]
3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. The Lancet 2016 Apr;387(10027):1513-1530. [doi: 10.1016/S0140-6736(16)00618-8]
4. Ministry of Health, Labour and Welfare. On measures to prevent severe diabetic nephropathy. 2018. URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000114064_13.pdf [accessed 2018-10-04]
5. Goto A, Arah OA, Goto M, Terauchi Y, Noda M. Severe hypoglycaemia and cardiovascular disease: systematic review and meta-analysis with bias analysis. BMJ 2013 Jul 29;347:f4533 [FREE Full text] [doi: 10.1136/bmj.f4533] [Medline: 23900314]
6. Khunti K, Davies M, Majeed A, Thorsted BL, Wolden ML, Paul SK. Hypoglycemia and risk of cardiovascular disease and all-cause mortality in insulin-treated people with type 1 and type 2 diabetes: a cohort study. Diabetes Care 2015 Feb;38(2):316-322. [doi: 10.2337/dc14-0920] [Medline: 25492401]
7. Sako A, Yasunaga H, Matsui H, Fushimi K, Hamasaki H, Katsuyama H, et al. Hospitalization for Hypoglycemia in Japanese Diabetic Patients. Medicine 2015;94(25):e1029. [doi: 10.1097/md.0000000000001029]
8. Ministry of Land, Infrastructure, Transport and Tourism. A guideline of cost-benefit analysis on public sector projects 2008. URL: http://www.mlit.go.jp/tec/hyouka/public/090601/shishin/shishin090601.pdf [accessed 2020-02-23]
9. Davis T, Salahi A, Welsh JB, Bailey TS. Automated insulin pump suspension for hypoglycaemia mitigation: development, implementation and implications. Diabetes Obes Metab 2015 Dec;17(12):1126-1132. [doi: 10.1111/dom.12542] [Medline: 26179879]
10. Poolsup N, Suksomboon N, Kyaw A. Systematic review and meta-analysis of the effectiveness of continuous glucose monitoring (CGM) on glucose control in diabetes. Diabetol Metab Syndr 2013;5(1):39 [FREE Full text] [doi: 10.1186/1758-5996-5-39] [Medline: 23876067]
11. Kamae I. Health Technology Assessment in Japan: Policy, Pharmacoeconomic Methods and Guidelines, Value, and Beyond. Singapore: ADIS; 2019.
12. Ministry of Health, Labour and Welfare. 2nd NDB Open Database. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000177221.html [accessed 2018-10-04]
13. Ohkubo Y, Kishikawa H, Araki E, Miyata T, Isami S, Motoyoshi S, et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. Diabetes Research and Clinical Practice 1995 May;28(2):103-117. [doi: 10.1016/0168-8227(95)01064-k]
14. Ministry of Internal Affairs and Communications. Chapter 1: Present and Future of Smartphone Economy. Survey on trend of communication tool. 2017. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2017/chapter-1.pdf#page=1 [accessed 2019-01-10]
15. Wong JC, Foster NC, Maahs DM, Raghinaru D, Bergenstal RM, Ahmann AJ, T1D Exchange Clinic Network. Real-time continuous glucose monitoring among participants in the T1D Exchange clinic registry. Diabetes Care 2014 Oct 10;37(10):2702-2709 [FREE Full text] [doi: 10.2337/dc14-0303] [Medline: 25011947]
16. Dentsu Digital. The percentage of domestic mobile app users who use health and fitness apps is a growth industry for the future. URL: https://fuller-inc.com/news/2018/12/healthcare-fitness-report/release.pdf [accessed 2018-10-04]
17. Adler A, Stevens R, Manley S, Bilous RW, Cull CA, Holman RR, UKPDS GROUP. Development and progression of nephropathy in type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS 64). Kidney Int 2003 Jan;63(1):225-232 [FREE Full text] [doi: 10.1046/j.1523-1755.2003.00712.x] [Medline: 12472787]
18. The Japanese Society for Dialysis Therapy. An Overview of Regular Dialysis Treatment in Japan. 2016. URL: https://www.jsdt.or.jp/english/2426.html [accessed 2020-05-16]
19. Ministry of Health, Labour and Welfare. Vital statistics monthly report, total for the year 2015. URL: http://www.mhlw.go.jp/toukei/saikin/hw/jinkou/geppo/nengai15/index.html [accessed 2018-10-04]
20. Ministry of Health, Labour and Welfare. National Health and Nutrition Survey. URL: https://www.nibiohn.go.jp/eiken/english/research/project_nhns.html [accessed 2020-02-19]
21. Ministry of Health, Labour and Welfare. Survey of Medical Care Activities in Public Health Insurance. URL: https://www.mhlw.go.jp/english/database/db-hss/smcaphi.html [accessed 2020-02-19]
22. Ministry of Health, Labour and Welfare. Summary of patient surveys. URL: https://www.mhlw.go.jp/english/database/db-hss/ps.html [accessed 2020-02-19]
23. Viana LV, Gross JL, Camargo JL, Zelmanovitz T, da Costa Rocha EPC, Azevedo MJ. Prediction of cardiovascular events, diabetic nephropathy, and mortality by albumin concentration in a spot urine sample in patients with type 2 diabetes. J Diabetes Complications 2012 Sep;26(5):407-412. [doi: 10.1016/j.jdiacomp.2012.04.014] [Medline: 22677793]

24. Sakamaki H, Ikeda S, Ikegami N, Uchigata Y, Iwamoto Y, Origasa H, et al. Measurement of HRQL using EQ-5D in patients with type 2 diabetes mellitus in Japan. Value Health 2006 Jan;9(1):47-53 [FREE Full text] [doi: 10.1111/j.1524-4733.2006.00080.x] [Medline: 16441524]
25. Hara K, Fukumura K, Kobayashi Y, Komuro I, Nakajima T. Cost-effective analysis of cardiac rehabilitation. Official Journal of the Japanese Association of Cardiac Rehabilitation 2015;20(1):156-159.
26. Takura T, Nakanishi T, Kawanishi H, Nitta K, Akizawa T, Hiramatsu M, et al. Cost-Effectiveness of Maintenance Hemodialysis in Japan. Ther Apher Dial 2015 Oct 21;19(5):441-449. [doi: 10.1111/1744-9987.12314] [Medline: 26387878]
27. Tanaka M, Ito H, Oshikiri K. Factors in Medical Costs for Type 2 Diabetes Mellitus in the Setting of Outpatient Clinic. Journal of the Japan Diabetes Society 2012;55(3):193-198.
28. Kahara T, Yakeda A, Hashimoto M, Nishio K. Annual Medical Expenses for Outpatients with Type 2 Diabetes Mellitus Requiring Insulin Therapy—Patients <70 versus ≥70 Years Old. Journal of the Japan Diabetes Society 2009;52(6):437-442. [doi: 10.11213/tonyobyo.52.437]
29. Kakihara H, Oishi M. The direct costs of diabetes care for outpatients. Journal of the Japan Diabetes Society 1999;42:909-916.
30. Ministry of Health, Labour and Welfare. Briefing session on further development of prevention of severe diabetic nephropathy. URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/05_1.pdf [accessed 2018-10-04]
31. Okubo R, Kai H, Kondo M, Saito C, Yoh K, Morito N, et al. Health-related quality of life and prognosis in patients with chronic kidney disease: a 3-year follow-up study. Clin Exp Nephrol 2014 Oct;18(5):697-703. [doi: 10.1007/s10157-013-0901-x] [Medline: 24198050]
32. Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost–effectiveness of interventions: alternative approaches. Bull. World Health Organ 2014 Dec 15;93(2):118-124. [doi: 10.2471/blt.14.138206]
33. Huang ES, O'Grady M, Basu A, Winn A, John P, Lee J, Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. The cost-effectiveness of continuous glucose monitoring in type 1 diabetes. Diabetes Care 2010 Jun 23;33(6):1269-1274 [FREE Full text] [doi: 10.2337/dc09-2042] [Medline: 20332354]
34. McQueen RB, Ellis SL, Campbell JD, Nair KV, Sullivan PW. Cost-effectiveness of continuous glucose monitoring and intensive insulin therapy for type 1 diabetes. Cost Eff Resour Alloc 2011 Sep 14;9(1):13 [FREE Full text] [doi: 10.1186/1478-7547-9-13] [Medline: 21917132]
35. Eastman R, Leptien AD, Chase HP. Cost-effectiveness of use of the GlucoWatch Biographer in children and adolescents with type 1 diabetes: a preliminary analysis based on a randomized controlled trial. Pediatr Diabetes 2003 Jun;4(2):82-86. [doi: 10.1034/j.1399-5448.2003.00011.x] [Medline: 14655264]
36. Yokoyama H, Okudaira M, Otani T, Sato A, Miura J, Takaike H, et al. Higher incidence of diabetic nephropathy in type 2 than in type 1 diabetes in early-onset diabetes in Japan. Kidney Int 2000 Jul;58(1):302-311 [FREE Full text] [doi: 10.1046/j.1523-1755.2000.00166.x] [Medline: 10886575]
37. Kubo M, Kiyohara Y, Kato I, Tanizaki Y, Arima H, Tanaka K, et al. Trends in the Incidence, Mortality, and Survival Rate of Cardiovascular Disease in a Japanese Community. Stroke 2003 Oct;34(10):2349-2354. [doi: 10.1161/01.str.0000090348.52943.a2]
38. Control Group, Turnbull FM, Abraira C, Anderson RJ, Byington RP, Chalmers JP, et al. Intensive glucose control and macrovascular outcomes in type 2 diabetes. Diabetologia 2009 Nov;52(11):2288-2298. [doi: 10.1007/s00125-009-1470-0] [Medline: 19655124]

Abbreviations

CGM: continuous glucose monitoring
CVD: cardiovascular disease
DM: diabetes mellitus
ESRD: end-stage renal disease
ICER: incremental cost-effectiveness ratio
MHLW: Ministry of Health, Labour and Welfare
QALY: quality-adjusted life year
WHO: World Health Organization

Edited by G Eysenbach; submitted 29.08.19; peer-reviewed by M Zolnoori, M Huang, M Pobiruchin; comments to author 28.10.19; revised version received 23.02.20; accepted 03.06.20; published 17.09.20.

Please cite as:
Tsuji S, Ishikawa T, Morii Y, Zhang H, Suzuki T, Tanikawa T, Nakaya J, Ogasawara K
Cost-Effectiveness of a Continuous Glucose Monitoring Mobile App for Patients With Type 2 Diabetes Mellitus: Analysis Simulation
J Med Internet Res 2020;22(9):e16053
URL: https://www.jmir.org/2020/9/e16053
doi:10.2196/16053
PMID:32940613

Original Paper

Assessing Digital Health Implementation for a Pediatric Chronic Pain Intervention: Comparing the RE-AIM and BIT Frameworks Against Real-World Trial Data and Recommendations for Future Studies

Rocio de la Vega[1], DPhil; Lee Ritterband[2], DPhil; Tonya M Palermo[1], DPhil

[1]Center for Child Health, Behavior and Development, Seattle Children's Research Institute, Seattle, WA, United States

[2]Center for Behavioral Health & Technology, Department of Psychiatry and Neurobehavioral Sciences, University of Virginia School of Medicine, Charlottesville, VA, United States

Corresponding Author:
Tonya M Palermo, DPhil
Center for Child Health, Behavior and Development
Seattle Children's Research Institute
P.O. Box 5371/SCRI
M/S CW8-6
Seattle, WA, 98145-5005
United States
Phone: 1 206 884 4208
Email: tonya.palermo@seattlechildrens.org

Abstract

Background: Digital health interventions have demonstrated efficacy for several conditions including for pediatric chronic pain. However, the process of making interventions available to end users in an efficient and sustained way is challenging and remains a new area of research. To advance this field, comprehensive frameworks have been created.

Objective: The aim of this study is to compare the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) and Behavior Interventions using Technology (BIT) frameworks with data collected from the web-based management of adolescent pain (WebMAP Mobile; WMM) randomized controlled trial (RCT).

Methods: We conducted a hybrid effectiveness-implementation cluster RCT with a stepped wedge design in which the intervention was sequentially implemented in 8 clinics, following a usual care period. Participants were 143 youths (mean age 14.5 years, SD 1.9; 117/143, 81.8% female) with chronic pain, from which 73 were randomized to receive the active intervention. Implementation outcomes were assessed using the RE-AIM and BIT frameworks.

Results: According to the RE-AIM framework, the WMM showed excellent reach, recruiting a sample 19% larger than the size originally planned and consenting 79.0% (143/181) of eligible referred adolescents. Effectiveness was limited, with only global impression of change showing significantly greater improvements in the treatment group; however, greater treatment engagement was associated with greater reductions in pain and disability. Adoption was excellent (all the invited clinics participated and referred patients). Implementation was acceptable, showing good user engagement and moderate adherence and positive attitudes of providers. Costs were similar to planned, with a 7% increase in funds needed to make the WMM publicly available. Maintenance was evidenced by 56 new patients downloading the app during the maintenance period and by all clinics agreeing to continue making referrals and all, but one, making new referrals. According to the BIT, 82% (60/73) of adolescents considered the treatment acceptable. In terms of adoption, 93% (68/73) downloaded the app, and all of them used it after their first log-in. In terms of appropriateness at the user level, 2 participants were unable to download the app. Perceptions of the appearance, navigation, and theme were positive. Providers perceived the WMM as a good fit for their clinic, beneficial, helpful, and resource efficient. In terms of feasibility, no technical issues were reported. In terms of fidelity, 40% (29/73) completed the treatment. Implementation costs were 7% above the budget. With regard to penetration, 56 new users accessed the app during the maintenance period. In terms of sustainability, 88% (7/8) of clinics continued recommending the WMM after the end of the study.

Conclusions: For the first time, a real-world digital health intervention was used as a proof of concept to test all the domains in the RE-AIM and BIT frameworks, allowing for comparisons.

International Registered Report Identifier (IRRID): RR2-10.1016/j.cct.2018.10.003

(J Med Internet Res 2020;22(9):e19898) doi:10.2196/19898

KEYWORDS

mobile health; eHealth; mHealth; implementation science; chronic pain; adolescents; mobile phone

Introduction

Background

Digital health interventions have demonstrated efficacy for a variety of conditions such as diabetes [1], cancer [2], chronic pain [3,4], depression, and anxiety [5]. These interventions are becoming better established and integrated within health care systems (eg, MindSpot Clinic in Australia [6]) and within clinic-based care (eg, integrating monitoring data from wearables in the digital health records) and are increasingly adopted by end users directly (eg, direct to consumer apps [7]).

However, as digital health solutions demonstrate efficacy in clinical research trials, the process of making them available to end users in an efficient way (ie, implementation in real-world settings) is still challenging and remains a new area of research. This is a key element of knowledge mobilization (ie, making evidence-based interventions available to those in need) and includes determining the best ways to design and deliver the interventions to make them easy to adapt, engaging, and low in burden for the users, while also addressing the need to secure funds for sustainability (ie, covering costs for human and technical resources).

To make advances in this field and find the most efficient ways of creating and disseminating digital health interventions, comprehensive frameworks have been created to assess both the effectiveness and implementation of these programs and to create benchmarks that allow for comparisons between different implementation and dissemination strategies. Traditional, well-established frameworks for evaluating the public health impact of interventions, such as the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework [8], have been extensively used [9]. For example, *Reach* has been assessed by quantifying the number of potential clinics reached out of all potential clinics in a nationwide intervention or with the number of workers reached out of the total workers in a company; *Effectiveness* has been measured as the change in the quality of life reported by the participants of a diabetes management intervention or as the percentage of learning objectives met in a health training intervention, etc. However, a general public health impact framework may not be detailed or tailored enough to capture the unique elements and nuances of the implementation process of digital health interventions.

Previous Work

Recently, a framework was put forth to test the implementation of "Behavior Interventions using Technology (the BIT framework, henceforth)" [10], which recharacterized implementation outcomes for behavioral intervention technology.

The BIT framework is an effort to clarify and better illustrate another widely used implementation framework, "implementation dimensions for health service interventions" [11], which includes the following domains: acceptability, adoption, appropriateness, feasibility, fidelity, implementation costs, penetration, and sustainability. The BIT framework could help fulfill the need for a more detailed and practical framework. However, although Proctor's framework is commonly cited in digital health implementation efforts, it is not yet widely used. To our knowledge, there is only one peer-reviewed publication illustrating the different domains of the framework with examples taken from different interventions [11]. Although this is an important starting point, to date, all the domains in the BIT framework have not been tested in a single digital health trial. Further, it is unknown how the BIT framework directly compares with other established frameworks such as the RE-AIM (eg, what specific advantages may be incurred). Thus, our aim was to conduct a comprehensive proof of concept study comparing the recently developed BIT framework with the RE-AIM framework using an effectiveness-implementation trial to assess the domains, to help test their empirical validity, and to guide future research in the field.

Aims

The primary aim of this study was to compare the RE-AIM and BIT frameworks with data collected from a hybrid effectiveness-implementation randomized controlled trial (RCT) evaluating a digital health intervention for youth with chronic pain called web-based management of adolescent pain (WebMAP Mobile, WMM; Trial Registration: ClinicalTrials.gov NCT03332563) [12]. Specifically, using this real-world trial, we aim to (1) compare and assess the strengths, limitations, and barriers of the BIT and RE-AIM frameworks and (2) make recommendations on how to design future studies to collect the data needed to test the different domains.

Methods

WMM Study

The WMM study is a hybrid effectiveness-implementation cluster RCT testing a mobile health intervention for self-management of adolescent chronic pain. In an effort to deliver evidence-based psychological interventions to adolescents with chronic pain, who have limited access to specialized pain clinics, a smartphone app was developed based on a well-validated web-based intervention: WebMAP; refer to the study by Palermo for details [13]). The mobile app has 6 main modules addressing the following: pain education; stress, emotions, and thoughts; relaxation and imagery; lifestyle and school interventions; staying active; and maintenance and relapse prevention and 2 supplemental modules that are assigned at baseline based on screening for sleep and mood problems.

Parents can access a related web-based cognitive behavioral intervention (WebMAP parent program) that contains 8 modules, which focus on education about chronic pain, recognizing stress and negative emotions, operant strategies (2 modules), modeling, sleep hygiene and lifestyle, communication, and relapse prevention [13]. We have published the clinical trial protocol for this trial [14] and have published the main outcome paper describing the efficacy of the WMM program among youth receiving the intervention versus those receiving usual care in pain and specialty clinics [12].

Methods, Procedures, and Participants

For the WMM RCT, a stepped wedge design was employed [14] in which pediatric pain or specialty (eg, gastroenterology) clinics were randomized into 4 waves to have access to the intervention. A total of 8 clinics across the United States participated (5 pain clinics and 3 specialty gastroenterology clinics). Each clinic began the trial in the usual care condition, during which all youth received usual care alone. Over the subsequent 8-month period, clinics were randomly assigned to begin the intervention period (2 clinics per wave) so that all the clinics ended being exposed to the intervention, allowing a period to test for maintenance. A total of 143 youth (aged between 10 and 17 years) with chronic pain and a caregiver participated in the study, with 73 assigned to the intervention group and 70 to the usual care group.

Inclusion criteria for the main trial were purposefully broad to enhance external validity and included the following: (1) being aged between 10 and 17 years, (2) having chronic pain defined as pain present for at least 3 months, and (3) child having access to a smartphone (iOS or Android) and participating parent having access to a web-enabled device. Exclusion criteria included the following: (1) non-English speaking child or parent, (2) presently in a psychiatric crisis (eg, recent inpatient admission or suicide attempt), and (3) inability to read at the fifth-grade level per parent report. No physical or other mental health comorbidities were excluded.

Measures

Provider surveys, parent and child surveys, clinic data, and administrative data were collected to assess implementation outcomes. Parents reported on sociodemographic characteristics (using a background form), presence of disease-related pain, duration of pain condition, and medication history. Adolescents reported on usual pain intensity (with an 11-point numerical rating scale), pain-related disability (using the Child Activity Limitations Interview [CALI-9] [15]), and patient global impression of change (PGIC, with a single-item question: “Since the start of the study my overall status is...” [1=“No change (or condition has gotten worse)” to 7=“A great deal better, and a considerable improvement that has made all the difference.”]).

Assessment of the Domains in the RE-AIM Framework

The RE-AIM framework has been used in different ways depending on the type of study and associated goals [9]. In this section, the metrics used to assess each of the domains and the sources used to retrieve the information for the WMM study are described.

Reach

Reach assesses participation in the study or intervention. It was defined as the percentage of patients giving consent to participate in the study out of the eligible patients referred by the providers at the clinics. This metric was calculated individually for each of the 8 clinics and averaged across all clinics. In addition, the planned sample size was compared with the final sample that was reached. As a goal of this study was to include a real-world population, the percentage of participants with comorbidities was tracked. Finally, acceptability of the treatment by the patients in the treatment group was assessed as a way to determine potential future reach (ie, if participants found treatment acceptable, it would be more likely to reach future patients). The Treatment Evaluation Inventory (TEI), a treatment acceptability measure used by our group in other trials [13], was administered online using Research Electronic Data Capture (REDCap) [16], a secure web-based survey app.

Effectiveness

Effectiveness was tested as the change from baseline to posttreatment and 3-month follow-up in patient-reported symptoms (pain and disability) using an intention-to-treat analysis and linear mixed effects (LME) regression models. Changes in PGIC were analyzed with one-way analyses of covariance. The effects of treatment engagement on treatment responses were also examined using LME regression models. The source for the effectiveness outcome measures was a battery of psychometrically sound questionnaires (described in the *Measures* section) administered online. The measures were completed independently by adolescents and their parents at the 3 time points (see WMM protocol [14] for more details).

Adoption

Adoption focuses on the delivery settings (ie, clinics) involved in the implementation of the intervention. It was defined as (1) the number of clinics agreeing to participate out of the clinics invited and (2) the percentage of clinics referring patients to the study out of the clinics agreeing to participate in the study.

Implementation

Implementation assesses the extent to which the intervention was delivered as intended. It was assessed at both the individual (ie, user and participant) and organization (ie, clinic) levels, including the cost of delivery and fidelity and consistency with how the intervention is delivered. At the *individual level*, engagement with the intervention (ie, using it as it was intended) was computed as the percentage of participants in the treatment group completing at least one module of the intervention, adherence was computed as the percentage of participants completing at least four modules, and symptom self-monitoring was also assessed as the average number of days participants registered their symptoms using the app during the treatment period. At the organization level, implementation was evaluated using a 6-item survey that assessed the attitudes of the providers of the different clinics (eg, “I think my patients would benefit from this app”) and by also assessing whether the actual costs of developing and implementing the intervention were similar to the budgeted costs or exceeded them. To test implementation at the *individual level*, app usage data, which were automatically

tracked and stored, were used to determine the number of modules completed per participant. To test implementation at the organization level, a web-based survey was administered to the providers using REDCap. The projected budget was also compared with the final expense report.

Maintenance

Maintenance was assessed at both the individual and organization levels. The *individual level* was evaluated by assessing the symptoms at the 3-month follow-up and tracking the number of new patients starting to use the app during the maintenance period. The organization level maintenance was computed as (1) the percentage of clinics agreeing to continue referring patients to the intervention (ie, agreeing to continue to use the app in the clinic) after the study had concluded, (2) the percentage of clinics actually providing referrals of their patients to the app, and (3) the attitude of the providers' willingness to continue prescribing the app (ie, "I will encourage my patients to use the app after the study is over"). For the *individual* and *organizational levels*, data from the web-based surveys and administrative tracking were used, respectively.

Assessment of the Domains in the BIT Framework

In this section, the metrics chosen to assess each domain of the BIT [10] and the sources used to retrieve that information for the WMM study are described.

Acceptability

Acceptability is defined as the perception of the treatment as useful or satisfactory. It was assessed at both the individual and organization levels. The *individual level* was evaluated using the TEI. Specifically, we included the percentage of participants in the treatment group above the TEI cutoff total score (ie, 27) for a moderately acceptable treatment. The organization level was assessed using a provider-completed questionnaire that assessed attitudes, barriers, and facilitators to recommending the app. All information was collected using REDCap.

Adoption

Adoption is defined as the initiation of use of the intervention. This was measured by assessing the following: (1) the percentage of participants downloading the app on their phones, (2) the percentage of participants using the app after their first log-in, and (3) the percentage of participants completing at least one module. These data were extracted from the app server database.

Appropriateness

Appropriateness is the perceived relevance of fit of the intervention within a context, its compatibility with practice, and its usability. At the *individual level*, appropriateness was assessed with adolescents' ratings on a measure of satisfaction completed posttreatment about the appearance, navigation, theme, and content of the app.

Feasibility

Feasibility is defined as the extent to which the intervention can be successfully used in a specific context at the individual and organization levels. At the *individual level*, we collected the number of technical issues reported during the study period. At the organization level, we collected the following: (1) the number of clinics agreeing to participate out of the clinics invited, (2) number of clinics referring patients out of the participating clinics, and (3) posttreatment provider feedback. Data on technical issues were retrieved from the administrative tracking system collected by the study personnel.

Fidelity

Fidelity is defined as the intended use versus the actual use of the intervention, or the adherence to it. We used the following: (1) the percentage of participants completing the treatment (at least four modules) and (2) the number of days tracking their symptoms in the app. Website and app back-end data were used as sources for this information.

Implementation Costs

Implementation costs can include any expenses related to the app or web-based program development (eg, researchers' and developers' salaries) or the implementation itself (eg, costs of the changes needed to adapt the intervention from a research tool to a stand-alone publicly available app). The actual costs compared with budgeted costs were computed. To assess this domain, budgets and administrative databases' tracking expenses were used.

Penetration

Penetration is the integration of the practice (the intervention, in this case) within the service or clinic. To test penetration at the *individual level*, the number of adolescents using the app during the maintenance period (ie, the period after study enrollment had finished and clinicians could refer patients to the app) was calculated. To test this domain at the organization level, how many new users were referred from each clinic was calculated. This metric was retrieved from the back-end data and from the study administrative tracking data.

Sustainability

Finally, sustainability is defined as the extent to which the practice is maintained, its ongoing use. This was assessed in 2 different ways: (1) by calculating the number of clinics agreeing to participate in the maintenance period and (2) the percentage of clinics making referrals. This was assessed using administrative tracking data, back-end data, and budget information.

Results

Overview

In this section, the outcomes for the different domains of the RE-AIM and BIT frameworks are reported. Detailed results of the RE-AIM and BIT domains can be found in Tables 1 and 2, respectively.

Table 1. Summary of implementation outcomes using the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework.

Domains	Metrices	Results	Sources
Reach			
User level	Final sample out of planned sample	n=143/120 (119%)	Administrative tracking data
User level	Consents out of eligible referred children	• Total N=143/181 (79%) • Clinic 1=4/4 (100%) • Clinic 2=10/15 (67%) • Clinic 3=4/5 (80%) • Clinic 4=6/7 (86%) • Clinic 5=15/17 (88%) • Clinic 6=45/55 (82%) • Clinic 7=15/20 (75%) • Clinic 8=44/68 (76%)	Administrative tracking data
User level	TEI[a] mean score and percentage above 27, the moderate acceptability cutoff	Mean 30.7, 86% moderate-to-high acceptability	Patient survey
Effectiveness			
User level	Change in treatment outcomes	Similar change in pain-related disability in both groups. Greater engagement associated with greater improvement in pain-related disability (Cohen *d*=-0.38 for high engagers and *d*=0.27 for low engagers). Greater improvement in global impression of change in the intervention group compared with the control group (*d*=0.54)	Patient survey
Adoption			
Organization level	Percentage of invited clinics agreeing to participate	All clinics agreed (100%)	Administrative tracking data
Organization level	Percentage of participating clinics referring patients	All clinics referred patients (100%)	Administrative tracking data
Implementation			
User level	1 module (engagement)	N=54 (74%)	Back-end data
User level	4+ modules (adherence)	N=29 (40%)	Back-end data
User level	Number of days self-monitoring pretreatment to posttreatment	Mean 30.5 (SD 29.4); median=19; range 2-56	Back-end data
Organization level	Provider attitudes toward the app (1 “Strongly disagree” to 5 “Strongly agree”)	• Helpful to provide CBT[b]: mean 4.6 • Patients would benefit: mean 4.6 • Improves quality of care: mean 4.5 • Better use of resources: mean 4.4 • Fills an important need: mean 4.3	Provider survey
Organization level	Actual costs compared with projected costs	The original budget was exceeded by 7%	Budget data
Maintenance			
User level	New patients using the app during the maintenance period and clinic they were referred from	• Total N=56 • Clinic 1=7 • Clinic 2=0 • Clinic 3=3 • Clinic 4=6 • Clinic 5=26 • Clinic 6=3 • Clinic 7=6 • Clinic 8=5	Administrative tracking data; app back-end data

Domains	Metrices	Results	Sources
Organization level	Percentage of clinics agreeing to continue making referrals	All clinics agreed (100%)	Administrative tracking data
Organization level	Percentage of clinics making referrals	7/8 clinics (88%) made referrals	Administrative tracking data
Organization level	Providers: “I will encourage my patients to use the app after the study is over”	92% agreed or strongly agreed with the item	Provider survey

[a]TEI: Treatment Evaluation Inventory.

[b]CBT: cognitive behavioral therapy.

Table 2. Summary of implementation using the Behavior Interventions using Technology framework.

Domains	Metrices	Results	Sources
Acceptability			
User level	TEI[a] mean score and percentage above the acceptability cutoff (>27)	Mean 30.7, 86% moderate-to-high acceptability	Patient survey
Organization level	Provider attitudes toward the app (1 "Strongly disagree" to 5 "Strongly agree")	• Helpful to provide CBT[b]: mean=4.6 • My patients would benefit: mean 4.6 • Improves the quality of care: mean 4.5 • Better use of resources: mean 4.4 • Fills in an important need: mean 4.3	Provider survey
Adoption			
User level	Percentage of participants who downloaded the app	n=68/73 (93%)	Back-end data
User level	Percentage of participants who used WMM[c] after first log-in	n=68 (100%)	Back-end data
User level	Percentage of participants who completed ≥1 module	n=54 (74%)	Back-end data
Appropriateness			
User level	Score on app perceptions (0 "Did not like it" to 5 "Liked it very much")	• Appearance: mean 3.6 • Navigation: mean 3.9 • Theme: mean 3.7 • Content: mean 3.3	Patient survey
Feasibility			
Organization level	Percentage of clinics agreeing to continue making referrals	All clinics agreed (100%)	Administrative tracking data
Organization level	Percentage of clinics making referrals	7/8 (88%)	Administrative tracking data
Organization level	Final sample out of planned sample	n=143/120 (119%)	Administrative tracking data
User level	Number of technical issues reported or complaints	n=0 (0%)	Administrative tracking data
User level	Participants comments	Not enough space to download the app, n=2 (3%)	Patient survey
Organization level	Providers comments	It was easy to refer patients; WMM is something useful that can be integrated in the practice	Provider survey
Fidelity			
User level	Number of days tracking symptoms	Mean 30.5 (SD 29.4); median=19; range 2-56	Back-end data
User level	Number of participants completing the treatment	n=29 (40%)	Back-end data
Implementation costs			
Organization level	App development costs	As planned	Budgets
Organization level	Making WMM publicly available	Exceed budget by 7%	Budgets
Penetration			
Organization level	New patients using the app during the referral period and clinic they were referred from	• Total N=56 • Clinic 1=7 • Clinic 2=0 • Clinic 3=3 • Clinic 4=6 • Clinic 5=26 • Clinic 6=3 • Clinic 7=6 • Clinic 8=5	Administrative tracking data; back-end data
Sustainability			
Organization level	Referrals made	100% of the clinics agreed; 88% kept referring	Administrative tracking data

Domains	Metrices	Results	Sources
Organization level	"I will encourage my patients to use the app after the study is over"	92% agreed or strongly agreed with the item	Provider survey

[a]TEI: Treatment Evaluation Inventory.

[b]CBT: cognitive behavioral therapy.

[c]WMM: WebMAP Mobile.

Outcomes of the Domains in the RE-AIM Framework

Reach

Originally, we planned to enroll 120 adolescents and parent dyads in a 1-year period, but due to high participation rates, we were able to slightly exceed enrollment by enrolling 143 participants (ie, the final number of participants enrolled was +19% of the planned one). The percentage of consented participants out of the eligible referrals was 79% on average and ranged from 66.7% to 100% for individual clinics. The treatment was considered acceptable by 85.7% of the participants. A total of 20 adolescents (14.3%) had disease-related pain, that is, chronic pain secondary to a comorbid disease, reflecting the inclusion of a heterogeneous sample.

Effectiveness

Effectiveness analyses showed that pain-related disability (the main outcome) and intensity decreased in both intervention and control groups at a similar rate from baseline to posttreatment and follow-up. Adolescents in the intervention group reported significantly greater perception of improvement on the PGIC compared with youth in the control group (Cohen d=0.54; P<.001). Participants with higher engagement with the WMM (1 SD above the mean) reported significantly greater improvements in pain and pain-related disability from pretreatment to 3-month follow-up, respectively (P=.01 and P=.02; see study by Palermo et al [12] for details).

Adoption

All the clinics invited to participate in the study agreed to participate, and all clinics referred participants to the study (referrals ranging from 6 to 80 participants per clinic).

Implementation

Implementation at the *individual level* was variable. A total of 93% of participants randomized to the intervention group downloaded the app, 74% engaged with the intervention (ie, completed at least one module), and 40% were adherent to it (ie, completed at least four modules). The average number of days participants registered their symptoms was 19.7, considering that the posttreatment assessment was conducted about 8 weeks after the baseline assessment; on average, participants registered symptoms on 35.2% of days during the treatment period.

At the *organization level*, providers' (n=27, 47% of provider sample) average scores on attitudes toward the WMM ranged from 4.3 to 4.6 out of 5, indicating strong agreement with the items related to (1) the usefulness and benefits of the WMM to patients, specifically 100% of participants agreed or strongly agreed, and (2) the cost-effectiveness of implementing the app within clinics, specifically 93% agreed or strongly agreed. The main outcomes paper includes the details [12]. Regarding budgets, the actual costs exceeded the original budget by 7% to perform a public release of the app.

Maintenance

At the *individual level*, symptoms at the 3-month follow-up are presented in the *Effectiveness* section. A total of 56 new patients downloaded the app during the maintenance period (with clinics referring from 0 to 26 patients each). At the organization level, 100% of the clinics agreed to continue referring patients to the intervention after the study and 88% (7/8 clinics) referred their patients to use the app during the maintenance period. Using app data tracking, we were also able to determine that in the 6-month implementation period, 56 adolescents had downloaded the app on their phones and opened it. Finally, most providers (92%) agreed or strongly agreed with the item: "I will encourage my patients to use WMM after the study is over." Some additional comments on the survey indicated potential barriers and facilitators of implementation. The main barrier reported was that the treatment content did not seem relevant to some families. The main reported facilitators were the following: (1) it was easy to make referrals and (2) the app was perceived as something useful that they could integrate into their practice.

Outcomes of the Domains in the BIT Framework

Acceptability

Treatment perceptions at the *individual level* on the TEI showed that 86% of adolescents rated the treatment as at least moderately acceptable (mean score of 31). The organization level was assessed in the same way as the *Implementation* outcome of the RE-AIM framework (ie, with the provider-completed questionnaire): all providers agreed or strongly agreed with items regarding the usefulness and benefits of the WMM to patients, and 93% of the providers agreed with the items about the cost-effectiveness of implementing the app within clinics.

Adoption

The percentage of participants randomized to active intervention who downloaded the app on their phones was high (68/73, 94%). All participants (100%) used the intervention after their first log-in; 74% of participants completed at least one module.

Appropriateness

At the *individual level*, adolescents' responses to questions about the appearance, navigation, theme, and content of the intervention were scored 3.3 to 3.9 out of 5, on average, on a scale of 0 "Did not like it" to 5 "Liked it very much."

Feasibility

Feasibility at the *individuallevel* was high: no technical issues were reported. A total of 2 adolescents (2.7%) reported not having enough space on their phones to download the app, which prevented them from using it. At the *organization level*, all clinics agreed to participate and referred participants to the study. The initial enrollment goal was exceeded by 19%. Posttreatment qualitative feedback was also positive: providers stated that it was easy to refer patients to the app and that it was a useful resource that they could integrate into their practice.

Fidelity

A total of 40% of the participants completed the treatment. Adolescents tracked their symptoms for 31 days on average (median 19) during the treatment, but this was highly variable (SD 29.4).

Implementation Costs

Implementing the app involved hiring a software development company and covering costs for code writing, graphic design, app testing, and app store fees. The original budget was US $97,500 to develop a custom app for each iOS and Android system. In addition, at the end of the study period, we decided to make the app publicly available and free of cost for the user. This last step involved some modifications (eg, eliminating the need for a password-protected log-in page and changing the privacy policy), new testing, and extending ongoing app maintenance, which exceeded our planned budget by 7% (US $12,000). This 7% was covered with donor funds and in-kind support from the app development company.

Penetration

Penetration resulted in 56 new users accessing the app during the maintenance period (that closed when the app was made publicly available in the stores), with the range per clinic being 0 to 26.

Sustainability

Finally, as a measure of sustainability, all clinics agreed to participate in the maintenance period, and we determined that 88% of them made referrals. At the end of the study, the app was made publicly available on the app stores (for Android and iOS users) of English-speaking countries free of cost for the user. Over the 2 years, the app will be maintained by the developers. As a final step for dissemination, a press release about the availability of the app was coordinated with the communication department of our institution.

Comparison Between the Barriers and Facilitators Identified by the RE-AIM and BIT Frameworks

In Table 3, we provide a summary of our experience assessing each domain using both frameworks, including facilitators and barriers to use. We also provide recommendations on how to overcome these barriers.

Table 3. Lessons learned: barriers and facilitators to assess the domains of each framework and recommendations for future studies.

Framework and domains, Barriers and facilitators	Recommendations and considerations
RE-AIM[a]	
Reach	
Organization level (number of consents obtained out of eligible referrals received), and overall N were easy to collect because our trial involved user level referral and tracking.	CONSORT (The Consolidated Standards of Reporting Trials) flow diagram will provide these data. For nonresearch contexts, tracking the users approached, interested, and participating would be needed.
Availability of a standardized acceptability measure with a threshold for defining moderate acceptability facilitated this measurement.	If data from previous studies with more traditional designs or epidemiologic data are available, other metrics such as changes on comorbidities or representativeness of participants can be included. If available, the distance between participants homes and the clinic can be another way to measure Reach.
Effectiveness	
Effectiveness data from primary and secondary outcome measures was facilitated by web-based survey administration.	This domain is almost always assessed in research studies. However, it might be challenging to assess in nonresearch contexts. Low-demand approaches, such as voluntary web-based surveys could help gather information.
Adoption	
The number of centers willing to participate and participating was easily assessed with administrative tracking. We originally planned to assess the number of referrals out of the eligible participants. However, clinics were unable to provide information on the age range of their patients or how many had chronic pain; thus, the number of eligible patients is unknown.	When defining adoption, it would be key to understand availability of information required to assess this domain. If it is not, alternative metrics should be planned and collected from the beginning of the intervention.
Implementation	
Being able to access and interpret back-end data for the app required working with the developers and having a data analyst transforming the databases.	It is important to plan the human resources needed and budget costs in advance.
Creating the web-based survey for the providers was an efficient way to collect data from multiple clinics located across the United States. Only half of providers participated in the survey, which was a barrier to understanding their perceptions.	Web-based surveys could be a cost-effective tool to assess this metric. However, they should be brief to avoid participant burden. Qualitative feedback can be collected in face-to-face interviews or open-ended surveys but requires additional cost to analyze and interpret.
Maintenance	
We originally planned to track number of referral flyers given per clinic, but providers did not use the flyers consistently. Thus, we used alternative metrics: using app data tracking to understand number of downloads and times the app was used.	Having access to the chosen metrics should be ensured from the beginning. Ideally, objective and subjective measures (eg, asking participants if they are still using the intervention and being able to track usage with back-end data) should be collected.
BIT[b]	
Acceptability	
Collecting web-based acceptability feedback facilitated this assessment as it was efficient and low burden. However, we were limited to quantitative data to understand perceptions.	Web-based surveys are recommended to assess acceptability, with the same considerations that the rich detail of user perceptions may not be possible to gather in this manner.
Adoption	
Information retrieved from the app needed several steps of cleaning and restructuring databases (and the involvement of personnel with 3 different profiles and skill sets: engineers, data manager, and research scientist) before being interpretable. The costs of this process may be a barrier if unplanned for.	Working with engineers and developers from the creation of the intervention and having a dialog about the information (metrics) needed is key to ensure that adoption can be properly assessed. Budget can be a barrier because it is often expensive to obtain some metrics in a "user friendly" way (eg, the systems may provide information in a way that is difficult to understand by the lay user).
Appropriateness	
A closed-ended patient survey allowed to collect perceptions about WMM[c] appearance, navigation, theme, and content.	For appropriateness, it would be ideal to be able to complement web-based surveys with additional qualitative assessments if costs permit their inclusion.
We were unable to capture qualitative perceptions or follow-up on the questions (they were anonymous).	
Feasibility	

Framework and domains, Barriers and facilitators	Recommendations and considerations
Technical issues and complaints were carefully tracked but it is possible that additional problems were unreported.	Participants should be able to easily report technical problems to maximize the chances of reporting. A phone number or contact email (that is attended) should be provided to the participants and included in the app or website.
Fidelity	
Resources needed to use back-end data also apply for this metric.	Defining what is "intended" and "actual" use beforehand would allow decisions to be made on the metrics to use and to plan on resources if back-end data are needed to be retrieved.
Implementation costs	
We decided to compare the planned budget and the real expenses as a way to determine efficiency of the resources.	Deciding whether implementation costs were adequate can be difficult without having a reference and is study specific.
Penetration	
At the *organization level*, we planned to assess the ratio of providers and users eligible per clinic out of the total number of providers in the clinic and total users; however, the clinics were unable to provide those numbers, and we assessed this domain by calculating how many new users were referred by each clinic.	Deciding how to assess the extent to which the practice is integrated within the system can be challenging. If unknown, a pilot study could help inform what information is feasible to obtain from clinics or organizations where the intervention is being implemented.
Sustainability	
Using back-end data and checking the activation codes used, we were able to determine the percentage of clinics making referrals during the maintenance period.	Ongoing use of the intervention after the study ends can be a challenging domain to assess, because contact with the participating centers should be minimal. Collecting information in a passive way (eg, tracking use with back-end data) or with brief web-based surveys would be preferred.

[a]RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance.

[b]BIT: Behavior Interventions using Technology.

[c]WMM: WebMAP Mobile.

Discussion

Principal Findings

This is the first time a real-world digital health intervention is used as a proof of concept to test all the domains of both the RE-AIM and BIT frameworks, demonstrating a full example of the process followed to collect information and assess the different domains from study inception to follow-up, and to compare both frameworks.

Using the RE-AIM framework, WMM [12] showed excellent *Reach*, as indicated by the high enrollment and participation rate. *Effectiveness* was limited, with only a global impression of change showing significantly greater improvements in the treatment group, although higher treatment engagement was associated with reduced pain and disability. *Adoption* was also excellent, as evidenced by the clinic participation. *Implementation* at the user level was moderate, with 93% of participants downloading the app and 40% being adherent (completing at least four modules). At the *organization level*, providers had positive attitudes toward WMM. Regarding budgets, a small increase in funds was needed to make WMM publicly available. Finally, *maintenance* at the user level was evidenced by new patients downloading the app during the maintenance period. At the *organization level*, maintenance was also excellent, with most of the clinics continuing to make new referrals.

Using the BIT framework, the *acceptability* of the WMM was evidenced by most of the users considering the treatment acceptable and the providers showing positive attitudes toward the app. *Adoption*, according to the BIT, was shown by most participants downloading the WMM and indicating its use. *Appropriateness* at the user level showed that the WMM had a good fit in this context. The *feasibility* of implementing the intervention was high, with most participants being able to download the app and not reporting any issues. At the *organization level*, the referrals exceeded the original goal; the providers mentioned that referrals were easy to make and that the WMM could be easily integrated in their practice. In addition, all clinics agreed to continue using the app with their patients, and most of them did so. *Fidelity* was moderate, given that less than half of the participants finished the treatment and only monitored symptoms on some of the days. *Implementation costs* regarding app development were as planned, but taking the extra step of making the WMM publicly available represented a small increase in costs. *Penetration* was also excellent, with new patients coming from most clinics during the implementation period. Finally, *sustainability* seems promising because most of the clinics agreed to keep using the app and did so with new patients.

Comparing the 2 frameworks, we can observe some overlap between them, but important differences exist. Specifically, RE-AIM has a strong emphasis on *Reach* and *Effectiveness* and assesses other aspects in a more limited way. In fact, an equation to calculate *Individual Impact* has been proposed [17]. It can be computed as the sum across target behaviors of the Reach domain×average of individual change at long-term follow-up (Effectiveness). Moreover, the organization level impact can

be computed as Adoption×Implementation. Maintenance, however, remains as an isolated domain that is explained more superficially in the RE-AIM framework. In contrast, within the BIT framework, the concept of *maintenance* is assessed by 2 domains: penetration (integration of the practice in the context) and sustainability (the extent to which that integration is maintained). The BIT framework is more detailed in the sense that it encompasses a larger number of domains that are pertinent to the technology field and allows for testing technology-focused aspects such as usability and feasibility (providing more room to integrate qualitative user feedback), and it also considers organizational metrics and costs (which may be important for intervention and budget planning). However, the BIT fails to incorporate the effectiveness dimension, which researchers need to evaluate separately, and there are no proposals on how to integrate the various domains. In general, both frameworks provide a comprehensive assessment of the implementation of the WMM study, which was estimated to be generally successful.

Although there were several different implementation components assessed by each framework, the choice to use one over the other might depend on different contexts (either because the information required is more accessible or because the information retrieved would be more relevant for specific implementation goals). It is also important to note that most often, every dimension will not be evaluated in one project, and choices will need to be made to prioritize the dimension most central to the research question. In addition to RE-AIM and BIT, there are many other frameworks and theories for the dissemination and implementation of evidence-based interventions (the RE-AIM workgroup cites 167 different theories). The advantage of using the same implementation framework is for comparison between eHealth and mobile health interventions and to develop a common language and criteria for implementation outcomes. Our experience suggests that the RE-AIM framework might be more appropriate and easier to use for research-based interventions at either planning or evaluation stages of a project and has the advantage of being well studied over several decades with many published evaluations. The BIT framework, however, is flexible to use to study the implementation of interventions with or without an effectiveness evaluation and would also be appropriate during the planning stage of a project. The primary advantage of BIT is that it evaluates the technical and cost aspects of implementation more comprehensively and has a more nuanced focus on sustainability. However, the goals of the implementation study should be the key factors to help guide the choice of the most appropriate framework.

Both frameworks provided a structure for assessing the different domains of implementation of a digital intervention by providing guidance on what to measure (numbers to track), how to measure (suggesting some standardized metrics and cutoff thresholds, using web-based tools), and when to measure (baseline and follow-up). This would be appropriate at the planning stage to consider how to budget accordingly and at the evaluation stage to identify strengths and areas of improvement after the intervention has been implemented. The main barriers emerging for both frameworks are the potential limitations in time and budget and the unavailability of certain information depending on the context of the intervention, low participation rates in surveys, and the need of having trained personnel to interpret the retrieved data.

A number of key recommendations can be made based on this proof of concept: (1) planning ahead regarding available budget and time is crucial because assessing implementation requires additional resources; (2) communication with all members of the team and participating centers (eg, hospitals and schools) from the beginning is key to ensure access to information and to budget for the implementation study; (3) using approaches that are efficient (ie, require little time or resources) such as web-based surveys should be prioritized, but complementing with more intensive techniques (such as phone or video interviews) may be needed to capture more detailed information in certain domains; (4) conducting multisite real-world trials entails conducting research in an uncontrolled environment that can be unpredictable, and when possible, use several metrics to assess each domain, that way, if one fails, back-up metrics would be available; and (5) it may be expensive or time consuming to assess all the domains in a given framework, and choosing the domains that are relevant for the study goals before starting the trial and assessing feasibility for data collection (eg, asking the participating centers beforehand about the information they have available and are willing to share and asking developers about availability of back-end data for the lay user) would maximize chances of successful data collection and help minimize costs.

Limitations

The limitations of this study should be considered when interpreting the findings. We were not able to assess all the domains at the *individual* and *organizational levels* for both frameworks. It is possible that studies conducted in different contexts can do so and find new barriers or facilitators not included here. Despite being a multicenter geographically diverse study, participants belonged to a specific type of population such as adolescents with chronic pain; results may be different when conducting studies with adults or with digital interventions for other health problems. Finally, the demographic characteristics of our sample of participants were predominantly White, female, and had a high socioeconomic status, which, despite being representative of adolescents attending pain clinics, may not be generalizable to samples with greater sociodemographic diversity.

Comparison With Previous Work

To date, the RE-AIM framework has been used to assess the implementation of nondigital interventions only [9], so we do not have other digital health examples for comparison. The BIT framework has been partially tested using different intervention studies to show examples of each domain [10]; consequently, all the examples represented cases of success in assessing these domains. In previous studies, there was no specific guidance to allow precision in measuring each domain or how to problem-solve any potential issues with measurement.

As mentioned before, there are many other dissemination and implementation frameworks to assess digital interventions, such

as the Non-adoption, Abandonment, Scale-up, Spread, Sustainability (NASSS) framework [18]. However, the focus of this framework is very different, being more centered on macro systems and organizations as the defining environments to test, and it is not a good fit to test implementation at the user (individual) level, missing relevant implementation information.

Future studies might consider using the structure presented here (ie, testing every domain in the frameworks using a single intervention) but including different populations and organization settings to test generalizability. It is important to test implementation in other countries that have different health care systems, because new barriers or facilitators may emerge.

Conclusions

The RE-AIM and BIT frameworks were tested with a proof of concept study, showing both to be a useful fit for assessing implementation with different strengths and weaknesses. Some recommendations for choosing a framework to assess implementation in digital health interventions were provided. Strategies to overcome the main barriers encountered when assessing the frameworks were suggested.

Acknowledgments

This research was supported by the American Pain Society and Pfizer Independent Grants for Learning and Change (grant ID #: 27971161, principal investigator: Palermo). The study sponsor and funders had no role in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication. 2Morrow.inc was responsible for software development. The authors thank the participating clinics for their involvement in the study: Seattle Children's Hospital (gastroenterology clinic and pain clinic), Children's Mercy Medical Center (abdominal pain clinic and pain clinic), C.S. Mott Children's Hospital (pain clinic), Nationwide Children's Hospital (pain clinic), and Connecticut Children's Medical Center (gastroenterology clinic and pain clinic).

Authors' Contributions

RV drafted the first version of the manuscript. All the authors made substantial and original contributions to the content and reviewed and approved the final version of the manuscript.

Conflicts of Interest

Some authors created WMM, the internet intervention used as an example for the proof of concept; however, they do not have any financial interest or benefit from the product.

References

1. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. QJM 2013 Dec;106(12):1067-1069 [FREE Full text] [doi: 10.1093/qjmed/hct203] [Medline: 24106313]
2. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Sainz-De-Abajo B. Comparison of mobile apps for the leading causes of death among different income zones: a review of the literature and app stores. JMIR Mhealth Uhealth 2014 Jan 9;2(1):e1 [FREE Full text] [doi: 10.2196/mhealth.2779] [Medline: 25099695]
3. Fisher E, Law E, Palermo T, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. Cochrane Database Syst Rev 2014 May 14;2014:- [FREE Full text] [doi: 10.1002/14651858.CD011118] [Medline: 25221436]
4. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. A systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
5. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
6. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
7. Jahns JR. mHealth App Developer Economics. Research2Guidance. 2014. URL: https://research2guidance.com/mhealth-app-developer-economics/ [accessed 2020-08-10]
8. Dzewaltowski DA, Glasgow RE, Klesges LM, Estabrooks PA, Brock E. RE-AIM: evidence-based standards and a web resource to improve translation of research into practice. Ann Behav Med 2004 Oct;28(2):75-80. [doi: 10.1207/s15324796abm2802_1] [Medline: 15454353]
9. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
10. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]

11. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
12. Palermo TM, Vega R, Murray C, Law E, Zhou C. A digital health psychological intervention (WebMAP mobile) for children and adolescents with chronic pain: results of a hybrid effectiveness-implementation stepped wedge cluster randomized trial. Pain 2020 Jul 10:- epub ahead of print. [doi: 10.1097/j.pain.0000000000001994] [Medline: 32658147]
13. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain 2016 Jan;157(1):174-185 [FREE Full text] [doi: 10.1097/j.pain.0000000000000348] [Medline: 26335910]
14. Palermo TM, de la Vega R, Dudeney J, Murray C, Law E. Mobile health intervention for self-management of adolescent chronic pain (WebMAP mobile): Protocol for a hybrid effectiveness-implementation cluster randomized controlled trial. Contemp Clin Trials 2018 Nov;74:55-60 [FREE Full text] [doi: 10.1016/j.cct.2018.10.003] [Medline: 30290276]
15. Holley AL, Zhou C, Wilson AC, Hainsworth K, Palermo TM. The CALI-9: a brief measure for assessing activity limitations in children and adolescents with chronic pain. Pain 2018 Jan;159(1):48-56 [FREE Full text] [doi: 10.1097/j.pain.0000000000001063] [Medline: 28957835]
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
17. Glasgow RE, Klesges LM, Dzewaltowski DA, Estabrooks PA, Vogt TM. Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. Health Educ Res 2006 Oct;21(5):688-694. [doi: 10.1093/her/cyl081] [Medline: 16945984]
18. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]

Abbreviations

BIT: Behavior Interventions using Technology
LME: linear mixed effects
PGIC: patient global impression of change
RCT: randomized controlled trial
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
REDCap: Research Electronic Data Capture
TEI: Treatment Evaluation Inventory
WebMAP: web-based management of adolescent pain
WMM: WebMAP Mobile

Edited by G Eysenbach; submitted 05.05.20; peer-reviewed by J Alvarez Pitti, S Smith; comments to author 29.06.20; revised version received 30.06.20; accepted 22.07.20; published 01.09.20.

Please cite as:
de la Vega R, Ritterband L, Palermo TM
Assessing Digital Health Implementation for a Pediatric Chronic Pain Intervention: Comparing the RE-AIM and BIT Frameworks Against Real-World Trial Data and Recommendations for Future Studies
J Med Internet Res 2020;22(9):e19898
URL: https://www.jmir.org/2020/9/e19898
doi:10.2196/19898
PMID:32870158

Original Paper

Artificial Intelligence and Its Effect on Dermatologists' Accuracy in Dermoscopic Melanoma Image Classification: Web-Based Survey Study

Roman C Maron[1], MSc; Jochen S Utikal[2,3], MD; Achim Hekler[1], MSc; Axel Hauschild[4], MD; Elke Sattler[5], MD; Wiebke Sondermann[6], MD; Sebastian Haferkamp[7], MD; Bastian Schilling[8], MD; Markus V Heppt[9], MD; Philipp Jansen[6], MD; Markus Reinholz[5], MD; Cindy Franklin[10], MD; Laurenz Schmitt[11], MD; Daniela Hartmann[5], MD; Eva Krieghoff-Henning[1], PhD; Max Schmitt[1], MSc; Michael Weichenthal[4], MD; Christof von Kalle[12], MD; Stefan Fröhling[13], MD; Titus J Brinker[1], MD

[1]Digital Biomarkers for Oncology Group (DBO), National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany

[2]Skin Cancer Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany

[3]Department of Dermatology, Venereology and Allergology, University Medical Center Mannheim, University of Heidelberg, Heidelberg, Germany

[4]Department of Dermatology, University Hospital Kiel, University of Kiel, Kiel, Germany

[5]Department of Dermatology, University Hospital Munich (LMU), Munich, Germany

[6]Department of Dermatology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

[7]Department of Dermatology, University Hospital Regensburg, Regensburg, Germany

[8]Department of Dermatology, University Hospital Würzburg, University of Würzburg, Würzburg, Germany

[9]Department of Dermatology, University Hospital Erlangen, University of Erlangen, Erlangen, Germany

[10]Department of Dermatology, University Hospital Cologne, Cologne, Germany

[11]Department of Dermatology, University Hospital Aachen, Aachen, Germany

[12]Department of Clinical-Translational Sciences, Charité University Medicine and Berlin Institute of Health (BIH), Berlin, Germany

[13]National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany

Corresponding Author:
Titus J Brinker, MD
Digital Biomarkers for Oncology Group (DBO)
National Center for Tumor Diseases (NCT)
German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 280
Heidelberg, 69120
Germany
Phone: 49 62213219304
Email: titus.brinker@nct-heidelberg.de

Abstract

Background: Early detection of melanoma can be lifesaving but this remains a challenge. Recent diagnostic studies have revealed the superiority of artificial intelligence (AI) in classifying dermoscopic images of melanoma and nevi, concluding that these algorithms should assist a dermatologist's diagnoses.

Objective: The aim of this study was to investigate whether AI support improves the accuracy and overall diagnostic performance of dermatologists in the dichotomous image–based discrimination between melanoma and nevus.

Methods: Twelve board-certified dermatologists were presented disjoint sets of 100 unique dermoscopic images of melanomas and nevi (total of 1200 unique images), and they had to classify the images based on personal experience alone (part I) and with the support of a trained convolutional neural network (CNN, part II). Additionally, dermatologists were asked to rate their confidence in their final decision for each image.

Results: While the mean specificity of the dermatologists based on personal experience alone remained almost unchanged (70.6% vs 72.4%; *P*=.54) with AI support, the mean sensitivity and mean accuracy increased significantly (59.4% vs 74.6%;

P=.003 and 65.0% vs 73.6%; *P*=.002, respectively) with AI support. Out of the 10% (10/94; 95% CI 8.4%-11.8%) of cases where dermatologists were correct and AI was incorrect, dermatologists on average changed to the incorrect answer for 39% (4/10; 95% CI 23.2%-55.6%) of cases. When dermatologists were incorrect and AI was correct (25/94, 27%; 95% CI 24.0%-30.1%), dermatologists changed their answers to the correct answer for 46% (11/25; 95% CI 33.1%-58.4%) of cases. Additionally, the dermatologists' average confidence in their decisions increased when the CNN confirmed their decision and decreased when the CNN disagreed, even when the dermatologists were correct. Reported values are based on the mean of all participants. Whenever absolute values are shown, the denominator and numerator are approximations as every dermatologist ended up rating a varying number of images due to a quality control step.

Conclusions: The findings of our study show that AI support can improve the overall accuracy of the dermatologists in the dichotomous image–based discrimination between melanoma and nevus. This supports the argument for AI-based tools to aid clinicians in skin lesion classification and provides a rationale for studies of such classifiers in real-life settings, wherein clinicians can integrate additional information such as patient age and medical history into their decisions.

(J Med Internet Res 2020;22(9):e18091) doi:10.2196/18091

KEYWORDS

artificial intelligence; machine learning; deep learning; neural network; dermatology; diagnosis; nevi; melanoma; skin neoplasm

Introduction

Melanoma detection and classification is a challenging task but diagnostic accuracy can be enhanced using aids such as dermoscopy, which has improved the examination of pigmented and nonpigmented skin lesions with the naked eye [1-3]. The combination of dermoscopy with reflectance confocal microscopy could further increase the diagnostic accuracy of melanocytic lesions, thereby highlighting the potential of the complementary approaches [4]. A comparatively new strategy falling within the realm of computer-aided diagnosis (CAD) is the use of trained convolutional neural networks (CNNs) to analyze macroscopic images of suspicious lesions. Studies have shown that, within certain limitations and considering a purely image-based setting, artificial intelligence (AI) can achieve on par or superior performance to dermatologists [5-9], thereby highlighting its potential as a decision-support system with immediate clinical implications.

Previous studies have shown that dermatologists perform better as a group [10] or as an ensemble of human and machine [11]. An imputation analysis of the results from the International Skin Imaging Collaboration (ISIC) 2017 challenge showed an increase in the dermatologists' performance when low-confidence decisions were replaced by a classifier's diagnosis [12]. However, in these studies, participants' answers were independent of each other and combined or modified retrospectively. The performance of the dermatologists with live information from CAD systems has to be investigated in more detail.

The CAD system MelaFind, intended for multispectral digital analysis of melanocytic lesions, has been evaluated in experimental settings and has shown to improve dermatologists' sensitivity, albeit at the cost of lower specificity [13,14], which has raised concerns about its benefits [15]. AI-based systems tend to be more balanced in terms of sensitivity and specificity scores and have shown to improve the classification accuracy of nondermatologists by using a content-based image retrieval algorithm [16]. Outside of dermatology, similar studies have shown the benefit of AI [17] while also highlighting the challenges AI assistance faces, as it can be both helpful and misleading at the same time [18]. This study builds on and investigates whether live AI support in the form of a classic CNN architecture is capable of improving the accuracy and the overall diagnostic performance of experts (dermatologists) in the dichotomous image–based discrimination between melanoma and nevus.

Methods

Study Design

This study was conducted from January 10, 2019 (design of study) to September 27, 2019 and was inspired by the study design of Sinz et al [2]. Questionnaires were sent out on June 12, 2019 and were completed by August 12, 2019. Ethical approval was waived by the ethics committee of the University of Heidelberg, Mannheim Faculty of Medicine as all the images were open source and anonymous and all participating dermatologists automatically became part of the study group.

The set of images used for the evaluation consisted of 1200 unique dermoscopic images split into 12 nonoverlapping individual test sets containing 100 images each (50 melanomas and 50 nevi). Each individual set was randomly assigned to exactly one dermatologist who diagnosed his/her set twice—first without AI support (part I) and afterwards with AI support (part II). Images for each part were presented to the participants in 2 separate surveys containing 50 images each (again 50:50 split) to investigate a possible learning effect. Thus, every participant completed 4 surveys (2 surveys for part I and 2 surveys for part II) where images seen in part I were identical to those seen in part II. Both surveys from part I were carried out before any survey from part II. The participants received detailed instructions by email in which the survey structure and the procedure were discussed and the lesion distribution was disclosed. For part II, participants were made aware of the classifier's performance, which was established beforehand on a separate validation set.

Classifier Training

Images were obtained from the ISIC archive [19], with a large fraction of images coming from the HAM10000 dataset [20].

The archive is publicly accessible and contains anonymous dermoscopic images from multiple sources and various camera systems. From the available pigmented lesions, only images showing a biopsy-verified melanoma (n=1633) or nevus (n=3311) were selected (total n=4944).

As each of the 12 individual test sets consisted of 2 subsets (with 50 images each), there were 24 test subsets. A designated training set was constructed for each test subset by removing its images from the total image data, resulting in 24 training sets, each containing 4894 images (ie, 4944 – 50). The class distribution across each training set was unevenly distributed with 1608 melanoma images in contrast to 3286 nevus images. To counter the class imbalance, the set of melanoma images was duplicated. Online data augmentation was used to increase the diversity of the training data and to modify the duplicated images. Before balancing, an evenly balanced validation set was removed from each training set for calibration and validation purposes later. A detailed description of the classifier training is given in Multimedia Appendix 1.

Classifier Integration

For AI support to be used in an effective manner, a confidence measure was displayed in addition to the binary decision (melanoma or nevus), which indicated the classifier's confidence in its decision. The measure was obtained by mapping the classifier's raw output probability for the predicted class into a percentage range from 0% to 100%.

Participants were made aware of the classifier's overall performance by establishing the overall performance of all 24 classifiers on their validation set and by taking the average, which resulted in mean sensitivity, mean specificity, and mean accuracy of 78.0%, 81.0%, and 80.0% respectively.

Electronic Survey and Participants

Fifteen participants were invited by personal invitation via their university email accounts by the principal investigator (TJB) based on previous cooperation. Three had to be excluded because one reported to not have time to do the second survey properly, and 2 did not attempt or finish the survey in the proposed timeframe, leaving 12 dermatologists (all board-certified and coauthors of this publication) from 9 German university hospitals for analysis. The Google Forms closed web-based survey was not anonymous and each participant had to disclose his or her full name and additional metadata. For part I, the dermatologists were shown images of biopsy-verified skin lesions and asked 3 mandatory questions. First, a personal rating of the image quality had to be given (excellent, good, sufficient, poor, other image problems, no image visible). Second, the participant was asked for a diagnosis of the displayed lesion with the options being benign or malignant. Third, the participant was requested to quantify his or her confidence in the decision on a scale of 0 (=very uncertain) to 10 (=very certain). Part II showed identical images and questions and was additionally labeled with the dermatologist's previous response and the diagnosis of the CNN together with its confidence level. For details on the participants, see Multimedia Appendix 1.

Performance and Statistical Analysis

The primary endpoint of this study was whether dermatologists' accuracy (overall proportion of the correct predictions among the total number of examined cases) would increase with AI support with the secondary endpoint investigating how sensitivity and specificity would change (tested for significance using Wilcoxon test). Additional analyses investigated whether dermatologists experienced a learning effect when working with AI support (tested for significance using Pearson's chi-squared test) and how they would respond to this form of computer aid. Statistical significance was considered at $P<.05$, corrected to 0.016 (Bonferroni correction for the primary and secondary endpoint) to account for multiple testing. The details on how survey answers were evaluated are listed in Multimedia Appendix 1. As a pair of classifiers was responsible for scoring the test set of one participant (one classifier for survey 1, the other for survey 2), the results of both classifiers were combined for analysis reasons so that every dermatologist had a respective counterpart. Values shown in the result sections are based on the mean of all the participants. If absolute values are shown, the denominator and numerator are approximated. In theory, every dermatologist rated the exact same number of images, but due to the quality control step, not every rating was counted. Therefore, each dermatologist ended up rating a varying number of images.

Results

Classification Results

In the first part of the study, the dermatologists had an overall mean sensitivity of 59.4% (95% CI 53.3%-65.5%), specificity of 70.6% (95% CI 62.3%-78.9%), and accuracy of 65.0% (95% CI 62.3%-67.6%, see Table 1).

In the second part of the study, where participants could integrate the results of the CNN-based classifier in their decision-making, their overall sensitivity increased significantly to 74.6% (95% CI 69.9%-79.3%; $P=.003$). Their mean specificity also showed a positive trend (72.4%; 95% CI 66.2%-78.6%; $P=.54$), so that the overall accuracy also increased significantly to 73.6% (95% CI 70.9%-76.3%; $P=.002$).

The CNN on its own had an even higher sensitivity (84.7%; 95% CI 81.9%-87.6%), specificity (79.1%; 95% CI 74.8%-83.4%), and accuracy (81.9%; 95% CI 79.7%-84.2%) on the test set, which was comparable to its performance on the validation set. The overall performance of the dermatologists and the CNN are summarized in Table 1 while individual performances are captured in Figure 1 and Figure 2, which show an overview of the various metrics for every participant with and without AI support.

Table 1. Overall mean performance of dermatologists without artificial intelligence (AI) support (-AI) compared to that of dermatologists with AI support (+AI) and AI on its own. Performance is split into 3 categories, that is, sensitivity, specificity, and accuracy.

Performance	AI	Dermatologist -AI	Dermatologist +AI
Sensitivity (95% CI)	84.7% (81.9%-87.6%)	59.4% (53.3%-65.5%)	74.6% (69.9%-79.3%)
Specificity (95% CI)	79.1% (74.8%-83.4%)	70.6% (62.3%-78.9%)	72.4% (66.2%-78.6%)
Accuracy (95% CI)	81.9% (79.7%-84.2%)	65.0% (62.3%-67.6%)	73.6% (70.9%-76.3%)

Figure 1. Combined and individual dermatologists' performance without and with artificial intelligence (AI) support. Every dot represents a single participant. A line between 2 dots connects the participants' metric without AI support to the corresponding metric with AI support. Highlighted dots represent the participants combined. Boxes indicate 25th and 75th percentile while the horizontal line within shows the median (50th percentile). Whiskers indicate the data range (1.5*IQR) where points beyond are considered as outliers.

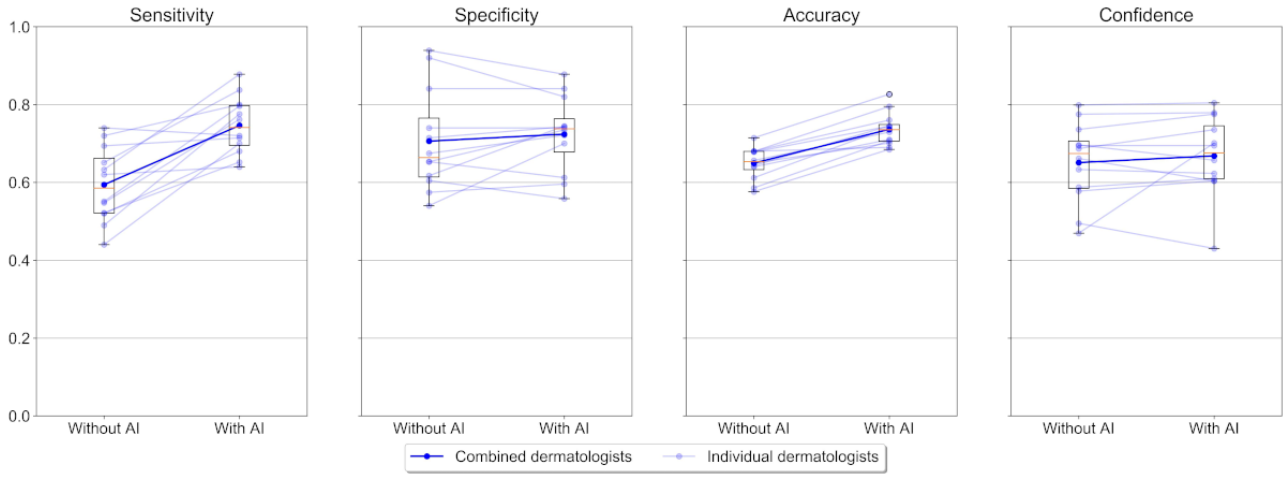
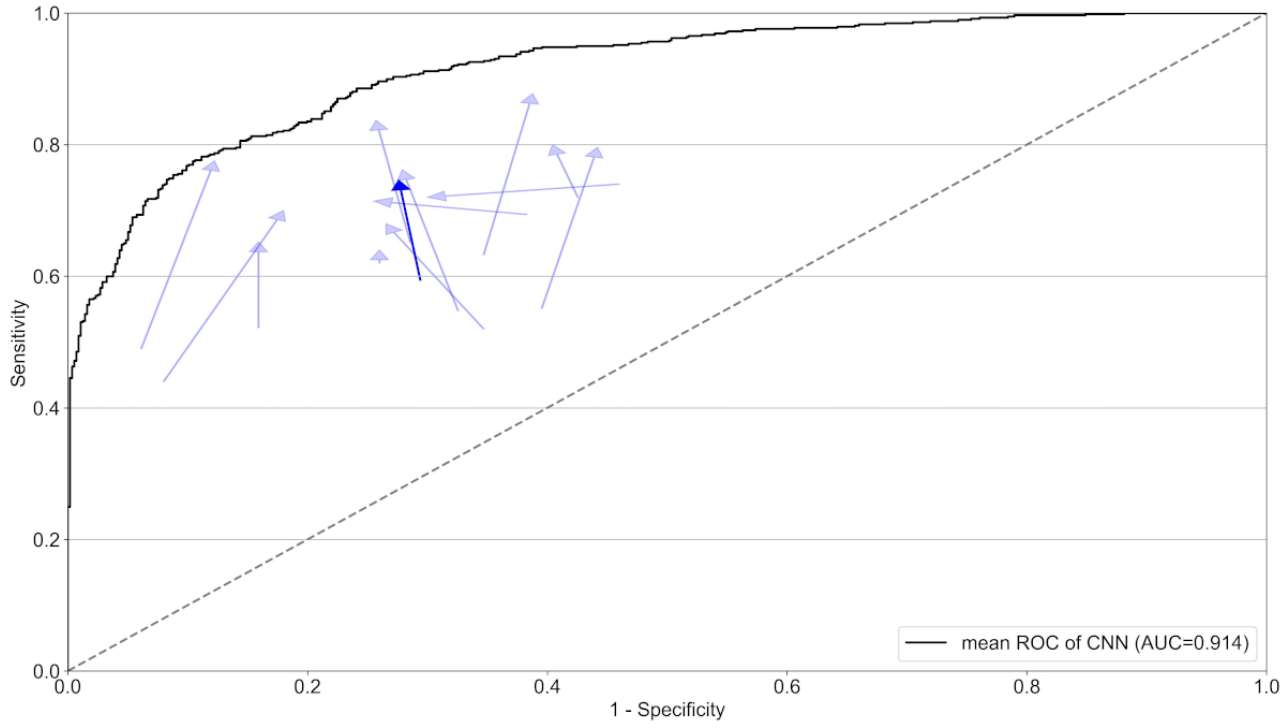


Figure 2. Combined and individual dermatologists' diagnostic accuracy without and with artificial intelligence (AI) support. Diagnostic accuracy is measured using sensitivity and specificity. Arrows represent the change in the diagnostic accuracy from without AI support to with AI support. Highlighted arrows represent the participants combined. In addition, the black curve denotes the mean receiver operating characteristic curve of the classifier. ROC: receiver operating characteristic; CNN: convolutional neural network; AUC: area under the curve.



Effects of the Use of the Classifier on the Decisions of the Dermatologists

The mean overall confidence of the dermatologists increased only marginally with the use of the classifier (65.0%; 95% CI 58.5%-71.6% without AI to 66.8%; 95% CI 60.2%-73.4% with AI). The confidence of the classifier in its diagnostic decisions was at 65.0% (95% CI 62.8%-67.4%). Upon detailed analysis, one can detect opposite effects on the dermatologists' confidence and switching behavior depending on agreement or disagreement with the classifier. On average, dermatologists and classifiers agreed in 63% (59/94; 95% CI 59.4%-66.1%) of cases. Among these, 8% (7/94) of cases were diagnosed incorrectly and 55% (52/94) of cases were classified correctly (see Table 2).

Table 2. Distribution of the correct and incorrect predictions by classifier and dermatologists without artificial intelligence (AI) support and switching in response to AI support. Percentages displayed below show the amount of times a switch did or did not occur for dermatologists when answering part II of the survey.

Groupings	Proportion, n (%)[a]	95% CI
Both incorrect, n=94	7 (8)	5.5%-10.4%
Dermatologist switched, n=7	0 (1)	0%-2.2%
Dermatologist stayed, n=7	7 (99)	97.8%-100%
AI correct, Dermatologist incorrect, n=94	25 (27)	24.0%-30.1%
Dermatologist switched, n=25	11 (46)	33.1%-58.4%
Dermatologist stayed, n=25	14 (54)	41.6%-66.9%
AI incorrect, Dermatologist correct, n=94	10 (10)	8.4%-11.8%
Dermatologist switched, n=10	4 (39)	23.2%-55.6%
Dermatologist stayed, n=10	6 (61)	44.4%-76.9%
Both correct, n=94	52 (55)	52.4%-57.1%
Dermatologist switched, n=52	0 (0)	0%-0.5%
Dermatologist stayed, n=52	52 (100)	99.5%-100%

[a]As the mean of all the participants was taken and every participant ended up rating a varying amount of images due to the quality control step, the reported absolute values are approximations.

Of the 37% (35/94; 95% CI 33.9%-40.6%) of cases in which dermatologists and classifier came to different conclusions, 10% (10/94) of cases were diagnosed correctly by the dermatologists and 27% (25/94) of cases were diagnosed correctly by the classifier.

In cases of agreement, the mean confidence of the dermatologists increased substantially from the first part of the study to the second part of the study (from 67.0%; 95% CI 60.6%-73.5% to 79.1%; 95% CI 75.0%-83.0%, respectively). In cases of disagreement, it decreased from 61.7% (95% CI 54.6%-68.9%) in the first to 44.3% (95% CI 31.8%-56.7%) in the second part of the study. This was also reflected in the dermatologists' switching behavior. In cases of agreement, the dermatologists basically never altered their classifications (0/59, 0%; 95% CI 0%-0.8%). In contrast, in cases of disagreement, the dermatologists altered their decision in 43% (15/35; 95% CI 30.8%-56.3%) of those cases. The dermatologists altered their diagnosis less often when they had initially diagnosed the lesion correctly (subsequently changed answer in 4/10, 39% of those cases) than when they had initially diagnosed it incorrectly (subsequently changed answer in 11/25, 46% of those cases, see Table 2). Altogether, this resulted in the observed overall increase in the correct diagnoses by the dermatologists. Out of all the occurring switches, dermatologists showed a higher willingness to switch from benign to malignant (11/15, 72%; 95% CI 53.4%-90.9%) than from malignant to benign (4/15, 28%; 95% CI 9.1%-46.6%).

The mean confidence levels of the classifier were much higher when the correct conclusion was reached (71.7%; 95% CI 69.3%-74.1%) than when the diagnosis was incorrect (34.8%; 95% CI 30.9%-38.7%). Regarding the dermatologists, a similar trend was less striking for part I (67.3%; 95% CI 61.1%-73.5% correct vs 60.8%; 95% CI 53.3%-68.2% incorrect) but became more pronounced for part II (71.8%; 95% CI 66.4%-77.1% correct vs 52.3%; 95% CI 41.0%-63.5% incorrect). A more detailed breakdown of the dermatologists' and classifier's confidence levels is shown in Table 3. Finally, the dermatologists altered their diagnoses in divergent cases more often when the CNN's confidence levels were high than when they were low (mean CNN confidence levels 63.3%; 95% CI 56.4%-70.1% with subsequent switch vs 53.7%; 95% CI 46.7%-60.7% with no subsequent switch).

Table 3. Confidence distribution of the classifier and dermatologists for part I and part II.

Groupings	Confidence (95% CI)
Both incorrect	
AI[a]	35.8% (27.6%-44.1%)
Dermatologist part I	62.3% (54.6%-69.9%)
Dermatologist part II	69.8% (62.4%-77.3%)
AI correct, Dermatologist incorrect	
AI	65.4% (60.7%-70.2%)
Dermatologist part I	60.8% (53.3%-68.3%)
Dermatologist part II	46.0% (33.6%-58.5%)
AI incorrect, Dermatologist correct	
AI	32.9% (28.1%-37.6%)
Dermatologist part I	63.2% (56.3%-70.2%)
Dermatologist part II	47.9% (36.4%-59.4%)
Both correct	
AI	74.5% (72.0%-76.9%)
Dermatologist part I	67.9% (61.5%-74.3%)
Dermatologist part II	80.3% (76.2%-84.3%)

[a]AI: artificial intelligence.

Learning Effect

The dermatologists received their results immediately after completing the first survey of the second part of the study. Thus, they had the opportunity to implement the lessons learned in the first AI-supported survey during the completion of the second survey in part II. A subanalysis showed that there was no detectable change in the dermatologists' performance from the first to the second survey in part I of the study (see figure in Multimedia Appendix 2) with no significant difference between sensitivity (P=.50) and specificity (P=.76). In contrast, the dermatologists tended to perform better in the second survey of part II compared to the first survey, suggesting that they had learned how to better incorporate the CNN results into their diagnostic procedures, albeit with differences in sensitivity (P=.21) and specificity (P=.43) being still insignificant. In divergent cases, participants switched their diagnosis in 49% (8/17; 95% CI 36.2%-62.3%) of cases in survey 2, whereas they had only done so in 38% (7/18; 95% CI 21.1%-54.5%) of these cases for survey 1. The tendency to switch at the correct moments in cases of disagreement was reinforced in the second survey wherein dermatologists switched far more often when they were incorrect (5/13, 39%; 95% CI 23.0%-55.7% in the first survey vs 6/12, 52%; 95% CI 39.8%-64.3% in the second survey).

Discussion

Principal Results

This study provides support for the value of CNN-based deep learning algorithms as complementary diagnostic tools in skin cancer detection as each of the 12 participating dermatologists experienced increased accuracy when working together with the classifier. In particular, the sensitivity of melanoma detection was improved. The specificity of the classification did not deteriorate substantially. This is remarkable because there is usually a trade-off between sensitivity and specificity in diagnostic tests. Thus, our classifier increased the overall accuracy of melanoma classification by the dermatologists. The one-sided performance improvement could be based on the classifier's higher confidence for the subset of melanoma images where a switch did occur coupled with the assumption that dermatologists tend to change their mind more readily toward malignant cases. Based on balanced accuracy, every physician experienced a performance boost when working with AI support. The increase in balanced accuracy varied among the participants, with participants having a lower balanced accuracy for part I, generally showing larger improvements. Interestingly, improvement was not solely determined by classifier performance as the top three most improved participants worked with a classifier performing worse than average. The observed switching rate by dermatologists when disagreeing with the classifier reinforces previous findings, which indicate physicians' susceptibility to recommendations of decision support systems [21].

The overall confidence of the dermatologists increased substantially when there was an agreement between the classifier and the dermatologist but decreased when there was a disagreement. The increased diagnostic confidence when there was an agreement could have an impact on treatment decisions, as the confidence to excise or not to excise increases, but it is not without downsides as confidence levels also increased when both agreed but were incorrect, reinforcing the participants' confidence in the wrong classification. The observed confidence trend when disagreeing is expected, as an opposite viewpoint

can create doubts in one's own diagnosis. This is however only beneficial when the classifier is correct but harmful when it is incorrect as it not only results in decreased confidence but also in participants switching from right to wrong. In the end, accuracy improvements were attained due to the combination of the classifier being right more often than wrong and dermatologists correctly being able to assess when to trust the classifier and when not to. Further improving such a system would therefore entail improving the classifier's performance as well as its integration so that participants are better capable to delineate when to trust the classifier and when to trust themselves.

The dermatologists' performance with AI support showed a nonsignificant trend of improvement between survey 1 and survey 2 for part II. This could indicate that after a single "practice session," dermatologists had gained enough experience with the use of the classifier and adapted their diagnostic procedure in a way that yielded better performance. Further practice and more detailed analysis of the way in which dermatologists interact with the classifier may improve performance even further.

The exact reason for dermatologists to switch is difficult to pin down. As each participant was shown his or her previous answer (part I) for questions for part II, the change of mind presumably occurred because of the classifier's answer. We cannot rule out that dermatologists would reconsider and switch on their own upon second viewing, but it is unlikely based on the fact that dermatologists almost never switched when agreeing with the classifier.

Limitations and Further Considerations

Currently, the algorithm on its own has a higher diagnostic accuracy than the dermatologists with AI support. However, the setting in which the dermatologists had to reach their decisions in this study does not completely reflect the real life situation wherein clinicians can integrate further information into their final decisions, such as age, patient history, or lesion localization, and an unbalanced distribution of melanomas and nevi. Therefore, adding the AI classifier to the routine diagnostic procedures already in place in the clinic and exploring its effects is an interesting next step. Completely taking the physician out of the loop is questionable as neural networks are known to have robustness issues and while patients may accept the usage of CAD systems, it is tied to the condition that physicians interpret the results of such systems and are not replaced by them [22]. Furthermore, the small sample size of the participants does not reflect the experience levels encountered in the clinic (eg, no trainee doctors) but was consciously chosen to ensure conscientious participation.

The comparatively low sensitivity attained by the participants could be due to the nature of the survey question, as participants were asked for a classification (melanoma/nevus) and not a therapy decision (eg, biopsy/reassuring patient). In real life, dermatologists will presumably tend to excise more lesions when they are unsure that the lesion in question is benign, thereby increasing their sensitivity. In a posthoc analysis, dermatologists' answers were converted from nevus to melanoma when their confidence was low, which showed an increase in excision sensitivity (see Multimedia Appendix 2). In addition, all the images selected for the study were verified by biopsy and are therefore more likely to represent edge cases, which are naturally more difficult to diagnose. This coupled with the nature of the question and the very large test set size could explain the lower diagnostic performance.

A limitation of our current classifier is that it can only distinguish melanomas from pigmented nevi. However, multiple studies have shown that it is possible to create CNN-based algorithms that can distinguish several classes of lesions, thus better reflecting the clinical reality [6,8,23]. Further, the classifier's performance on an external test set would likely be lower; however, classifier performance was not the primary aspect of this study, but rather if and how dermatologists are influenced by such a system.

Conclusions

Our results support further research into AI-based classifiers as diagnostic aids in skin cancer classification. We show that clinicians can improve their overall accuracy through improving sensitivity at constant specificity by learning to optimize their interactions with a classifier. While users switched to the correct answer more often than to the incorrect one, minimizing incorrect switches is a challenge that requires further investigation. Our study also has some limitations such as a comparatively artificial setting. In future, clinical trials should be performed to investigate how AI-based classifiers affect skin cancer classification in a real-life setting in which an improved classifier is incorporated in the diagnostic routine.

Acknowledgments

We would like to thank Dr. M. Petri for aiding in the data acquisition by participating in the survey and for his critical review of the manuscript. This study was funded by the Federal Ministry of Health, Berlin, Germany (grant: Skin Classification Project; grant holder: Titus J. Brinker, German Cancer Research Center, Heidelberg, Germany). The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflicts of Interest

TJB reports owning a company that develops mobile apps (Smart Health Heidelberg GmbH, Handschuhsheimer Landstr. 9/1, 69120 Heidelberg). SH reports advisory roles for or has received honoraria from Pierre Fabre Pharmaceuticals, Novartis, Roche, BMS, Amgen and MSD outside the submitted work. AH reports clinical trial support, speaker's honoraria, or consultancy fees

from the following companies: Amgen, BMS, Merck Serono, MSD, Novartis, Oncosec, Philogen, Pierre Fabre, Provectus, Regeneron, Roche, OncoSec, Sanofi-Genzyme, and Sun Pharma. ES reports advisory roles for Heine Optotechnik GmbH and has received honoraria or travel support from Heine Optotechnik, LaRoche Posay, Naos and Abbott outside the submitted work. BS reports advisory roles for or has received honoraria from Pierre Fabre Pharmaceuticals, Incyte, Novartis, Roche, BMS and MSD, research funding from BMS, Pierre Fabre Pharmaceuticals and MSD, and travel support from Novartis, Roche, BMS, Pierre Fabre Pharmaceuticals and Amgen outside the submitted work. JSU is on the advisory board or has received honoraria and travel support from Amgen, Bristol Myers Squibb, GlaxoSmithKline, LeoPharma, Merck Sharp and Dohme, Novartis, Pierre Fabre, Roche, outside the submitted work.

Multimedia Appendix 1
Detailed methods description.
[DOC File , 30 KB - jmir_v22i9e18091_app1.doc]

Multimedia Appendix 2
Additional figures and tables.
[DOC File , 145 KB - jmir_v22i9e18091_app2.doc]

References

1. Rosendahl C, Tschandl P, Cameron A, Kittler H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. J Am Acad Dermatol 2011 Jun;64(6):1068-1073. [doi: 10.1016/j.jaad.2010.03.039] [Medline: 21440329]
2. Sinz C, Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, et al. Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. J Am Acad Dermatol 2017 Dec;77(6):1100-1109. [doi: 10.1016/j.jaad.2017.07.022] [Medline: 28941871]
3. Vestergaard ME, Macaskill P, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. Br J Dermatol 2008 Sep;159(3):669-676. [doi: 10.1111/j.1365-2133.2008.08713.x] [Medline: 18616769]
4. Tschandl P, Wiesner T. Advances in the diagnosis of pigmented skin lesions. Br J Dermatol 2018 Jan;178(1):9-11. [doi: 10.1111/bjd.16109] [Medline: 29357612]
5. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Collaborators. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer 2019 May;113:47-54 [FREE Full text] [doi: 10.1016/j.ejca.2019.04.001] [Medline: 30981091]
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
7. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader study level-Ilevel-II Groups, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842 [FREE Full text] [doi: 10.1093/annonc/mdy166] [Medline: 29846502]
8. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. J Invest Dermatol 2018 Jul;138(7):1529-1538 [FREE Full text] [doi: 10.1016/j.jid.2018.01.028] [Medline: 29428356]
9. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. J Med Internet Res 2018 Oct 17;20(10):e11936 [FREE Full text] [doi: 10.2196/11936] [Medline: 30333097]
10. Kurvers RHJM, Krause J, Argenziano G, Zalaudek I, Wolf M. Detection Accuracy of Collective Intelligence Assessments for Skin Cancer Diagnosis. JAMA Dermatol 2015 Dec 01;151(12):1346-1353. [doi: 10.1001/jamadermatol.2015.3149] [Medline: 26501400]
11. Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, Collaborators. Superior skin cancer classification by the combination of human and artificial intelligence. Eur J Cancer 2019 Oct;120:114-121 [FREE Full text] [doi: 10.1016/j.ejca.2019.07.019] [Medline: 31518967]
12. Marchetti MA, Liopyris K, Dusza SW, Codella NCF, Gutman DA, Helba B, International Skin Imaging Collaboration. Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017. J Am Acad Dermatol 2020 Mar;82(3):622-627. [doi: 10.1016/j.jaad.2019.07.016] [Medline: 31306724]
13. Hauschild A, Chen SC, Weichenthal M, Blum A, King HC, Goldsmith J, et al. To excise or not: impact of MelaFind on German dermatologists' decisions to biopsy atypical lesions. J Dtsch Dermatol Ges 2014 Jul;12(7):606-614. [doi: 10.1111/ddg.12362] [Medline: 24944011]

14. Rigel DS, Roy M, Yoo J, Cockerell CJ, Robinson JK, White R. Impact of guidance from a computer-aided multispectral digital skin lesion analysis device on decision to biopsy lesions clinically suggestive of melanoma. Arch Dermatol 2012 Apr;148(4):541-543. [doi: 10.1001/archdermatol.2011.3388] [Medline: 22351788]
15. Cukras AR. On the comparison of diagnosis and management of melanoma between dermatologists and MelaFind. JAMA Dermatol 2013 May;149(5):622-623. [doi: 10.1001/jamadermatol.2013.3405] [Medline: 23677101]
16. Sadeghi M, Chilana PK, Atkins MS. How users perceive content-based image retrieval for identifying skin images. In: Understanding and Interpreting Machine Learning in Medical Image Computing Applications. Cham: Springer International Publishing; 2018:141-148.
17. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11591-11596 [FREE Full text] [doi: 10.1073/pnas.1806905115] [Medline: 30348771]
18. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med 2020;3:23 [FREE Full text] [doi: 10.1038/s41746-020-0232-8] [Medline: 32140566]
19. Codella NCF, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). arxiv. 2017. URL: https://arxiv.org/abs/1710.05006 [accessed 2020-04-05]
20. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018 Aug 14;5:180161 [FREE Full text] [doi: 10.1038/sdata.2018.161] [Medline: 30106392]
21. Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artif Intell Med 2005 Jan;33(1):25-30. [doi: 10.1016/j.artmed.2004.07.007] [Medline: 15617979]
22. Fink C, Uhlmann L, Hofmann M, Forschner A, Eigentler T, Garbe C, et al. Patient acceptance and trust in automated computer-assisted diagnosis of melanoma with dermatofluoroscopy. J Dtsch Dermatol Ges 2018 Jul;16(7):854-859. [doi: 10.1111/ddg.13562] [Medline: 29927518]
23. Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, Collaborators. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur J Cancer 2019 Sep;119:57-65 [FREE Full text] [doi: 10.1016/j.ejca.2019.06.013] [Medline: 31419752]

Abbreviations

AI: artificial intelligence
CAD: computer-aided diagnosis
CNN: convolutional neural network
ISIC: International Skin Imaging Collaboration

Edited by G Eysenbach; submitted 04.02.20; peer-reviewed by L Zhang, CT Cheng; comments to author 23.03.20; revised version received 28.03.20; accepted 14.05.20; published 11.09.20.

Please cite as:
Maron RC, Utikal JS, Hekler A, Hauschild A, Sattler E, Sondermann W, Haferkamp S, Schilling B, Heppt MV, Jansen P, Reinholz M, Franklin C, Schmitt L, Hartmann D, Krieghoff-Henning E, Schmitt M, Weichenthal M, von Kalle C, Fröhling S, Brinker TJ
Artificial Intelligence and Its Effect on Dermatologists' Accuracy in Dermoscopic Melanoma Image Classification: Web-Based Survey Study
J Med Internet Res 2020;22(9):e18091
URL: https://www.jmir.org/2020/9/e18091
doi:10.2196/18091
PMID:32915161

Original Paper

An Innovative Artificial Intelligence–Based App for the Diagnosis of Gestational Diabetes Mellitus (GDM-AI): Development Study

Jiayi Shen[1,2*], MBBS; Jiebin Chen[3*], BSc; Zequan Zheng[4], MBBS; Jiabin Zheng[4], MBBS; Zherui Liu[3], BSc; Jian Song[5], BA; Sum Yi Wong[4], MBBS; Xiaoling Wang[6], BA; Mengqi Huang[6], BA; Po-Han Fang[4], MBBS; Bangsheng Jiang[4], MBBS; Winghei Tsang[1], MBBS; Zonglin He[4], MBBS; Taoran Liu[7], BSc; Babatunde Akinwunmi[8,9], MD, MPH, MMSc; Chi Chiu Wang[10], MBBS, PhD; Casper J P Zhang[11], MPH, PhD; Jian Huang[12], MPH, PhD; Wai-Kit Ming[1], MD, MPH, PhD, MMSc

[1]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[2]Sun Yat-sen University Cancer Center, Guangzhou, China

[3]College of Information Science and Technology, Jinan University, Guangzhou, China

[4]International School, Jinan University, Guangzhou, China

[5]School of International Studies, Sun Yat-sen University, Guangzhou, China

[6]School of Journalism and Communication, Jinan University, Guangzhou, China

[7]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

[8]Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, United States

[9]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Harvard University, Boston, MA, United States

[10]Department of Obstetrics & Gynaecology, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[11]School of Public Health, The University of Hong Kong, Hong Kong, Hong Kong

[12]Multidisciplinary, Collaborative Research Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, St Mary's Campus, Imperial College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, MPH, PhD, MMSc
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
Guangzhou
China
Phone: 86 14715485116
Email: wkming@connect.hku.hk

Abstract

Background: Gestational diabetes mellitus (GDM) can cause adverse consequences to both mothers and their newborns. However, pregnant women living in low- and middle-income areas or countries often fail to receive early clinical interventions at local medical facilities due to restricted availability of GDM diagnosis. The outstanding performance of artificial intelligence (AI) in disease diagnosis in previous studies demonstrates its promising applications in GDM diagnosis.

Objective: This study aims to investigate the implementation of a well-performing AI algorithm in GDM diagnosis in a setting, which requires fewer medical equipment and staff and to establish an app based on the AI algorithm. This study also explores possible progress if our app is widely used.

Methods: An AI model that included 9 algorithms was trained on 12,304 pregnant outpatients with their consent who received a test for GDM in the obstetrics and gynecology department of the First Affiliated Hospital of Jinan University, a local hospital in South China, between November 2010 and October 2017. GDM was diagnosed according to American Diabetes Association (ADA) 2011 diagnostic criteria. Age and fasting blood glucose were chosen as critical parameters. For validation, we performed k-fold cross-validation (k=5) for the internal dataset and an external validation dataset that included 1655 cases from the Prince of Wales Hospital, the affiliated teaching hospital of the Chinese University of Hong Kong, a non-local hospital. Accuracy, sensitivity, and other criteria were calculated for each algorithm.

Results: The areas under the receiver operating characteristic curve (AUROC) of external validation dataset for support vector machine (SVM), random forest, AdaBoost, k-nearest neighbors (kNN), naive Bayes (NB), decision tree, logistic regression (LR), eXtreme gradient boosting (XGBoost), and gradient boosting decision tree (GBDT) were 0.780, 0.657, 0.736, 0.669, 0.774, 0.614, 0.769, 0.742, and 0.757, respectively. SVM also retained high performance in other criteria. The specificity for SVM retained 100% in the external validation set with an accuracy of 88.7%.

Conclusions: Our prospective and multicenter study is the first clinical study that supports the GDM diagnosis for pregnant women in resource-limited areas, using only fasting blood glucose value, patients' age, and a smartphone connected to the internet. Our study proved that SVM can achieve accurate diagnosis with less operation cost and higher efficacy. Our study (referred to as GDM-AI study, ie, the study of AI-based diagnosis of GDM) also shows our app has a promising future in improving the quality of maternal health for pregnant women, precision medicine, and long-distance medical care. We recommend future work should expand the dataset scope and replicate the process to validate the performance of the AI algorithms.

(J Med Internet Res 2020;22(9):e21573) doi:10.2196/21573

KEYWORDS

AI; application; disease diagnosis; maternal health care; artificial intelligence; app; women; rural; innovation; diabetes; gestational diabetes; diagnosis

Introduction

Gestational diabetes mellitus (GDM), common in pregnancy, exerts negative effects on both mothers and their newborns, including cesarean delivery, shoulder dystocia, macrosomia, neonatal hypoglycemia, post-GDM type 2 diabetes mellitus, cardiovascular disease of pregnant women, and increased risk of obesity and type 2 diabetes mellitus on the offspring [1]. However, if GDM can be diagnosed at an early stage, early interventions can be implemented to maximally reduce its adverse consequences [2,3]. Although GDM prevalence in some developing African countries is high (eg, 8.2% in Nigeria and 9.5% in Tanzania) [4], pregnant women are less likely to receive adequate health care due to the lack of skilled health workers [5]. Other factors, such as poverty, inadequate medical services, long distance to hospitals, less access to information, and culture and traditions also prevent women from seeking care during pregnancy.

Artificial intelligence (AI) has been widely used in disease diagnosis in recent years [6,7]. Several advanced AI algorithms, such as deep learning, support vector machine (SVM), and convolutional neural network, have shown comparable performance to clinicians [8]. Major advanced AI approaches yield significant discriminative performance with relatively high sensitivity, specificity, and accuracy in object-identifying tasks [9,10]. At the same time, the world has witnessed the instantaneity of reporting and the consistency of producing results by AI [11]. AI is becoming more suitable for use in clinical daily practice [12] and offers the advantage of greater accuracy and efficiency [13]. An AI-driven dietary platform has been developed for diabetes management [14], and AI tools can enhance diabetes care for individuals and societal health [15,16].

Due to the advantages above, AI is expected to be further studied and implemented in the GDM diagnosis field to maximize social and economic benefits. A systematic review and meta-analysis on telemedicine technologies for diabetes in pregnancy conducted by our team in 2016 showed that telemedicine technologies can streamline clinical care delivery and improve maternal satisfaction [17]. We also evaluated the current state of GDM diagnosis programs by searching Scopus, Web of Science, PubMed, and Embase for studies published in English from inception up to November 17, 2019, using the keywords "gestational diabetes mellitus," "GDM," "GDM screening," "GDM detection," "GDM diagnosis," "machine learning," "artificial intelligence (AI)," and "deep learning." Although some papers applied AI on screening or early diagnosis of GDM [18,19], they only used the expert system or risk score model instead of up-to-date AI algorithms such as random forest. Recently, a team from Israel applied top 20 contributing features such as baseline risk score and glucose challenge test results of previous pregnancy. A machine learning model based on national electronic health records reached high accuracy for GDM diagnosis [20]. Therefore, we intend to establish a GDM diagnosis tool using AI technology for women in low-resource areas. As our app targets to serve patients in resource-limited areas, it would be more practical and accessible if we can use only fasting glucose value and other patient's basic health information such as age, body weight, and height. This study (referred to as GDM-AI study, ie, the study of AI-based diagnosis of GDM) aims to validate and rank the performance and applicability of AI algorithms in diagnosing GDM and to develop an innovative AI application for maternal health care. This paper will also present the ideas behind our app, as well as its contributions and prospects.

Methods

Recruitment

Our retrospective study initially involved 12,316 pregnant women who delivered a singleton at the First Affiliated Hospital of Jinan University (Guangzhou, Guangdong, China) from November 1, 2010, to October 31, 2017. We obtained ethics review and approval from the research ethics committee of the Jinan University. Medical records were used, and all data was confidential with anonymized numbers. The study excluded 12 pregnant women, as their profiles were not complete. A total of 12,304 pregnant women with full profile after admission were used as the development set and were diagnosed with or without GDM according to the International Association of Diabetes and Pregnancy Study Groups (IADPSG) diagnosis

criteria. Patients were excluded if they met any of the following criteria: non-Chinese, multiple gestations, oral glucose tolerance test performed before 12 weeks, delivery in another hospital, major fetal malformation, or without patient clinical outcome in the electronic medical records. We extracted clinical data from the database of the First Affiliated Hospital of Jinan University (Guangzhou, Guangdong, China) as the development set, including the clinical baseline characteristics, maternal and neonatal complications, along with their clinical outcomes. The demographics and clinical characteristics of the patients are presented in Table 1. Another dataset of 1655 cases was obtained with the same criteria as an external validation set from Prince of Wales Hospital, the affiliated teaching hospital of the Chinese University of Hong Kong.

Table 1. Baseline characteristics.

Demographic and clinical variables	Developmental (Training) set (N=12,304)			External validation set (N=1655)		
	GDM[a] (n=2761) (mean, SD)	NGT[b] (n=9543) (mean, SD)	*P* value	GDM (n=240) (mean, SD)	NGT (n=1415) (mean, SD)	*P* value
Age (year)	30.21 (4.42)	28.50 (3.98)	<.001	32.87 (4.71)	30.33 (4.82)	<.001
Fasting Glucose Value (mmol/L)	4.89 (0.73)	4.37 (0.36)	<.001	4.74 (0.58)	4.32 (0.28)	<.001
1-h postload plasma glucose (mmol/L)	9.82 (1.84)	7.33 (1.38)	<.001	10.16 (1.63)	7.25 (1.35)	<.001
2-h postload plasma glucose (mmol/L)	8.53 (1.65)	6.47 (1.04)	<.001	8.63 (1.23)	6.24 (1.02)	<.001

[a]GDM: gestational diabetes mellitus

[b]NGT: normal glucose tolerance

Study Design

All pregnant women received 2-hour 75 g oral glucose tolerance test according to the American Diabetes Association (ADA) 2011 criteria. As per the ADA 2011 diagnostic criteria for GDM, the upper limits of the blood glucose for fasting, 1-hour postprandial, and 2-hour postprandial blood glucose are 5.1 mmol/L, 10.0 mmol/L, and 8.5 mmol/L, respectively. Those with one or more abnormal value(s) will be diagnosed as GDM.

Each case was carefully reviewed by 2 experts individually. The patient's evaluation result, which would be finished within a week, was assigned to GDM cohort and non-GDM cohort, depending on the laboratory data change, clinical manifestation, clinical intervention, and final diagnosis. If a discrepancy occurred, the case was reviewed by another third expert and labeled after consensus was reached. In the developmental dataset, cases were labeled as either normal or GDM, according to the ADA 2011 criteria.

We compared baseline characteristics between the 2 cohorts (nonGDM and GDM) as shown in Table 1 and found that there were significant differences between them, except height. Therefore, we tested several times different combinations of baseline characteristics in AI algorithms to get the best combination for distinguishing GDM from nonGDM. Eventually, we came to a combination of age and fasting blood glucose.

We tested 9 advanced AI algorithms, including SVM, random forest, AdaBoost, k-nearest neighbors (kNN), naive Bayes (NB), decision tree, logistic regression (LR), eXtreme gradient boosting (XGBoost), and gradient boosting decision tree (GBDT), by utilizing age and fasting blood glucose in the collected datasets. We used the development set for model development and carried out internal validation with 5-fold cross-validation. The development set was randomly split into 5 folds. Within each fold, we used 1 fold for validation and the rest for training the model. Results were calculated by averaging the results from the 5 separate experiments.

GDM classification performances of the trained models were validated using the internal validation with 5-fold cross-validation and the external validation set, which was evaluated using the area under the receiver operating characteristic curve (AUROC).

Statistical Analysis

We hypothesized that the AI model is at least comparable to the ADA 2011 criteria. Thus, we compared performances of the advanced AI models to ADA 2011 diagnosis results. Accuracy, sensitivity, specificity, area under the curve (AUC), positive predictive value (PPV), negative predictive value (NPV), Brier score, positive likelihood ratio, and negative likelihood ratio were calculated for each algorithm.

For the validation datasets, performance was evaluated from the probability values by using the AUROC curve analysis for GDM detection with python 3.6.8 based on Jupyter Notebook (Project Jupyter).

Performance evaluation was achieved via receiver operating characteristic (ROC) curve analysis, and calculation of AUC using the "pROC" package. The cut-off value (0.501) for the model was determined by the "OptimalCutpoints" package using the Youden method. Asymptotic 2-sided 95% CIs were computed for the logit transform of each proportion (ie, sensitivity and specificity). All analyses were performed using Stata (version 14.0).

Results

A total of 12,304 outpatient cases were included in the development dataset. Among these women, 77.6% (9543/12304) were non-GDM as per the ADA 2011 criteria, and 22.4% (2761/12,304) women were diagnosed as GDM. GDM was found in 14.5% (240/1655) cases among the external validation dataset.

A 5-fold cross-validation was applied for the internal dataset, and accuracy and false positives were recorded. All the evaluation indices are presented in Table 2 and Table 3.

ROC curves are shown in Figure 1, which indicates the higher performance of AUC of SVM, AdaBoost, NB, LR, XGBoost, and GBDT.

For the internal validation dataset, the best performance of AUC was for GBDT and XGBoost, which had relatively higher accuracy, specificity, and PPV, while their Brier scores were lower than those of others (Table 2).

For the external validation dataset, the AUC of SVM for GDM was 0.78, with 88.7% accuracy and 100% specificity (Table 3). The specificity of NB was 98.2% for diagnosing GDM with AUC of 0.774 (Table 3). Demonstration of AI Application is shown in Table 4.

Table 2. The detection performance of 9 algorithms for the internal validation dataset.

Algorithms	Accuracy	Sensitivity	Specificity	PPV[a]	NPV[b]	Brier score	AUC[c]
SVM[d]	0.849	0.377	0.985	0.880	0.845	0.151	0.766
Random forest	0.833	0.432	0.949	0.709	0.852	0.167	0.728
AdaBoost	0.860	0.376	1	1	0.847	0.140	0.763
kNN[e]	0.841	0.415	0.964	0.768	0.851	0.159	0.723
NB[f]	0.845	0.367	0.983	0.860	0.843	0.155	0.768
Decision tree	0.838	0.431	0.956	0.738	0.853	0.162	0.706
LR[g]	0.844	0.363	0.984	0.865	0.842	0.156	0.765
XGBoost[h]	0.860	0.377	1	1	0.847	0.140	0.771
GBDT[i]	0.860	0.376	1	1	0.847	0.140	0.772

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the curve.

[d]SVM: support vector machine.

[e]kNN: k-nearest neighbors.

[f]NB: naive Bayes.

[g]LR: logistic regression.

[h]XGBoost: eXtreme gradient boosting.

[i]GBDT: gradient boosting decision tree.

Table 3. The detection performance of 9 algorithms for the external validation dataset.

Algorithms	Accuracy	Sensitivity	Specificity	PPV[a]	NPV[b]	Brier score	AUC[c]
SVM[d]	0.887	0.221	1	1	0.883	0.113	0.780
Random forest	0.838	0.263	0.936	0.409	0.882	0.162	0.655
AdaBoost	0.882	0.183	1	1	0.878	0.118	0.736
kNN[e]	0.862	0.254	0.965	0.550	0.884	0.138	0.669
NB[f]	0.878	0.263	0.982	0.716	0.887	0.122	0.774
Decision tree	0.841	0.242	0.942	0.414	0.880	0.159	0.614
LR[g]	0.877	0.258	0.983	0.713	0.887	0.123	0.769
XGBoost[h]	0.882	0.183	1	1	0.878	0.118	0.742
GBDT[i]	0.882	0.183	1	1	0.878	0.118	0.757

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the curve.

[d]SVM: support vector machine.

[e]kNN: k-nearest neighbors.

[f]NB: naive Bayes.

[g]LR: logistic regression.

[h]XGBoost: eXtreme gradient boosting.

[i]GBDT: gradient boosting decision tree.

Figure 1. Overall area under the receiver operating characteristic curves for internal validation dataset. SVM: support vector machine; knn: k-nearest neighbors; NB: naive Bayes; LR: logistic regression; XGBoost: eXtreme gradient boosting; GBDT: gradient boosting decision tree.

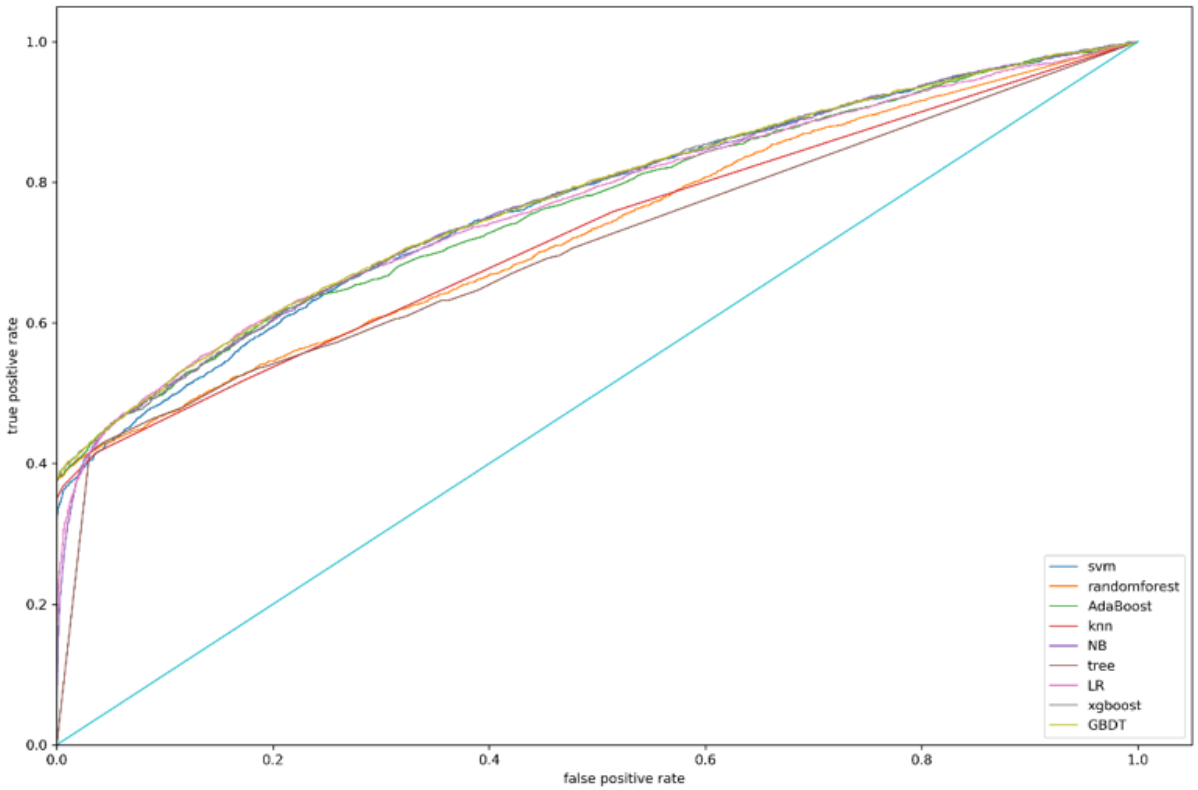


Table 4. Demonstration of AI application.

Sample	Age	Fasting glucose (mmol/L)	Result with AI application
1	35	5.2	GDM[a]
2	25	4.5	No GDM
3	27	4.8	No GDM
4	33	3.5	No GDM
5	37	5.6	GDM
6	30	4.3	No GDM
7	30	6.7	GDM
8	27	5.4	GDM

[a]GDM: gestational diabetes mellitus.

The goal of this preliminary research was to develop an app and demonstrate the soundness of the methods. The users can put data into the app, which will be transferred simultaneously to the doctors and other medical personnel. In this way, the app provides a platform for the doctor to get informed of the users' health status and give intervention to the GMD patients, while the app user can keep track of their physical conditions.

To be more specific, our mobile app is expected to quantify the daily lives of pregnant women and optimize diet, exercise, and sleep to help them maximize their well-being. The app serves as an "online intelligent nurse" that can answer simple questions to reduce obstetric doctors' workload. The built-in model was chosen for our app, which will still be tested, adjusted, and improved continually. There is a need to collect data regarding pregnant women's daily habits and body conditions and to use machine learning algorithms to study these data. Such data will also be sent to a cloud database and analyzed again to determine the relationship between the amount of exercise and caloric intake to help users balance exercise and diet. Moreover, it will give a prediction of the probability of users having other diseases and their next-day blood glucose levels. The actual blood glucose uploaded by the user will be used to make comparisons with the predicted value to verify, correct, or improve our model.

The detailed process of app development is illustrated in Figure 2 and Figure 3.

The app interface is shown in Figure 4.

Figure 2. How the AI app works.

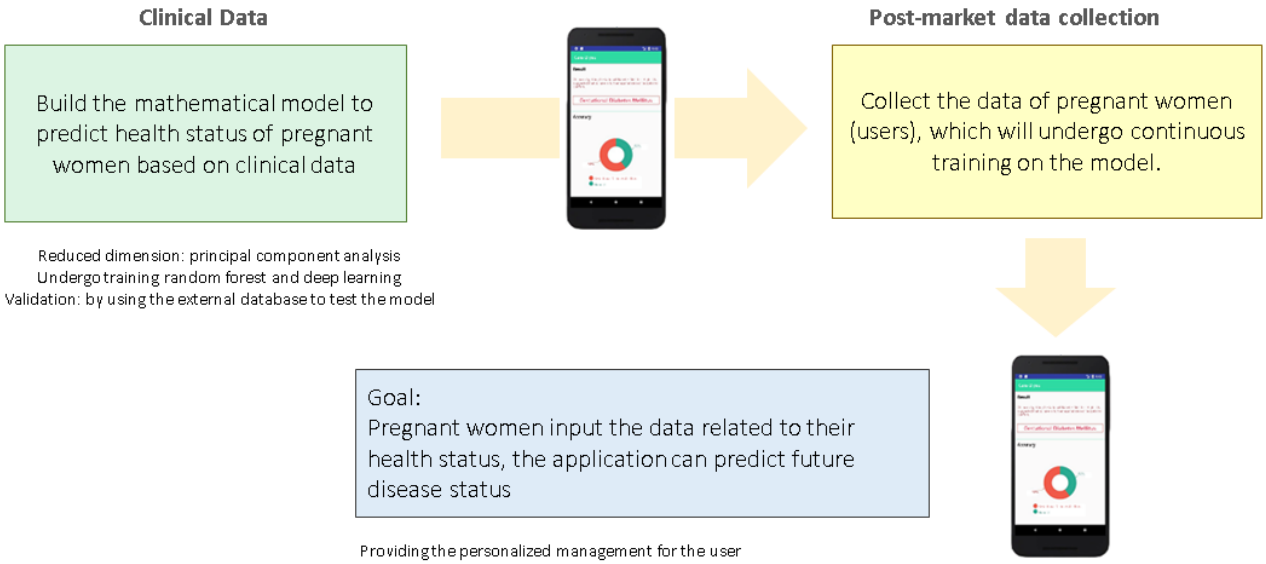


Figure 3. Structure of the app.

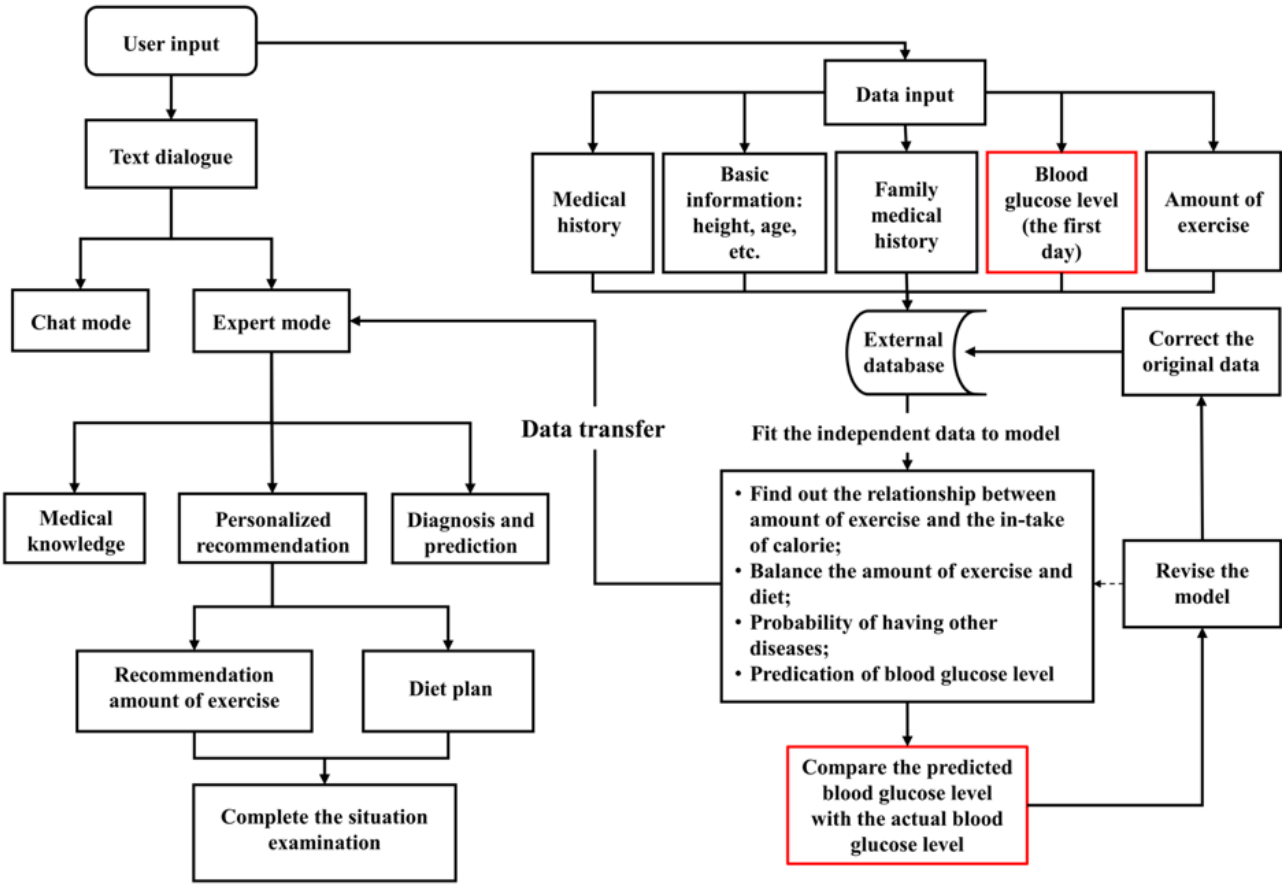
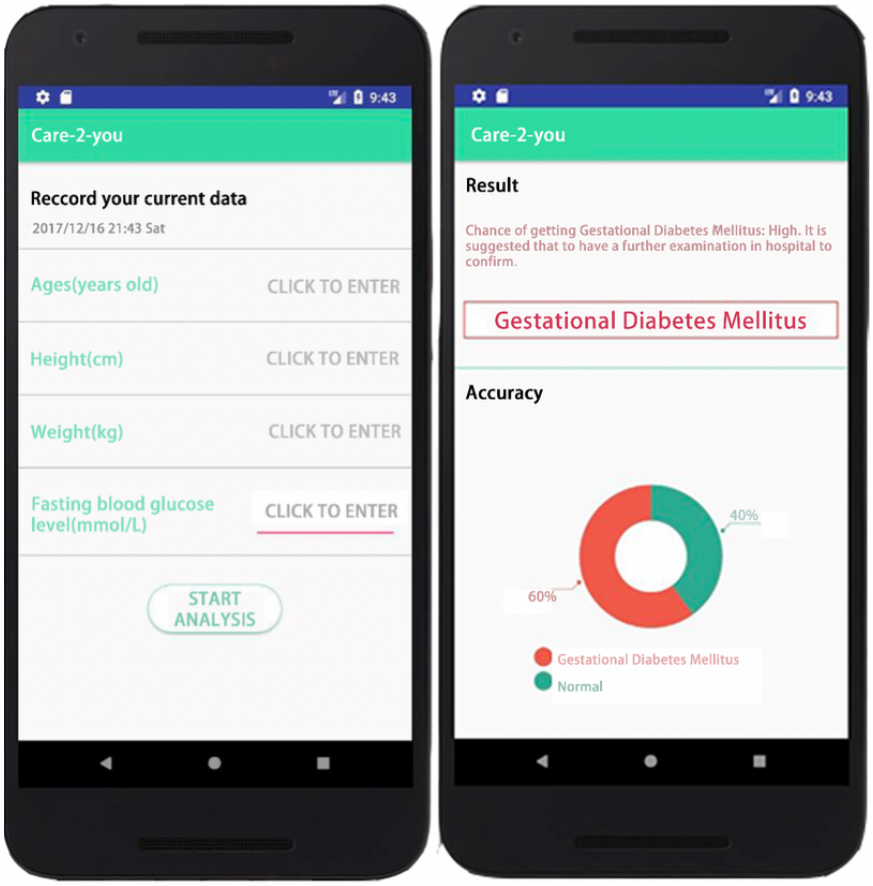


Figure 4. Interface of the app.



Discussion

Principal Results

In this study, we described the use of 9 AI algorithms for the diagnosis of GDM, utilizing only the age and fasting blood glucose. Internal and external validation demonstrated that AI algorithms provide strong or moderate evidence to rule in diagnosing GDM under the same medical resource conditions. Moreover, SVM retained high performance in accuracy, specificity, PPV, positive likelihood ratio, and AUROC for achieving the correct diagnosis, which suggests the potential of the SVM algorithm to make highly accurate diagnosis decisions. NB also provides moderate evidence to rule in diagnosing GDM under the same medical resource conditions.

Established diagnostic tools such as ADA 2011 criteria have been used to decide whether or not a pregnant woman has GDM. However, ADA 2011 criteria are resource-demanding, and thus may not be utilized in undeveloped areas in China. Moreover, these diagnostic tests are both expensive and difficult for pregnant women given their body conditions during pregnancy. Compared with those established diagnostic tools in the ADA 2011 criteria, the automatic diagnosis algorithm can provide a real-time and accurate diagnosis with fewer medical resources. Also, such algorithm-based diagnosis would be less expensive since it requires fewer equipment and professional medical staff. As GDM puts a major economic burden on the public health care system, the government and health policymakers can evaluate the economic benefits of our free app (which can inform patients of their diagnoses and facilitate early medical intervention, if needed) based on the results of this study and seek international cooperation [21-26].

We believe that the future focus of AI in medicine should be directed towards solving medical problems in resource-insufficient areas, and this application will likely help address the shortage of medical resources. In addition to GDM, we believe that AI could be applied to other diseases.

Limitations

First, the development set all comes from one hospital, but the external validation set has fixed this problem. Second, the data is retrospective and therefore, not up to date. Third, the datasets pertain to the Guangdong and Hong Kong populations, both in regard to patient and system characteristics. Applicability of our findings to other populations with distinct health care systems may need further investigations. Fourth, there is no algorithm that performs well in sensitivity compared with human experts, illustrating that the capability of AI-diagnosing GDM requires improvement.

Since the scope of our dataset is relatively small, the next step of our study will be to expand our internal dataset and repeat the process to validate that SVM performs well in different datasets. Databases from different jurisdictions can be included in our test. Although SVM outperformed in overall criteria, NB outperformed SVM according to Brier Score in both datasets. Therefore, a further investigation into the differences between SVM and NB will be carried out.

Comparison With Existing Literature

Diagnosis for initial GDM is performed to aid further observation of pregnant women and to guide interventions. Several tools have been previously developed for diagnosing and predicting diabetes mellitus. However, these tools require detailed information of a pregnant woman, for example, all important factors regulating the patient's blood glucose level [27], demographic information [18], or various blood glucose values [28]. Therefore, they are not suitable for pregnant women living in resource-limited rural areas. In contrast, our diagnosis mechanism only requires the fasting glucose value and the patient's age, which enhances its application.

Contribution of the Application

A core feature of the app is intelligent medical care. The process of collecting, extracting, processing, and presenting data and renewing the app are both automated. Therefore, it can save a lot of time and effort while performing at high efficacy and accuracy.

Moreover, if our app can be interconnected with wearable devices, the app can monitor, in real-time, the heart rates and blood pressure levels of pregnant women to inform them of their physical conditions with data and illustrations, such as whether they should continue to exercise or rest. Such intelligent health care will be beneficial for mothers-to-be in rural areas where medical resources might be in short supply, reassuring both pregnant women and their families.

Precision medicine is a modern branch of medicine that can give individualized medical care according to a patient's genetic, biomarker, or psychosocial characteristics. Precision medicine remains expensive and difficult to deliver for most pregnant women in rural areas. However, our app provides the possibility for affordable precision medicine providers since it can respond with medical advice and individualized arrangements of daily exercises and diets.

The application also makes the long-distance medicine possible for those pregnant women living in rural areas, where medical services are insufficient and transportation is underdeveloped. Doctors can track the patient's glucose level and other health information through the application in real-time and give diagnosis and suggestions through real-time online communication. The application can also be used for urban pregnant women who are too busy to attend time-consuming examinations.

In addition, through the realization of precision medicine and long-distance medical care, this app can use available obstetric resources to detect a relationship between health data and pregnant women's health to search for new ways to manage risks during pregnancy and to provide more effective management of GDM. In this way, our app can improve the efficiency and quality of maternal health care, particularly in rural areas, and help revitalize the current global medical system.

The AI-based app can promote long-distance medical care by making timely and accurate diagnosis in low-resource conditions while possibly lowering the cost of GDM diagnosis and improving the quality and efficiency of maternal health care in

rural areas. First, the introduction of this application will effectively be improving diagnostic rate of gestational diabetes in low-resource areas and thus preventing high risks in pregnancies. Second, AI makes intervention for high-risk women possible. Third, AI application might be complementary solutions to reduce diagnostic delay and delay of getting prevention advice and possible treatment in underserved rural GDM population globally.

Conclusion

There are many challenges associated with inadequate obstetric services in rural areas around the world, yet this also provides an opportunity for the development of AI in the medical field. In our study, 9 algorithms (SVM, random forest, AdaBoost, kNN, NB, decision tree, LR, XGBoost, and GBDT) were tested to identify the best-performing algorithm in the diagnosis of GDM. SVM performed best and was adopted to develop a mobile app. Although further experiments are needed, we believe the developed app will promote precision medicine and long-distance medical care while improving the quality and efficiency of maternal health care in rural areas.

Authors' Contributions

J Shen and JC contributed equally to this article. J Shen and JC conceived the original idea and designed the study. JC did the experimental studies and drew figures; J Shen did the data analysis and data interpretation and wrote the first version of the manuscript. ZL contributed to programming. Both ZZ and JZ contributed to the final version of the manuscript. J Song helped write the manuscript. SYW, XW, PHF, BJ, WT, CCW, and MH collected clinical data. BA, CJPZ, JH, and TL contributed to the manuscript draft and data analysis. WKM conceived the original idea, designed the study, and supervised the project. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Conflicts of Interest

None declared.

References

1. Chiefari E, Arcidiacono B, Foti D, Brunetti A. Gestational diabetes mellitus: an updated overview. J Endocrinol Invest 2017 Sep;40(9):899-909 [FREE Full text] [doi: 10.1007/s40618-016-0607-5] [Medline: 28283913]
2. Ming W, Ding W, Zhang CJP, Zhong L, Long Y, Li Z, et al. The effect of exercise during pregnancy on gestational diabetes mellitus in normal-weight women: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2018 Nov 12;18(1):440 [FREE Full text] [doi: 10.1186/s12884-018-2068-7] [Medline: 30419848]
3. Shepherd E, Gomersall J, Tieu J, Han S, Crowther C, Middleton P. Combined diet and exercise interventions for preventing gestational diabetes mellitus. Cochrane Database Syst Rev 2017 Nov 13;11:CD010443 [FREE Full text] [doi: 10.1002/14651858.CD010443.pub3] [Medline: 29129039]
4. Zhu Y, Zhang C. Prevalence of Gestational Diabetes and Risk of Progression to Type 2 Diabetes: a Global Perspective. Curr Diab Rep 2016 Jan;16(1):7 [FREE Full text] [doi: 10.1007/s11892-015-0699-x] [Medline: 26742932]
5. Bradley S, Kamwendo F, Chipeta E, Chimwaza W, de Pinho H, McAuliffe E. Too few staff, too many patients: a qualitative study of the impact on obstetric care providers and on quality of care in Malawi. BMC Pregnancy Childbirth 2015 Mar 21;15:65 [FREE Full text] [doi: 10.1186/s12884-015-0492-5] [Medline: 25880644]
6. Das N, Topalovic M, Janssens W. Artificial intelligence in diagnosis of obstructive lung disease: current status and future potential. Curr Opin Pulm Med 2018 Mar;24(2):117-123 [FREE Full text] [doi: 10.1097/MCP.0000000000000459] [Medline: 29251699]
7. Ozkan IA, Koklu M, Sert IU. Diagnosis of urinary tract infection based on artificial intelligence methods. Comput Methods Programs Biomed 2018 Nov;166:51-59 [FREE Full text] [doi: 10.1016/j.cmpb.2018.10.007] [Medline: 30415718]
8. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]
9. Parajuli R, Doneys P. Exploring the role of telemedicine in improving access to healthcare services by women and girls in rural Nepal. Telematics and Informatics 2017 Nov;34(7):1166-1176. [doi: 10.1016/j.tele.2017.05.006]
10. Schölkopf B, Smola A, Müller KR. Kernel principal component analysis. In: Gerstner W, Germond A, Hasler M, Nicoud JD, editors. Artificial Neural Networks — ICANN'97. ICANN 1997. Lecture Notes in Computer Science, vol 1327. Berlin, Heidelberg: Springer; 1997:583-588.
11. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. Prog Retin Eye Res 2018 Nov;67:1-29 [FREE Full text] [doi: 10.1016/j.preteyeres.2018.07.004] [Medline: 30076935]
12. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
13. Dankwa-Mullan I, Rivo M, Sepulveda M, Park Y, Snowdon J, Rhee K. Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here. Popul Health Manag 2019 Jun;22(3):229-242 [FREE Full text] [doi: 10.1089/pop.2018.0129] [Medline: 30256722]

14. Park SW, Kim G, Hwang Y, Lee WJ, Park H, Kim JH. Validation of the effectiveness of a digital integrated healthcare platform utilizing an AI-based dietary management solution and a real-time continuous glucose monitoring system for diabetes management: a randomized controlled trial. BMC Med Inform Decis Mak 2020 Jul 10;20(1):156 [FREE Full text] [doi: 10.1186/s12911-020-01179-x] [Medline: 32650771]
15. Rigla M, García-Sáez G, Pons B, Hernando ME. Artificial Intelligence Methodologies and Their Application to Diabetes. J Diabetes Sci Technol 2018 Mar;12(2):303-310 [FREE Full text] [doi: 10.1177/1932296817710475] [Medline: 28539087]
16. Sriram RD, Reddy SSK. Artificial Intelligence and Digital Tools: Future of Diabetes Care. Clin Geriatr Med 2020 Aug;36(3):513-525 [FREE Full text] [doi: 10.1016/j.cger.2020.04.009] [Medline: 32586478]
17. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine Technologies for Diabetes in Pregnancy: A Systematic Review and Meta-Analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
18. Polak S, Mendyk A. Artificial intelligence technology as a tool for initial GDM screening. Expert Systems with Applications 2004 May;26(4):455-460. [doi: 10.1016/j.eswa.2003.10.005]
19. Lakshmi K. Modeling an Expert System for Diagnosis of Gestational Diabetes Mellitus Based On Risk Factors. IOSR-JCE 2013;8(3):29-32. [doi: 10.9790/0661-0832932]
20. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nat Med 2020 Jan;26(1):71-76 [FREE Full text] [doi: 10.1038/s41591-019-0724-8] [Medline: 31932807]
21. Martin B, Sacks DA. The global burden of hyperglycemia in pregnancy - Trends from studies in the last decade. Diabetes Res Clin Pract 2018 Nov;145:17-19 [FREE Full text] [doi: 10.1016/j.diabres.2018.04.003] [Medline: 29679621]
22. Dall TM, Yang W, Gillespie K, Mocarski M, Byrne E, Cintina I, et al. The Economic Burden of Elevated Blood Glucose Levels in 2017: Diagnosed and Undiagnosed Diabetes, Gestational Diabetes Mellitus, and Prediabetes. Diabetes Care 2019 Sep;42(9):1661-1668 [FREE Full text] [doi: 10.2337/dc18-1226] [Medline: 30940641]
23. Xu T, Dainelli L, Yu K, Ma L, Silva Zolezzi I, Detzel P, et al. The short-term health and economic burden of gestational diabetes mellitus in China: a modelling study. BMJ Open 2017 Dec 03;7(12):e018893 [FREE Full text] [doi: 10.1136/bmjopen-2017-018893] [Medline: 29203507]
24. Sosa-Rubi SG, Dainelli L, Silva-Zolezzi I, Detzel P, Espino Y Sosa S, Reyes-Muñoz E, et al. Short-term health and economic burden of gestational diabetes mellitus in Mexico: A modeling study. Diabetes Res Clin Pract 2019 Jul;153:114-124 [FREE Full text] [doi: 10.1016/j.diabres.2019.05.014] [Medline: 31108135]
25. Meregaglia M, Dainelli L, Banks H, Benedetto C, Detzel P, Fattore G. The short-term economic burden of gestational diabetes mellitus in Italy. BMC Pregnancy Childbirth 2018 Feb 23;18(1):58 [FREE Full text] [doi: 10.1186/s12884-018-1689-1] [Medline: 29471802]
26. Chen Y, Quick WW, Yang W, Zhang Y, Baldwin A, Moran J, et al. Cost of gestational diabetes mellitus in the United States in 2007. Popul Health Manag 2009 Jun;12(3):165-174 [FREE Full text] [doi: 10.1089/pop.2009.12303] [Medline: 19534581]
27. Liszka-Hackzell JJ. Prediction of blood glucose levels in diabetic patients using a hybrid AI technique. Comput Biomed Res 1999 Apr;32(2):132-144 [FREE Full text] [doi: 10.1006/cbmr.1998.1506] [Medline: 10337495]
28. Wang K, Adrian AM, Chen K, Wang K. An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus. J Biomed Inform 2015 Apr;54:220-229 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.001] [Medline: 25677947]

Abbreviations

ADA: American Diabetes Association
AI: artificial intelligence
AUC: area under the curve
AUROC: areas under the receiver operating characteristic
GBDT: gradient boosting decision tree
GDM: gestational diabetes mellitus
GDM-AI: AI-based diagnosis of GDM
IADPSG: International Association of Diabetes and Pregnancy Study Group
kNN: k-nearest neighbors
LR: logistic regression
NB: naïve Bayes
NPV: negative predictive value
PPV: positive predictive value.
ROC: receiver operating characteristic
SVM: support vector machine
XGBoost: eXtreme gradient boosting

Edited by G Eysenbach; submitted 18.06.20; peer-reviewed by A Al-Hasan, T Muto; comments to author 10.07.20; revised version received 23.07.20; accepted 27.07.20; published 15.09.20.

Please cite as:
Shen J, Chen J, Zheng Z, Zheng J, Liu Z, Song J, Wong SY, Wang X, Huang M, Fang PH, Jiang B, Tsang W, He Z, Liu T, Akinwunmi B, Wang CC, Zhang CJP, Huang J, Ming WK
An Innovative Artificial Intelligence–Based App for the Diagnosis of Gestational Diabetes Mellitus (GDM-AI): Development Study
J Med Internet Res 2020;22(9):e21573
URL: https://www.jmir.org/2020/9/e21573
doi:10.2196/21573
PMID:32930674

Review

Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy

Chang Seok Bang[1,2,3,4], MD, PhD; Jae Jun Lee[3,4,5], MD; Gwang Ho Baik[1,2], MD

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

[2]Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Republic of Korea

[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Republic of Korea

[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea

[5]Department of Anesthesiology and Pain Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

Corresponding Author:
Chang Seok Bang, MD, PhD
Department of Internal Medicine
Hallym University College of Medicine
Sakju-ro 77
Chuncheon,
Republic of Korea
Phone: 82 33 240 5000
Fax: 82 33 241 8064
Email: csbang@hallym.ac.kr

Abstract

Background: *Helicobacter pylori* plays a central role in the development of gastric cancer, and prediction of *H pylori* infection by visual inspection of the gastric mucosa is an important function of endoscopy. However, there are currently no established methods of optical diagnosis of *H pylori* infection using endoscopic images. Definitive diagnosis requires endoscopic biopsy. Artificial intelligence (AI) has been increasingly adopted in clinical practice, especially for image recognition and classification.

Objective: This study aimed to evaluate the diagnostic test accuracy of AI for the prediction of *H pylori* infection using endoscopic images.

Methods: Two independent evaluators searched core databases. The inclusion criteria included studies with endoscopic images of *H pylori* infection and with application of AI for the prediction of *H pylori* infection presenting diagnostic performance. Systematic review and diagnostic test accuracy meta-analysis were performed.

Results: Ultimately, 8 studies were identified. Pooled sensitivity, specificity, diagnostic odds ratio, and area under the curve of AI for the prediction of *H pylori* infection were 0.87 (95% CI 0.72-0.94), 0.86 (95% CI 0.77-0.92), 40 (95% CI 15-112), and 0.92 (95% CI 0.90-0.94), respectively, in the 1719 patients (385 patients with *H pylori* infection vs 1334 controls). Meta-regression showed methodological quality and included the number of patients in each study for the purpose of heterogeneity. There was no evidence of publication bias. The accuracy of the AI algorithm reached 82% for discrimination between noninfected images and posteradication images.

Conclusions: An AI algorithm is a reliable tool for endoscopic diagnosis of *H pylori* infection. The limitations of lacking external validation performance and being conducted only in Asia should be overcome.

Trial Registration: PROSPERO CRD42020175957; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=175957

(J Med Internet Res 2020;22(9):e21983) doi:10.2196/21983

KEYWORDS

artificial intelligence; convolutional neural network; deep learning; machine learning; endoscopy; Helicobacter pylori

Introduction

More than half of the world's population is infected with the *Helicobacter pylori* bacteria [1], which is associated with various disorders, such as gastritis, peptic ulcer, mucosa-associated lymphoid tissue lymphoma, gastric adenocarcinoma, and immune thrombocytopenic purpura [2,3]. The infection causes chronic atrophic gastritis, intestinal metaplasia, dysplasia, and gastric cancer in sequence [4]. The International Agency for Research on Cancer has categorized *H pylori* as a group 1 carcinogen [5]. Elimination of this pathogen is considered the most promising strategy for the prevention of gastric cancer [6,7].

An important aspect of endoscopy is the ability to predict *H pylori*–induced gastritis by visual inspection of the gastric mucosa to identify patients at high risk for gastric cancer. Representative features of *H pylori*–induced gastritis have been reported in the literature, including mucosal edema, atrophy, diffuse erythema, enlargement of mucosal folds, or mucosal nodularity [8,9]. The regular arrangement of collecting venules and fundic gland polyps has been suggested as a predictive marker of the *H pylori*–naïve stomach. Also, map-like redness under white-light imaging (WLI) or a cracked pattern under blue-laser imaging (BLI) have been suggested as features of a posteradicated gastric mucosa [8,9].

These endoscopic features do not have objective indicators, and there is the potential for interobserver or intraobserver variability in the optical diagnosis of *H pylori*–infected mucosa [10]. Although expert endoscopists might reliably identify an *H pylori* infection with meticulous visual inspection of the mucosa during endoscopic examination, novice endoscopists require substantial time to perform this task efficiently. Image-enhanced endoscopy (IEE), such as narrow-band imaging (NBI), BLI, or linked color imaging (LCI), with or without magnification, has been developed. Previous studies have indicated increased diagnostic accuracy of gastrointestinal neoplasms with the application of these modalities during endoscopic examination [11,12]. This also requires considerable training and prolonged procedure time. There are no uniform features of *H pylori* infection in IEE [12]. Therefore, there are currently no established methods of optical endoscopic diagnosis of *H pylori* infection. Definitive diagnosis continues to require endoscopic biopsy, which is categorized as an invasive diagnostic test.

Artificial intelligence (AI) has been increasingly adopted in clinical practice, especially for image recognition and classification [13]. This technique has shown promising diagnostic performance using endoscopic images, such as detecting cancer or neoplastic lesions and classifying neoplastic or nonneoplastic lesions in the gastrointestinal tract [14]. Application of AI in endoscopic examination is expected to be useful. It can help detect *H pylori* infection in real time and determine the optimum definitive test for *H pylori* infection. There has been no diagnostic test accuracy meta-analysis of AI for the prediction of *H pylori* infection using endoscopic images.

This study aimed to evaluate the diagnostic performance of AI for the diagnosis of *H pylori* infection using endoscopic images.

Methods

Ethics

This study adhered to the guidelines of the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) [15]. The protocol of this study was registered at the International Prospective Register of Systematic Reviews (PROSPERO) [CRD42020175957] on March 2019 before initiating the study. Approval of the institutional review board was exempted as only anonymized data was collected from the literature.

Literature Searching Strategy

Two independent evaluators (CSB and JJL) having published 23 systematic reviews and 11 PROSPERO protocols searched PubMed, Embase, and the Cochrane Library using common keywords relevant to *H pylori* infection and AI (inception to March 2020). The abstracts of all identified studies were reviewed to exclude irrelevant articles. Full-text reviews were conducted to determine whether the inclusion criteria were satisfied in all the studies. Bibliographies were also reviewed to identify additional relevant articles. Disagreements between the evaluators were resolved by consultation with a third evaluator (GHB). The details are presented in Multimedia Appendix 1.

Selection Criteria

We included studies that met the following criteria: (1) studies with endoscopic images of *H pylori* infection as a case group and endoscopic images without *H pylori* infection as a negative control group; (2) application of the AI algorithm for the prediction of *H pylori* infection; (3) inclusion of diagnostic performance indices of the AI algorithm, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), or accuracy, which enable an estimation of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values for the prediction of *H pylori* infection using endoscopic images; (4) prospective or retrospective study design; (5) human adult subjects; and (6) full-text publications written in English. The exclusion criteria included (1) narrative reviews; (2) letters, comments, editorials, or protocol studies; (3) guidelines; and (4) systematic reviews and meta-analyses. Studies meeting at least one of the exclusion criteria were excluded from the analysis.

Methodological Quality

The Quality Assessment of Diagnostic Accuracy Studies–2 (QUADAS-2) tool was used to determine the methodological quality of the included articles. This tool contains 4 domains: patient selection, index test, reference standard, and flow and timing [16]. Each domain was assessed in terms of high, low, or unclear risk of bias, and the first 3 domains were also assessed in terms of high, low, or unclear concerns regarding applicability [16]. Review Manager version 5.3.3 (RevMan for Windows 7, Nordic Cochrane Centre) was used to generate the summary figure of the methodological quality evaluation. Data extraction,

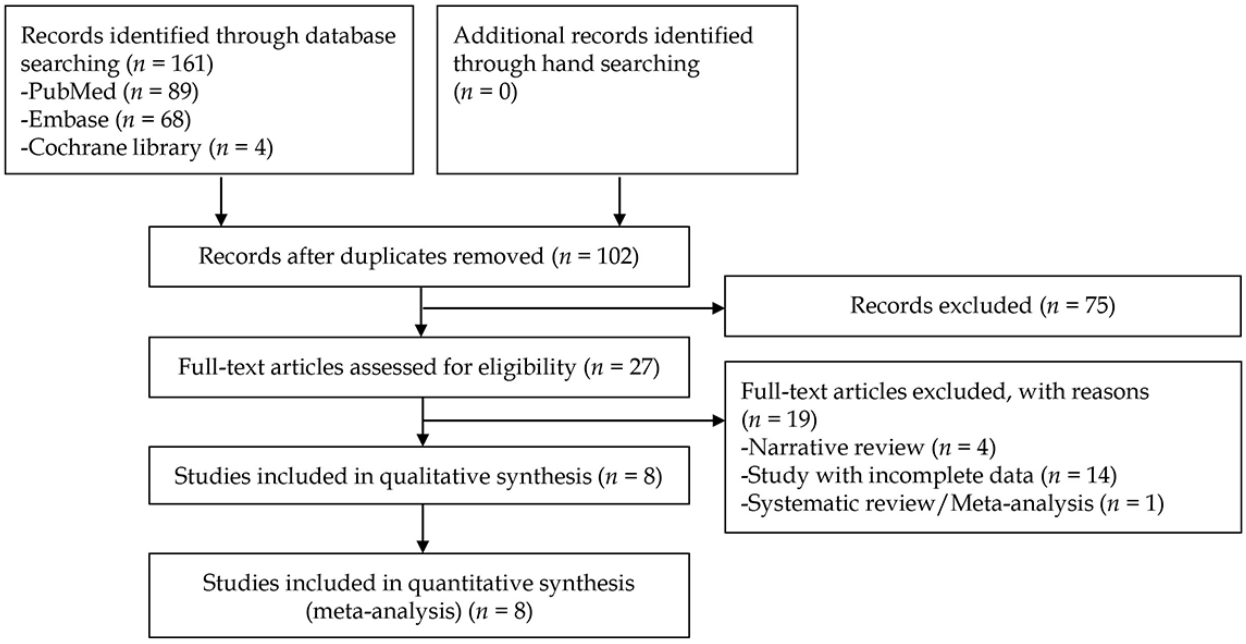
primary and modifier-based analyses, and statistical analysis are described in Multimedia Appendix 2 [17-20].

Results

Identification of Relevant Studies

In total, 161 articles were identified by searching 3 electronic databases. Among them, 59 were duplicate studies, and 75 were excluded during the initial screening by reviewing titles and abstracts. Full texts of the remaining 27 articles were thoroughly reviewed. Among these, 19 studies were excluded from the final analysis due to the following reasons: narrative review (n=4), incomplete data (n=14), and systematic review or meta-analysis (n=1; the topic of this systematic review was the role of nonmagnified endoscopy for the assessment of *H pylori* infection) [8]. The remaining 8 studies [9,10,21-26] were included in the final analysis. Figure 1 illustrates a flow diagram showing the process used to identify the relevant articles.

Figure 1. Flow diagram of the identification of relevant studies.



Characteristics of the Included Studies

The included studies could be categorized by analysis based on the number of enrolled patients [9,10,22,23,25,26] and number of enrolled images [9,10,21,24]. Two studies [9,10] presented both patient-based and image-based analyses. Enrolled studies presented performance of the AI algorithm with test dataset (internal validation), and there was no study that presented external validation performance.

Among the 8 studies [9,10,21-26] included for the prediction of *H pylori* infection using endoscopic images, we identified 1719 patients (385 patients with *H pylori* infection vs 1334 controls). Additionally, 2855 endoscopic images with *H pylori* infection and 2287 control images including 514 posteradicated images were identified.

Among the studies, 5 were retrospectively conducted [9,10,21,22,25], and 3 [23,24,26] were prospectively conducted.

All studies were conducted in Asia, and the age of the enrolled population ranged from a mean of 48.6 years to a median of 64 years. Most studies [9,21-24,26] established the AI algorithm based on the convolutional neural network (CNN), whereas 2 studies [10,25] established support vector machine (SVM)-based algorithms. Most studies [9,21,22,24-26] used endoscopic images with WLI, whereas a study by Yasuda et al [10] used endoscopic images with LCI, and Nakashima et al [23] used LCI and BLI images in addition to endoscopic images with WLI. While most studies [9,10,21,22,24-26] presented the performance of the AI algorithm as a single primary outcome, one study [23] also presented a feature map, which implies visualizing where established AI algorithms pay attention to and indicate a region of interest.

These characteristics (modifiers) were evaluated as potential sources of heterogeneity through the subgroup analysis and meta-regression. Detailed characteristics of the studies are presented in Table 1.

Table 1. Clinical characteristics of the included studies.

Study, format, nationality	Type of AI[a]	Type of endoscopy, diagnostic method of *Helicobacter pylori* infection	Number of cases in test dataset	Number of controls in test dataset	Age of patients in test dataset; gender in patients in test dataset (M/F[b])	TP[c]	FP[d]	FN[e]	TN[f]	Unit of analysis
Yasuda et al [10], retrospective, Japan	Support vector machine	LCI[g]; more than 2 different tests in each case (histology, serum antibody, stool antigen, urea breath test)	42 *H pylori* patients	63 controls (46 posteradication patients and 17 uninfected patients)	Median 64 years (range 26-88); (61/44)	38	9	4	54	Patient-based
—	—	—	210 *H pylori*–positive images	315 control images (230 posteradication and 85 uninfected images)	—	161	70	49	245	Image-based
—	—	—	210 *H pylori*–positive images	85 uninfected images (*H pylori*–naïve)	—	161	9	49	76	Image-based (infected vs uninfected)
—	—	—	210 *H pylori*–positive images	230 posteradication images	—	161	61	49	169	Image-based (infected vs after-eradication)
—	—	—	85 uninfected images	230 posteradication images	—	76	61	9	169	Image-based (uninfected vs after-eradication)
Zheng et al [21], retrospective, China	CNN[h]	WLI[i]; histology with immunohistochemistry (if negative, urea breath test was done)	2575 *H pylori*–positive images	1180 control images (whether posteradication or uninfected images is unknown)	Mean 48.6 years (SD 12.9); (220/232)	2359	17	216	1163	Image-based
Shichijo et al [9], retrospective, Japan	CNN	WLI; serum or urine antibody, stool antigen, urea breath test	70 *H pylori*–positive patients	777 controls (284 posteradication and 493 uninfected images)	—	44	47	26	730	Patient-based
—	—	—	59 *H pylori*–positive images	477 uninfected images (*H pylori*–naïve)	—	44	12	15	465	Image-based (infected vs uninfected)
—	—	—	55 *H pylori*–positive images	182 posteradication images	—	44	35	11	147	Image-based (infected vs after-eradication)
—	—	—	481 uninfected images	249 posteradication images	—	465	102	16	147	Image-based (uninfected vs after-eradication)
Nakashima et al [23], prospective, Japan	CNN	WLI; serum antibody (*H pylori* IgG ≥10 U/mL was considered positive)	30 *H pylori* patients	30 controls (uninfected patients; *H pylori*–naïve)	—	20	12	10	18	Patient-based
—	—	LCI	—	—	—	29	1	1	29	Patient-based
—	—	BLI[j]-bright	—	—	—	29	4	1	26	Patient-based

Study, format, nationality	Type of AI[a]	Type of endoscopy, diagnostic method of *Helicobacter pylori* infection	Number of cases in test dataset	Number of controls in test dataset	Age of patients in test dataset; gender in patients in test dataset (M/F[b])	TP[c]	FP[d]	FN[e]	TN[f]	Unit of analysis
Itoh et al [24], prospective, Japan	CNN	WLI; serum antibody (*H pylori* IgG ≥10 U/mL was considered positive)	15 *H pylori*–positive images	15 control images (uninfected patients; *H pylori*–naïve)	—	13	2	2	13	Image-based
Shichijo et al [22], retrospective, Japan	CNN	WLI; serum or urine antibody, stool antigen, urea breath test	72 *H pylori* patients	325 controls (uninfected patients; *H pylori*–naïve)	mean 50.4 (SD 11.2), (168/226)	64	41	8	284	Patient-based
Huang et al [25], retrospective, Taiwan	Sequential forward floating selection with SVM[k]	WLI; histology (3 pairs of samples from the topographic sites, including antrum, body, and cardia were obtained in a uniform way)	130 *H pylori* patients	106 controls (whether posteradication or uninfected patients is unknown)	—	128	21	2	85	Patient-based
Huang et al [26], prospective, Taiwan	Refined feature selection with neural network	WLI; histology (3 pairs of samples from the topographic sites, including antrum, body, and cardia were obtained in a uniform way)	41 *H pylori* patients	33 controls (whether posteradication or uninfected patients is unknown)	—	35	3	6	30	Patient-based

[a]AI: artificial intelligence.

[b]M/F: make/female.

[c]TP: true positive.

[d]FP: false positive.

[e]FN: false negative.

[f]TN: true negative.

[g]LCI: linked color imaging.

[h]CNN: convolutional neural network.

[i]WLI: white-light imaging.

[j]BLI: blue-laser imaging.

[k]SVM: support vector machine.

Methodological Quality of the Studies

Among the 8 studies [9,10,21-26] in the final analysis, 6 studies [9,10,21,22,25,26] showed low risk of bias, and 2 studies [23,24] showed high risk of bias in patient selection.

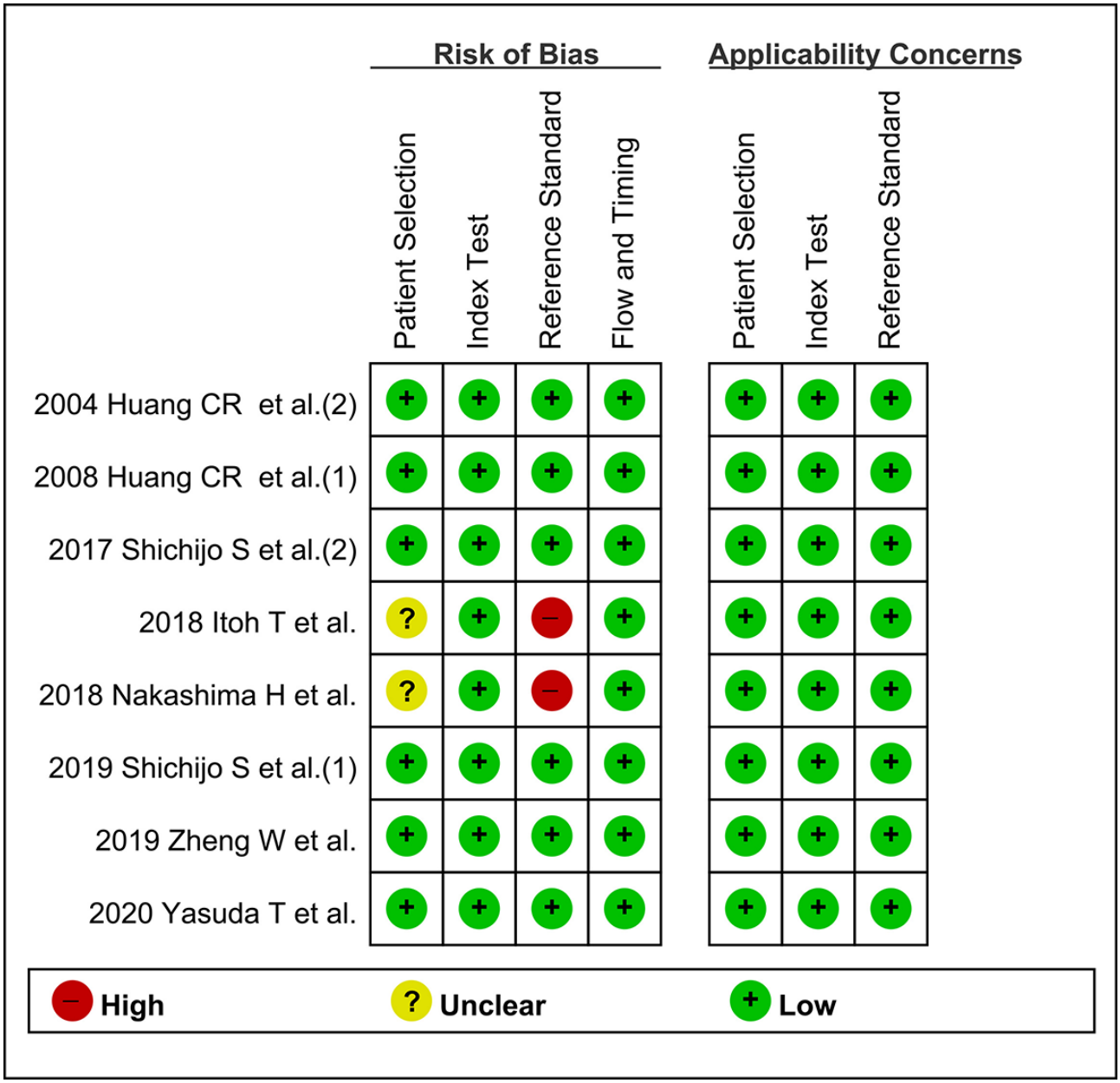
In terms of the patient selection, 4 studies [9,10,21,22] used multiple tests, including a biopsy, serology (serum anti–*H pylori* IgG titer), stool antigen test, urine examination (urine anti–*H pylori* IgG titer), or a urea breath test for the determination of *H pylori* infection. Two studies [25,26] used only gastric biopsy; however, 3 pairs of samples from the topographic sites, including the antrum, body, and cardia were obtained in a uniform way. The remaining 2 studies [23,24] used only serology (serum anti–*H pylori* IgG titer) for the determination of *H pylori* infection. Although a serology test is convenient and widely used in Japan, local validation is essential to determine the best cutoff values. A recent Cochrane review suggested that serology is less accurate for the diagnosis of *H pylori* infection compared with the urea breath test [27].

For concerns regarding image selection, most studies [9,10,21,22,25,26] did not limit the specific topographic area of the endoscopic still images for enrollment in the study. However, 2 studies [23,24] used still images limited to the lesser curvature of the stomach. Considering that topographic distribution and density of *H pylori* is different according to the stage of gastritis, the results of these studies may include a risk of bias.

Considering the commonly detected pitfalls in patient and image selection described above, these 2 studies [23,24] were rated as high risk in the patient selection domain in the risk of bias evaluation.

Overall, studies [23,24] with high risk in at least 1 of the 7 domains were rated as low methodological quality in the subgroup analysis (Figure 2).

Figure 2. Quality Assessment of Diagnostic Accuracy Studies–2 for the assessment of the methodological qualities of all the enrolled studies. (+) denotes low risk of bias, (?) denotes unclear risk of bias, (-) denotes high risk of bias.



Diagnostic Test Accuracy of Artificial Intelligence for the Prediction of Helicobacter pylori Infection

Among the 6 studies [9,10,22,23,25,26] of patient-based analysis, the sensitivity, specificity, PLR, NLR, DOR, and area under the curve (AUC) with 95% CI of AI for the prediction of *H pylori* infection were 0.87 (95% CI 0.72-0.94), 0.86 (95% CI 0.77-0.92), 6.2 (95% CI 3.8-10.1), 0.15 (95% CI 0.07-0.34), 40 (95% CI 15-112), and 0.92 (95% CI 0.90-0.94), respectively (Table 2, Figure 3). The SROC curve, with a 95% confidence region and prediction region, is illustrated in Figure 4. To investigate the clinical utility of AI, a Fagan nomogram was generated. Assuming 50% prevalence of *H pylori* infection, the Fagan nomogram shows that the posterior probability of *H pylori* infection was 86% if the test was positive, and the posterior probability of absence of *H pylori* infection was 13% if the test was negative (Figure 5).

Table 2. Summary of diagnostic test accuracy and subgroup analysis of the included studies with patient-based analysis.

Subgroup	Number of included studies	Sensitivity (95% CI)	Specificity (95% CI)	PLR[a]	NLR[b]	DOR[c]	AUC[d]
Value of meta-analysis in all included studies	6	0.87 (0.72-0.94)	0.86 (0.77-0.92)	6.2 (3.8-10.1)	0.15 (0.07-0.34)	40 (15-112)	0.92 (0.90-0.94)
Methodological quality of included studies[e]							
High quality	5	0.89 (0.75-0.96)	0.88 (0.83-0.92)	7.7 (5.6-10.6)	0.12 (0.05-0.28)	64 (32-129)	0.94 (0.91-0.95)
Low quality	1	Null	Null	Null	Null	Null	Null
Total number of included patients[e]							
≤100	4	0.90 (0.73-0.97)	0.88 (0.81-0.93)	7.6 (5.3-10.9)	0.11 (0.04-0.32)	68 (29-158)	0.94 (0.91-0.95)
<100	2	Null	Null	Null	Null	Null	Null
Format of study							
Retrospective	4	0.90 (0.73-0.97)	0.88 (0.81-0.93)	7.6 (5.3-10.9)	0.11 (0.04-0.32)	68 (29-158)	0.94 (0.91-0.95)
Prospective	2	Null	Null	Null	Null	Null	Null
Published year							
After 2010	4	0.80 (0.64-0.90)	0.86 (0.73-0.93)	5.6 (2.8-11.3)	0.24 (0.13-0.45)	23 (8-72)	0.90 (0.87-0.92)
Before 2010	2	Null	Null	Null	Null	Null	Null
Type of AI[f]							
Neural network–based	4	0.78 (0.64-0.87)	0.87 (0.74-0.94)	6.0 (2.7-13.0)	0.26 (0.15-0.44)	23 (7-73)	0.89 (0.86-0.91)
SVM[g]-based	2	Null	Null	Null	Null	Null	Null
Type of endoscopic image							
WLI[h]	5	0.86 (0.67-0.95)	0.86 (0.75-0.92)	6.1 (3.4-10.9)	0.16 (0.06-0.42)	37 (11-124)	0.92 (0.89-0.94)
LCI[i]	1	Null	Null	Null	Null	Null	Null
Classifying performance between *Helicobacter pylori*–positive vs *H pylori*–naïve patients	2	0.82 (0.74-0.89)	0.85 (0.81-0.89)	3.5 (0.8-14.3)	0.27 (0.05-1.41)	13 (0.8-229)	Null

[a]PLR: positive likelihood ratio.

[b]NLR: negative likelihood ratio.

[c]DOR: diagnostic odds ratio.

[d]AUC: area under the curve.

[e]These modifiers were significant in the meta-regression analysis.

[f]AI: artificial intelligence.

[g]SVM: support vector machine.

[h]WLI: white-light imaging.

[i]LCI: linked color imaging.

Figure 3. Forest plots of sensitivity and specificity of artificial intelligence algorithm for the prediction of Helicobacter pylori infection in endoscopic images.

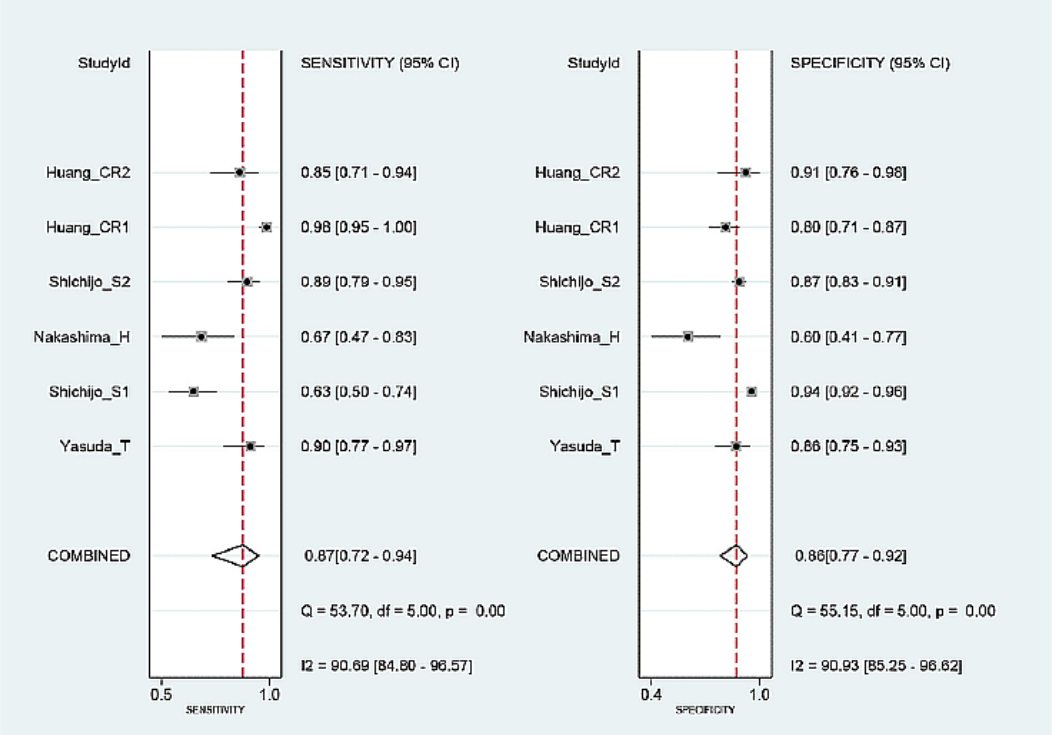


Figure 4. Summary receiver operating characteristic curve with 95% confidence region and prediction region for the prediction of Helicobacter pylori infection in endoscopic images.

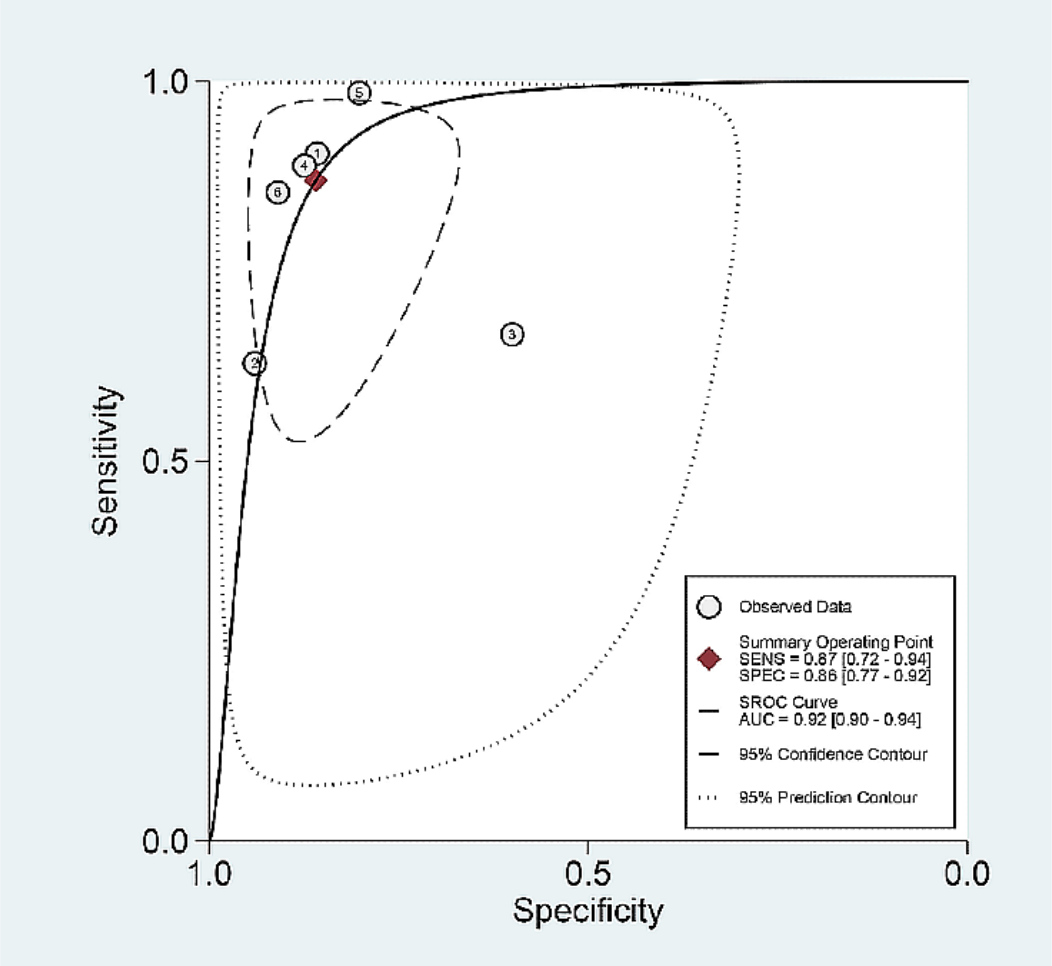
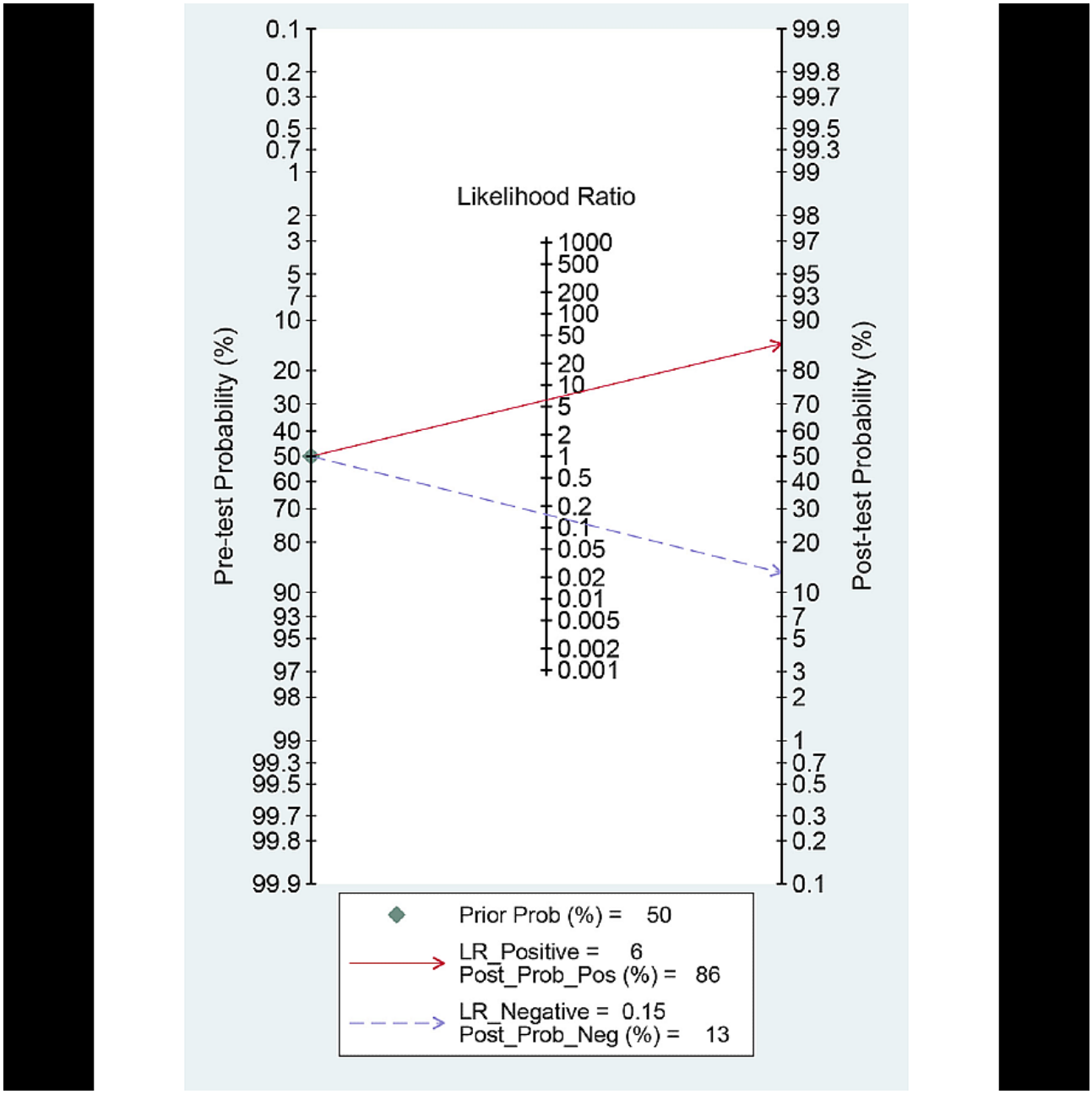


Figure 5. Fagan normogram for the prediction of Helicobacter pylori infection in endoscopic images.



Among the 4 studies [9,10,21,24] of image-based analysis, sensitivity, specificity, PLR, NLR, DOR, and AUC with 95% CI of AI for the prediction of *H pylori* infection were 0.81 (95% CI 0.68-0.90), 0.93 (95% CI 0.82-0.98), 12.3 (95% CI 3.8-39.2), 0.20 (95% CI 0.11-0.38), 61 (95% CI 11-322), and 0.93 (95% CI 0.90-0.95), respectively (Table 3).

Only 2 studies [9,10] reported outcomes related to discrimination between noninfected images and posteradication images. Therefore, a meta-analysis was not possible. Pooled analysis of the crude value of TP, FP, FN, and TN revealed that accuracy of the AI algorithm reached 82.01% (857/1045).

Additionally, only 2 studies [9,10] reported outcomes regarding discrimination between images showing *H pylori* infection and posteradication images. Therefore, a meta-analysis was not possible. However, pooled analysis of the crude value of TP, FP, FN, and TN revealed that accuracy of the AI algorithm reached 77.0% (521/677).

Regarding comparison of the performance between AI and endoscopists, only 2 studies presented outcomes [10,22]. In the study by Yasuda et al [10], the diagnostic accuracy of an SVM-based AI algorithm was superior to that of inexperienced endoscopists. However, there was no significant difference between experienced endoscopists and the AI algorithm [10]. The accuracy of a CNN-based AI algorithm reached 87.7% in the study by Shichijo et al [22], while the accuracy achieved by endoscopists was 82.4%. The difference was statistically significant between the AI algorithm and endoscopists (5.3%, 95% CI 0.3-10.2) [22].

Exploring Heterogeneity With Meta-Regression and Subgroup Analysis

For the prediction of *H pylori* infection using endoscopic images, the SROC curve was generated in the patient-based studies. The shape of the curve was symmetric (Figure 4). We observed a negative correlation coefficient between logit transformed sensitivity and specificity (–0.22) and an asymmetric parameter, β, with a nonsignificant *P* value (*P*=.29) indicating no heterogeneity among the studies. However, the 95% prediction region in the SROC curve was wide, and the methodological quality among the included studies (*P*<.001) and total number of included patients (*P*=.03) were found to be the source of heterogeneity in the joint model of meta-regression (published year [*P*=.41], study format [*P*=.10], type of endoscopic image [*P*=.92], and type of AI [*P*=.07]; Figure 6). Subgroup analyses, based on the modifiers of heterogeneity, showed higher AUCs or DORs in studies with a large population of patients (≤100) or those demonstrating high methodological quality (Table 2).

In terms of the image-based analysis, the overall number of included studies was 4, and subgroup analysis was possible with only 3 studies. Studies with CNN (vs SVM) and studies with WLI (vs LCI) showed higher AUCs or DORs (Table 3). However, these modifiers (type of AI and type of endoscopic imaging) were not a significant covariate in the meta-regression analysis (total number of included patients [*P*=.06], methodological quality [*P*=.68], published year [*P*=.78], study format [*P*=.68], type of endoscopic image [*P*=.72], or type of AI [*P*=.72]).

The enrolled studies included various types of control groups. The fundamental question of this study was whether the AI algorithm could differentiate endoscopic images between an *H pylori*–positive and a naïve gastric mucosa. Table 1 shows the types of control group included in each study. Two studies clearly presented the classifying performance of an AI algorithm discriminating *H pylori*–positive and *H pylori*–naïve in a patient-based analysis, and there were 3 with image-based analysis. Subgroup analysis was also performed and showed slightly lower AUCs or DORs in patient-based or image-based analysis (Table 2 and 3). However, this factor (studies with clearly presented classifying performance data discriminating *H pylori*–positive and *H pylori*–naïve group) was not a significant modifier in the meta-regression analysis (*P*=.21 in the patient-based analysis, and *P*=.10 in the image-based analysis).

Figure 6. Meta-regression for the reason of heterogeneity in the diagnostic test accuracy meta-analysis. nopt: number of patients.

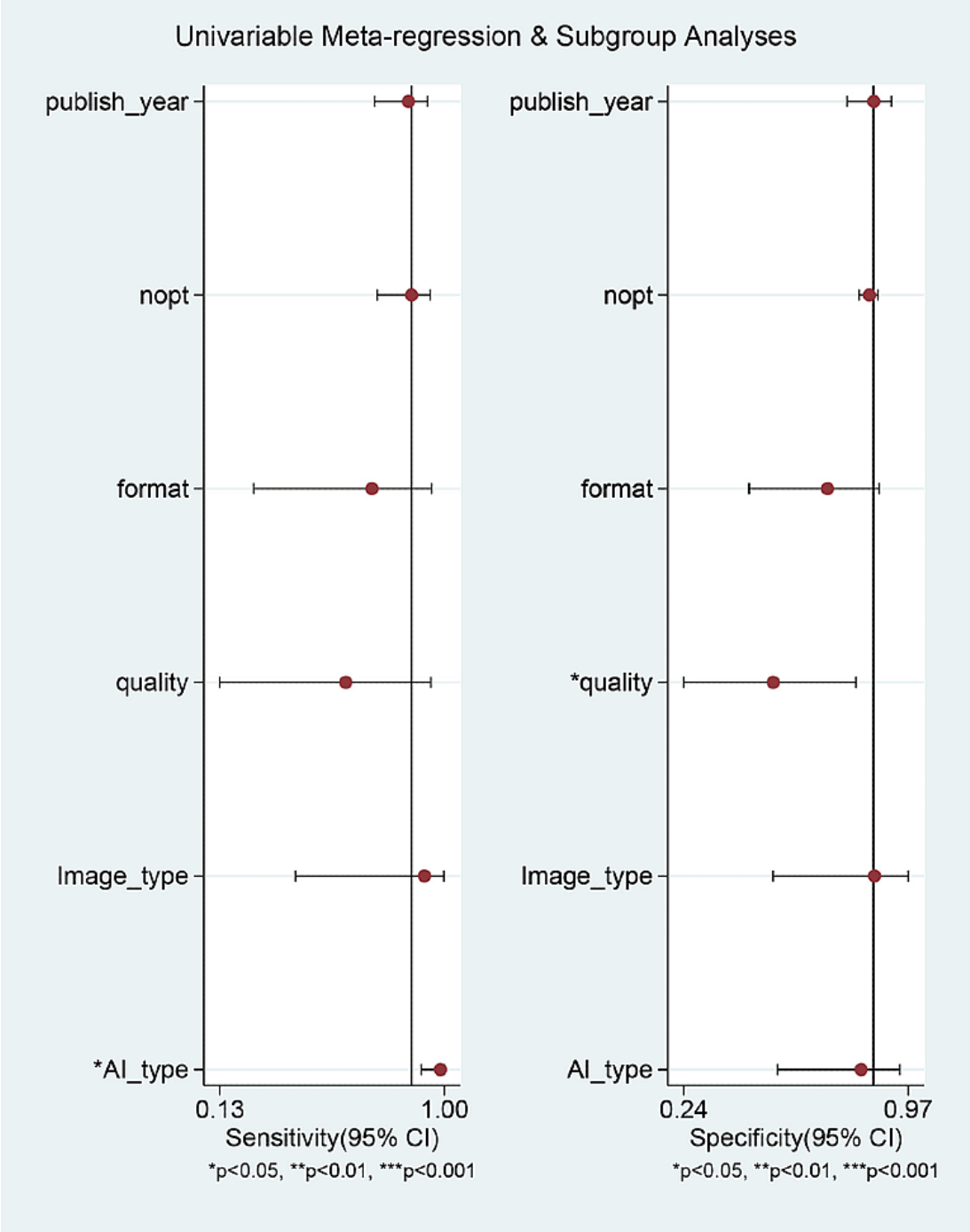


Table 3. Summary of diagnostic test accuracy and subgroup analysis of the included studies with image-based analysis.

Subgroup	Number of included studies	Sensitivity (95% CI)	Specificity (95% CI)	PLR[a]	NLR[b]	DOR[c]	AUC[d]
Value of meta-analysis in all the included (bivariate and HSROC[e] method)	4	0.81 (0.68-0.90)	0.93 (0.82-0.98)	12.3 (3.8-39.2)	0.20 (0.11-0.38)	61 (11-322)	0.93 (0.90-0.95)
Value of meta-analysis in all the included (Moses-Shapiro-Littenberg method)		0.90 (0.89-0.91)	0.94 (0.93-0.95)	11.1 (1.6-76.2)	0.20 (0.08-0.52)	56 (5-591)	0.90 (0.71-0.99)
Methodological quality of included studies							
High quality	3	0.90 (0.87-0.91)	0.94 (0.93-0.95)	13.1 (1.4-124.5)	0.22 (0.08-0.62)	61 (4-919)	0.87 (0.43-0.99)
Low quality	1	Null	Null	Null	Null	Null	Null
Total number of included patients							
≤100	3	0.90 (0.87-0.91)	0.94 (0.93-0.95)	13.1 (1.4-124.5)	0.22 (0.08-0.62)	61 (4-919)	0.87 (0.43-0.99)
<100	1	Null	Null	Null	Null	Null	Null
Format of study							
Retrospective	3	0.90 (0.87-0.91)	0.94 (0.93-0.95)	13.1 (1.4-124.5)	0.22 (0.08-0.62)	61 (4-919)	0.87 (0.43-0.99)
Prospective	1	Null	Null	Null	Null	Null	Null
Published year							
After 2010	4	0.90 (0.89-0.91)	0.94 (0.93-0.95)	11.1 (1.6-76.2)	0.20 (0.08-0.52)	56 (5-591)	0.90 (0.71-0.99)
Before 2010	0						
Type of AI[f]							
Neural network–based	3	0.91 (0.90-0.92)	0.97 (0.96-0.97)	16.8 (2.0-141.7)	0.17 (0.05-0.61)	98 (6-1640)	0.95 (0.75-0.99)
SVM[g]–based	1	Null	Null	Null	Null	Null	Null
Type of endoscopic image							
WLI[h]	3	0.91 (0.90-0.92)	0.97 (0.96-0.97)	16.8 (2.0-141.7)	0.17 (0.05-0.61)	98 (6-1640)	0.95 (0.75-0.99)
LCI[i]	1	Null	Null	Null	Null	Null	Null
Classifying performance between *Helicobacter pylori*–positive vs *H pylori*–naïve images	3	0.77 (0.71-0.82)	0.96 (0.94-0.98)	11.8 (3.7-38.3)	0.26 (0.21-0.32)	53 (17-161)	0.88 (0.79-0.96)

[a]PLR: positive likelihood ratio.

[b]NLR: negative likelihood ratio.

[c]DOR: diagnostic odds ratio.

[d]AUC: area under the curve.

[e]HSROC: hierarchical summary receiver operating characteristic.

[f]AI: artificial intelligence.

[g]SVM: support vector machine.

[h]WLI: white-light imaging.

[i]LCI: linked color imaging.

Publication Bias

Figure 7 shows the Deek funnel plot of studies of patient-based analysis and Figure 8 shows the Deek funnel plot of studies of image-based analysis. The plot was grossly symmetrical with respect to the regression line. The Deek funnel plot asymmetry test showed no evidence of publication bias (P=.38 in the patient-based analysis, and P=.27 in the image-based analysis).

Figure 7. Deek funnel plot for the studies of patient-based analysis.

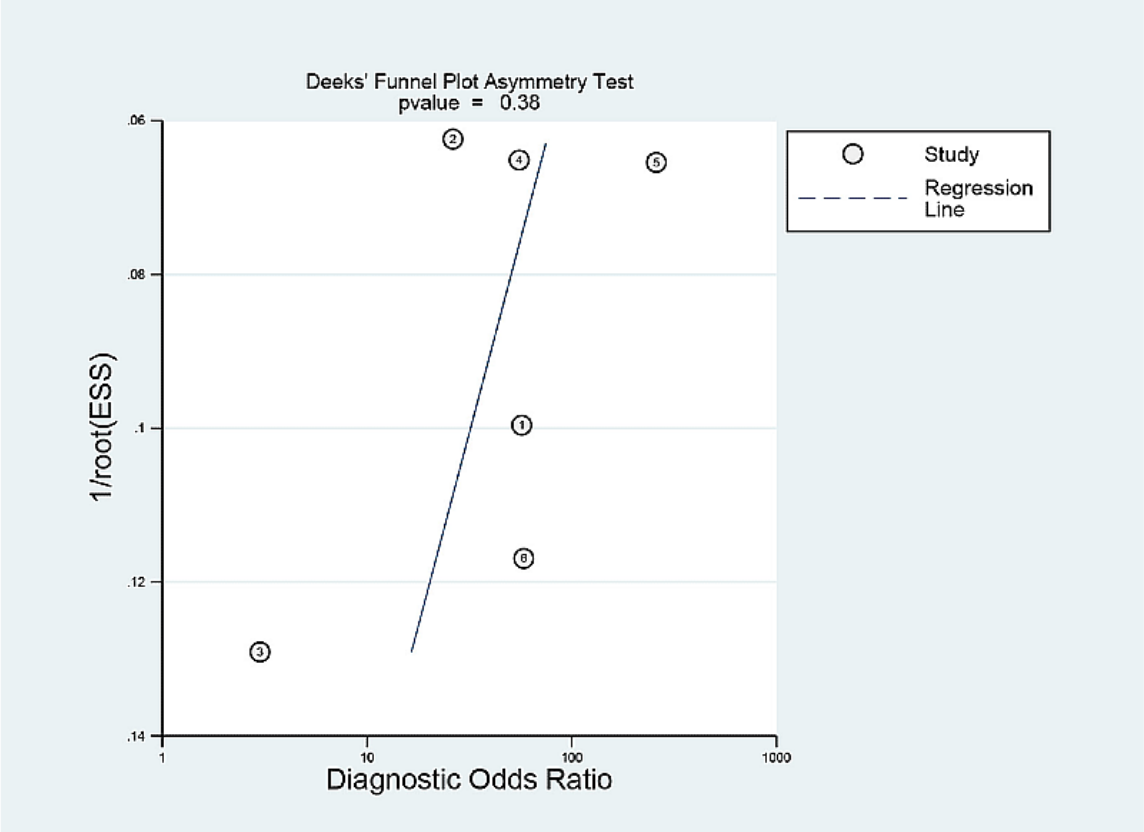
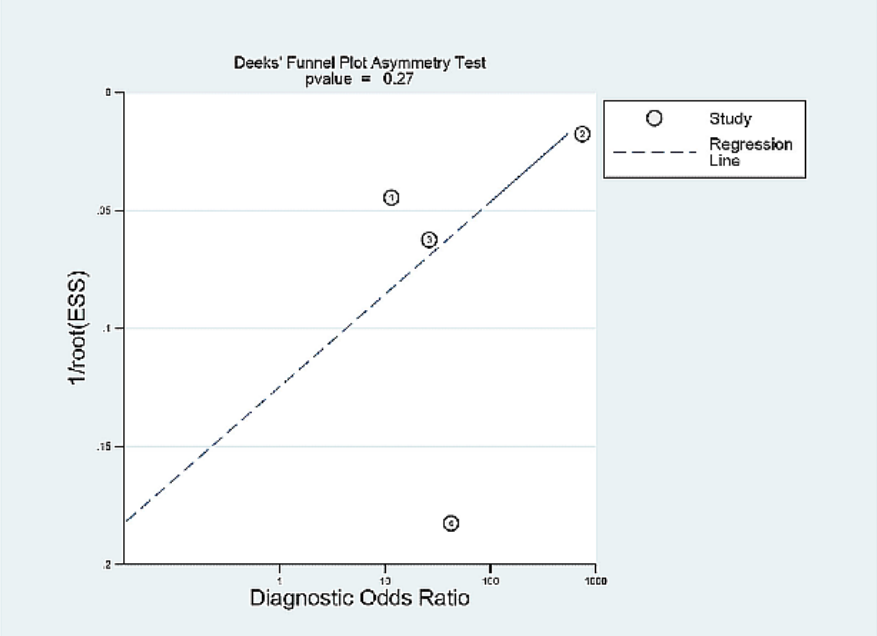


Figure 8. Deek funnel plot for the studies of image-based analysis.



Discussion

Principal Findings

This study presented the good performance of the AI algorithm applied to endoscopic diagnosis of *H pylori* infection, indicating that AI-assisted endoscopy is feasible in clinical practice. Indeed, this approach might be characterized as a computer-aided diagnosis, and the most important benefit consists of the improvement in diagnostic accuracy of conventional endoscopy with WLI [28]. Optical endoscopic diagnosis has operator-dependent characteristics, and the diagnostic process is completely subjective. However, AI-assisted endoscopy could be helpful in providing a second opinion and may help avoid operator dependency in diagnostic endoscopy [28]. Currently, it is unclear how endoscopists would react to a diagnosis made using AI (examples from the literature include approval, a learning opportunity, or "presenting an indolent attitude") [28,29]. Therefore, a prospective study based on the application of AI in clinical practice (more specifically, in diagnostic endoscopy) is essential [30,31]. However, providing robust answers using an AI algorithm irrespective of the endoscopists' inspection would be helpful to increase the likelihood of identifying important findings in diagnostic endoscopy. As endoscopic biopsy is an invasive procedure, application of a highly accurate AI algorithm in endoscopic examination may reduce the need for unnecessary biopsies in a substantial proportion of patients.

Another important finding of this study is the robustness of the diagnostic performance of the AI algorithm, irrespective of the modifiers detected during the systematic review process. Although studies based on a large population of patients presenting high methodological quality demonstrated higher diagnostic performance, this difference in diagnostic performance was not substantial. Neither the type of AI, such as CNN or SVM, nor the type of endoscopic images used, such as WLI, LCI, or BLI, affected overall diagnostic performance. Studies with patient-based analysis and image-based analysis commonly presented a good performance of AI for the diagnosis of *H pylori* infection (Tables 2 and 3).

AI is generally characterized as being of a black-box nature due to the difficulty in explaining the determination of the AI algorithm. The class activation map is a technique for visualizing the locations to which established AI algorithms pay attention and indicating a region of interest. This technique offers the possibility of explaining the determination of the AI algorithm. Although only one study [23] included in this systematic review adopted this type of feature map with the AI algorithm, this technique has now been widely adopted for the establishment of the AI algorithm and could be useful for the work of endoscopists, specifically for targeted biopsy in *H pylori* detection.

In terms of the IEE, the ultimate goal of this technique would be optical biopsy replacing invasive histologic examination with the aid of discrete differentiation and enhancement of surface mucosal features. Previous studies on the diagnosis of *H pylori* infection with WLI showed low sensitivity and poor interobserver agreement [11,32-34]. However, studies with IEE commonly showed increased diagnostic accuracy of premalignant or malignant lesions during endoscopic examination [11,12]. Previous studies with IEE also indicated the usefulness of LCI for the diagnosis of *H pylori* infection [35,36]. Although a recently published systematic review concluded that currently no established uniform findings exist for optical endoscopic diagnosis of *H pylori* infection [8], IEE continues to have potential for the differentiation of *H pylori* infection. The development of standardized validated indicators is required. The additive effect of magnifying endoscopy in NBI also showed promising results for the diagnosis of *H pylori* infection [37,38]. Due to insufficient data on IEE for the application of AI in this study, the real value of IEE with AI could not be evaluated. Further studies using various types of IEE with AI applications is essential.

Limitations

Although, this review rigorously investigated the diagnostic accuracy of the AI algorithm for *H pylori* infection in endoscopic images, our analysis has several inevitable limitations originating from potential bias in each study. First, the diagnostic performance of AI could have been exaggerated. It is more likely that the endoscopic images in each included study may have distinct features of *H pylori* infection and a clear and focused view, leading to a selection bias [28]. Second, the overfitting (modeling error that occurs when a certain learning model is excessively tailored to the training dataset and predictions are not well generalized to new datasets) of the AI algorithm cannot be excluded [31]. The diagnostic performance of the AI algorithms can only be valid for the population under evaluation and depends on the prevalence of target conditions for the selected population (so-called spectrum bias or class imbalance). The best and only way to prove the real performance of an AI algorithm is external (prospective) validation using unused datasets for model development, collected in a way that minimizes the spectrum bias [31]. However, there is no single study that adopted external validation for the performance of an established AI algorithm in this systematic review. Moreover, all the enrolled studies were conducted at a single center, which limits the generalization of the results. Third, there were little data regarding posteradication images, thus increasing the difficulty of the analysis of performance in the discrimination of uninfected and posteradicated images of *H pylori* infection. In real clinical practice, patients are not divided into only 2 categories of infected or noninfected patients. Indeed, there are many posteradicated patients, and this aspect should be reflected in the establishment of an AI algorithm. However, only 2 studies considered this category and conducted a separate analysis [9,10]. Because there were only 4 studies that conducted multiple tests in enrolling *H pylori*–infected patients, there may be a concern for selection bias. However, this factor is not expected to affect the overall results because there is a high probability of actual infection if any type of test is positive. Moreover, this factor was reflected in the methodological quality, and authors verified the effect of this bias through additional meta-regression. All the included studies were conducted in Asia, and no study confirmed the diagnostic validity of AI using external validation. Since the age of the

enrolled population ranged from a mean of 48.6 years to a median of 64 years, excluding a younger population, further studies are required to understand the real value of the widespread use of this algorithm. Considering the high accuracy and real-time diagnostic characteristics, the results of this study indicate the clinical utility of using an AI algorithm as an additive tool for the prediction of *H pylori* infection during endoscopic procedures. It is highly likely that AI could replace endoscopists' diagnoses of *H pylori* infections as guessed by visual inspection based on the evidence of this study. The real potential would be elucidated through the clinical application studies.

Conclusion

In conclusion, an AI algorithm can be considered a reliable tool for endoscopic diagnosis of *H pylori* infection. The limitations of lacking external validation performance and being conducted only in Asia should be overcome.

Acknowledgments

Funding for this research was provided by the Bio & Medical Technology Development Program of the National Research Foundation and the Korean government, Ministry of Science and ICT (grant number NRF2017M3A9E8033253).

Authors' Contributions

CSB was responsible for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing the original draft, and reviewing and editing the final draft. JJL was responsible for data curation, formal analysis, investigation, and resources. GHB was responsible for data curation, formal analysis, investigation, and resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used to find relevant articles.
[DOCX File , 20 KB - jmir_v22i9e21983_app1.docx]

Multimedia Appendix 2
Data extraction, primary- and modifier-based analyses, and statistical analysis.
[DOCX File , 23 KB - jmir_v22i9e21983_app2.docx]

References

1. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global prevalence of Helicobacter pylori infection: systematic review and meta-analysis. Gastroenterology 2017 Aug;153(2):420-429 [FREE Full text] [doi: 10.1053/j.gastro.2017.04.022] [Medline: 28456631]
2. Chey WD, Leontiadis GI, Howden CW, Moss SF. ACG clinical guideline: treatment of Helicobacter pylori infection. Am J Gastroenterol 2017 Feb;112(2):212-239. [doi: 10.1038/ajg.2016.563] [Medline: 28071659]
3. Bang CS, Lee JJ, Baik GH. The most influential articles in Helicobacter pylori research: a bibliometric analysis. Helicobacter 2019 Aug;24(4):e12589. [doi: 10.1111/hel.12589] [Medline: 31033071]
4. Correa P. A human model of gastric carcinogenesis. Cancer Res 1988 Jul 01;48(13):3554-3560 [FREE Full text] [Medline: 3288329]
5. IARC monographs on the evaluation of carcinogenic risks to humans. In: International Agency for Research on Cancer, ed Shistosomes, Liver Flukes and Helicobacter pylori. Vol. 61. Lyon: International Agency for Research on Cancer; 1994.
6. IARC working group reports. In: International Agency for Research on Cancer, ed Helicobacter pylori Eradication as a Strategy for Preventing Gastric Cancer. Vol. 8. Lyon: International Agency for Research on Cancer; 2014.
7. Bang CS, Baik GH, Shin IS, Kim JB, Suk KT, Yoon JH, et al. Helicobacter pylori eradication for prevention of metachronous recurrence after endoscopic resection of early gastric cancer. J Korean Med Sci 2015 Jun;30(6):749-756 [FREE Full text] [doi: 10.3346/jkms.2015.30.6.749] [Medline: 26028928]
8. Glover B, Teare J, Patel N. A systematic review of the role of non-magnified endoscopy for the assessment of infection. Endosc Int Open 2020 Feb;8(2):E105-E114 [FREE Full text] [doi: 10.1055/a-0999-5252] [Medline: 32010741]
9. Shichijo S, Endo Y, Aoyama K, Takeuchi Y, Ozawa T, Takiyama H, et al. Application of convolutional neural networks for evaluating Helicobacter pylori infection status on the basis of endoscopic images. Scand J Gastroenterol 2019 Feb;54(2):158-163. [doi: 10.1080/00365521.2019.1577486] [Medline: 30879352]
10. Yasuda T, Hiroyasu T, Hiwa S, Okada Y, Hayashi S, Nakahata Y, et al. Potential of automatic diagnosis system with linked color imaging for diagnosis of Helicobacter pylori infection. Dig Endosc 2020 Mar;32(3):373-381. [doi: 10.1111/den.13509] [Medline: 31398276]

11. Dohi O, Majima A, Naito Y, Yoshida T, Ishida T, Azuma Y, et al. Can image-enhanced endoscopy improve the diagnosis of Kyoto classification of gastritis in the clinical setting? Dig Endosc 2020 Jan;32(2):191-203. [doi: 10.1111/den.13540] [Medline: 31550395]
12. Kim J. Usefulness of narrow-band imaging in endoscopic submucosal dissection of the stomach. Clin Endosc 2018 Oct;51(6):527-533 [FREE Full text] [doi: 10.5946/ce.2018.186] [Medline: 30453446]
13. Cho B, Bang CS. Artificial intelligence for the determination of a management strategy for diminutive colorectal polyps: hype, hope, or help. Am J Gastroenterol 2020 Jan;115(1):70-72. [doi: 10.14309/ajg.0000000000000476] [Medline: 31770118]
14. Cho B, Bang CS, Park SW, Yang YJ, Seo SI, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019 Dec;51(12):1121-1129. [doi: 10.1055/a-0981-6133] [Medline: 31443108]
15. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, PRISMA-DTA Group, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 2018 Jan 23;319(4):388-396. [doi: 10.1001/jama.2017.19163] [Medline: 29362800]
16. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
17. Bang CS, Lee JJ, Baik GH. Prediction of chronic atrophic gastritis and gastric neoplasms by serum pepsinogen assay: a systematic review and meta-analysis of diagnostic test accuracy. J Clin Med 2019 May 10;8(5):657 [FREE Full text] [doi: 10.3390/jcm8050657] [Medline: 31083485]
18. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005 Oct;58(10):982-990. [doi: 10.1016/j.jclinepi.2005.02.022] [Medline: 16168343]
19. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001 Oct 15;20(19):2865-2884. [doi: 10.1002/sim.942] [Medline: 11568945]
20. Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. The Stata Journal 2018 Nov 19;9(2):211-229. [doi: 10.1177/1536867x0900900203]
21. Zheng W, Zhang X, Kim JJ, Zhu X, Ye G, Ye B, et al. High accuracy of convolutional neural network for evaluation of Helicobacter pylori infection based on endoscopic images: preliminary experience. Clin Transl Gastroenterol 2019 Dec;10(12):e00109 [FREE Full text] [doi: 10.14309/ctg.0000000000000109] [Medline: 31833862]
22. Shichijo S, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, et al. Application of convolutional neural networks in the diagnosis of Helicobacter pylori infection based on endoscopic images. EBioMedicine 2017 Nov;25:106-111 [FREE Full text] [doi: 10.1016/j.ebiom.2017.10.014] [Medline: 29056541]
23. Nakashima H, Kawahira H, Kawachi H, Sakaki N. Artificial intelligence diagnosis of infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. Ann Gastroenterol 2018;31(4):462-468 [FREE Full text] [doi: 10.20524/aog.2018.0269] [Medline: 29991891]
24. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. Endosc Int Open 2018 Feb;6(2):E139-E144 [FREE Full text] [doi: 10.1055/s-0043-120830] [Medline: 29399610]
25. Huang C, Chung P, Sheu B, Kuo H, Popper M. Helicobacter pylori-related gastric histology classification using support-vector-machine-based feature selection. IEEE Trans Inf Technol Biomed 2008 Jul;12(4):523-531. [doi: 10.1109/TITB.2007.913128] [Medline: 18632332]
26. Huang C, Sheu B, Chung P, Yang H. Computerized diagnosis of Helicobacter pylori infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network. Endoscopy 2004 Jul;36(7):601-608. [doi: 10.1055/s-2004-814519] [Medline: 15243882]
27. Best LM, Takwoingi Y, Siddique S, Selladurai A, Gandhi A, Low B, et al. Non-invasive diagnostic tests for Helicobacter pylori infection. Cochrane Database Syst Rev 2018 Mar 15;3:CD012080 [FREE Full text] [doi: 10.1002/14651858.CD012080.pub2] [Medline: 29543326]
28. Hoogenboom SA, Bagci U, Wallace MB. Artificial intelligence in gastroenterology. The current state of play and the potential. How will it affect our practice and when? Tech Innov Gastrointest Endosc 2020 Apr;22(2):42-47. [doi: 10.1016/j.tgie.2019.150634]
29. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):e17620 [FREE Full text] [doi: 10.2196/17620] [Medline: 32406857]
30. Tian Y, Liu X, Wang Z, Cao S, Liu Z, Ji Q, et al. Concordance between Watson for oncology and a multidisciplinary clinical decision-making team for gastric cancer and the prognostic implications: retrospective study. J Med Internet Res 2020 Feb 20;22(2):e14122 [FREE Full text] [doi: 10.2196/14122] [Medline: 32130123]
31. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]
32. Bah A, Saraga E, Armstrong D, Vouillamoz D, Dorta G, Duroux P, et al. Endoscopic features of Helicobacter pylori-related gastritis. Endoscopy 1995 Oct;27(8):593-596. [doi: 10.1055/s-2007-1005764] [Medline: 8608753]

33. Laine L, Cohen H, Sloane R, Marin-Sorensen M, Weinstein WM. Interobserver agreement and predictive value of endoscopic findings for H. pylori and gastritis in normal volunteers. Gastrointest Endosc 1995 Nov;42(5):420-423. [doi: 10.1016/s0016-5107(95)70043-9] [Medline: 8566631]
34. Redéen S, Petersson F, Jönsson K, Borch K. Relationship of gastroscopic features to histological findings in gastritis and Helicobacter pylori infection in a general population sample. Endoscopy 2003 Nov;35(11):946-950. [doi: 10.1055/s-2003-43479] [Medline: 14606018]
35. Dohi O, Yagi N, Onozawa Y, Kimura-Tsuchiya R, Majima A, Kitaichi T, et al. Linked color imaging improves endoscopic diagnosis of active Helicobacter pylori infection. Endosc Int Open 2016 Jul;4(7):E800-E805 [FREE Full text] [doi: 10.1055/s-0042-109049] [Medline: 27556101]
36. Takeda T, Asaoka D, Nojiri S, Nishiyama M, Ikeda A, Yatagai N, et al. Linked color imaging and the Kyoto classification of gastritis: evaluation of visibility and inter-rater reliability. Digestion 2019 Jul 12:1-10. [doi: 10.1159/000501534] [Medline: 31302654]
37. Yagi K, Saka A, Nozawa Y, Nakamura A. Prediction of Helicobacter pylori status by conventional endoscopy, narrow-band imaging magnifying endoscopy in stomach after endoscopic resection of gastric cancer. Helicobacter 2014 Apr;19(2):111-115. [doi: 10.1111/hel.12104] [Medline: 24372729]
38. Kanzaki H, Uedo N, Ishihara R, Nagai K, Matsui F, Ohta T, et al. Comprehensive investigation of areae gastricae pattern in gastric corpus using magnifying narrow band imaging endoscopy in patients with chronic atrophic fundic gastritis. Helicobacter 2012 Jun;17(3):224-231 [FREE Full text] [doi: 10.1111/j.1523-5378.2012.00938.x] [Medline: 22515361]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
BLI: blue-laser imaging
CNN: convolutional neural network
DOR: diagnostic odds ratio
FN: false negative
FP: false positive
IEE: image-enhanced endoscopy
LCI: linked color imaging
NBI: narrow-band imaging
NLR: negative likelihood ratio
NPV: negative predictive value
PLR: positive likelihood ratio
PRISMA-DTA: Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies
PROSPERO: International Prospective Register of Systematic Reviews
PPV: positive predicted value
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies–2
SROC: summary receiver operating characteristic
SVM: support vector machine
TN: true negative
TP: true positive
WLI: white-light imaging

Edited by G Eysenbach; submitted 30.06.20; peer-reviewed by J Chung, J Frausto-Solis; comments to author 31.07.20; revised version received 02.08.20; accepted 03.08.20; published 16.09.20.

Please cite as:
Bang CS, Lee JJ, Baik GH
Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy
J Med Internet Res 2020;22(9):e21983
URL: http://www.jmir.org/2020/9/e21983/
doi:10.2196/21983
PMID:32936088

Original Paper

Marrying Medical Domain Knowledge With Deep Learning on Electronic Health Records: A Deep Visual Analytics Approach

Rui Li[1*], BS; Changchang Yin[1*], MS; Samuel Yang[1,2], MD; Buyue Qian[3], PhD; Ping Zhang[1], PhD

[1]The Ohio State University, Columbus, OH, United States

[2]Nationwide Children's Hospital, Columbus, OH, United States

[3]Xi'an Jiaotong University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Ping Zhang, PhD
The Ohio State University
Lincoln Tower 310A
1800 Cannon Drive
Columbus, OH, 43210
United States
Phone: 1 614 293 9286
Email: zhang.10631@osu.edu

Abstract

Background: Deep learning models have attracted significant interest from health care researchers during the last few decades. There have been many studies that apply deep learning to medical applications and achieve promising results. However, there are three limitations to the existing models: (1) most clinicians are unable to interpret the results from the existing models, (2) existing models cannot incorporate complicated medical domain knowledge (eg, a disease causes another disease), and (3) most existing models lack visual exploration and interaction. Both the electronic health record (EHR) data set and the deep model results are complex and abstract, which impedes clinicians from exploring and communicating with the model directly.

Objective: The objective of this study is to develop an interpretable and accurate risk prediction model as well as an interactive clinical prediction system to support EHR data exploration, knowledge graph demonstration, and model interpretation.

Methods: A domain-knowledge–guided recurrent neural network (DG-RNN) model is proposed to predict clinical risks. The model takes medical event sequences as input and incorporates medical domain knowledge by attending to a subgraph of the whole medical knowledge graph. A global pooling operation and a fully connected layer are used to output the clinical outcomes. The middle results and the parameters of the fully connected layer are helpful in identifying which medical events cause clinical risks. DG-Viz is also designed to support EHR data exploration, knowledge graph demonstration, and model interpretation.

Results: We conducted both risk prediction experiments and a case study on a real-world data set. A total of 554 patients with heart failure and 1662 control patients without heart failure were selected from the data set. The experimental results show that the proposed DG-RNN outperforms the state-of-the-art approaches by approximately 1.5%. The case study demonstrates how our medical physician collaborator can effectively explore the data and interpret the prediction results using DG-Viz.

Conclusions: In this study, we present DG-Viz, an interactive clinical prediction system, which brings together the power of deep learning (ie, a DG-RNN–based model) and visual analytics to predict clinical risks and visually interpret the EHR prediction results. Experimental results and a case study on heart failure risk prediction tasks demonstrate the effectiveness and usefulness of the DG-Viz system. This study will pave the way for interactive, interpretable, and accurate clinical risk predictions.

(J Med Internet Res 2020;22(9):e20645) doi:10.2196/20645

KEYWORDS

electronic health records; interpretable deep learning; knowledge graph; visual analytics

Introduction

Clinical risk prediction is an important task in electronic health record (EHR) analysis aiming to predict the current and future states of patients based on their historical diagnosis codes, laboratory results, clinical notes, and other medical events. Recurrent neural networks (RNNs), as a successful extension of standard feed-forward networks, have recently been shown to leverage the superior computational power of neural networks and gain good performance in clinical tasks, such as diagnostic code prediction [1-8], disease progression modeling [9], patient subtyping [10], clinical relation identification [11], and imputation of missing values [12,13]. To pursue better performance, some approaches [2,3] attempt to integrate medical domain knowledge (eg, the hierarchical structure of International Classification of Diseases, Ninth Revision [ICD-9] codes) to learn better medical code representations and achieve much better diagnosis prediction accuracies. The resulting improvement substantially benefits clinical care applications such as clinical decision support systems [14].

Despite the superior performance from RNNs, optimizing, interpreting, and applying such models in clinical practice remain to be challenges to domain experts [1,4]. First, when deploying these accurate yet complicated models in clinical practice, there is an increasing trend for the domain experts to focus more on trust and interpretability issues. For example, in the context of a heart failure prediction task, which factors do models consider more important in determining a high prediction risk? Second, it is also quite challenging for presenting the results and interacting with the models. For example, medical experts may be eager to know what will happen to the prediction results if we add a new drug on a specific date. However, it is difficult to ask doctors to interact with a complex model without any interface design. Thus, it is worthwhile to develop a robust, interpretable, and interactive system to address the above limitations.

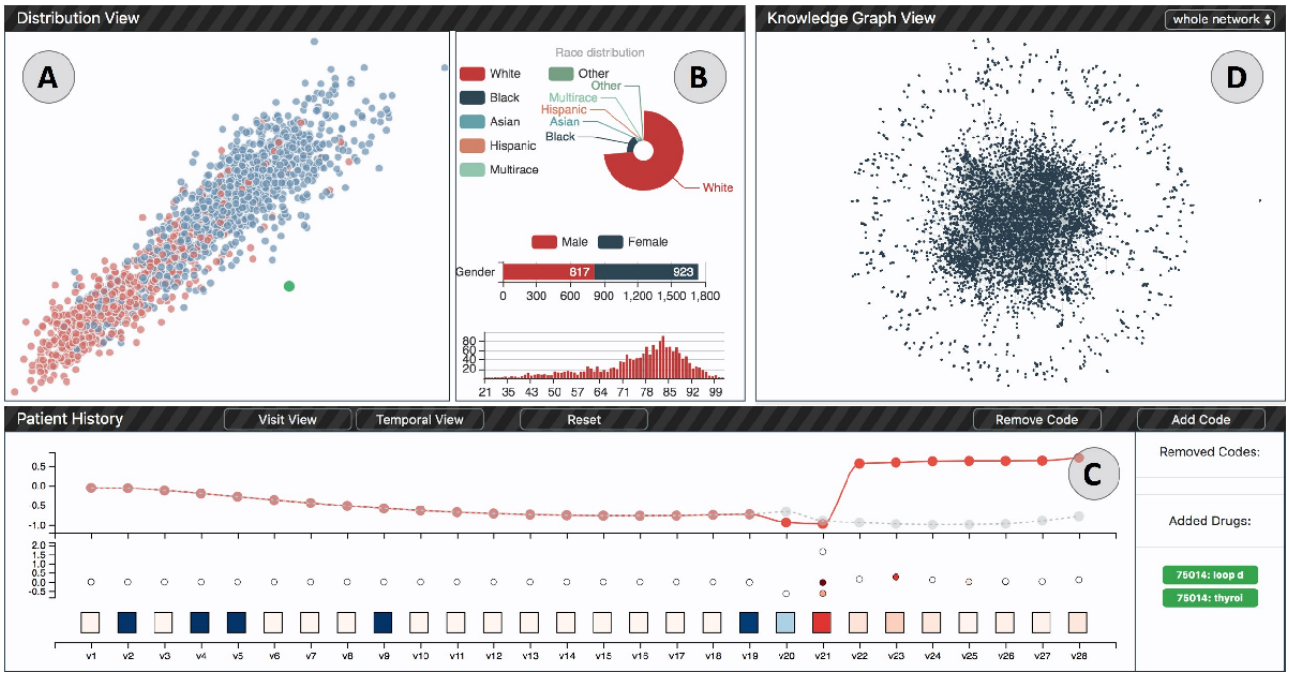
Recently, there has been an increasing interest in applying visual analytic techniques to interpret the RNN model for EHR prediction tasks. For example, RetainVis [4] improved the reverse time attention model (RETAIN) [1] with additional features (eg, temporal information) and visualized the contribution of both visit-level and code-level using multiple visualization views. Similarly, CarePre [15] is designed to interpret the prediction results in the context of a group of similar patients based on the RETAIN model. In this study, we present DG-Viz, which brings together the power of deep learning (ie, an interpretable RNN model) and visual analytics to predict clinical risks and visually interpret the EHR prediction results. Specifically, we develop an interpretable RNN model, called domain-knowledge–guided recurrent neural network (DG-RNN), which incorporates medical knowledge from a public medical knowledge graph KnowLife [16] with a graph-based attention mechanism. Then, the output vectors are concatenated, and a global max-pooling layer is followed to generate a fixed-size vector. Next, a fully connected layer is used to generate clinical outcomes. Following this, based on the model and EHR data, we design and implement the DG-Viz system, as shown in Figure 1, which consists of a projection view, a patient history view, and a knowledge graph view to present an overview of EHR data; the prediction results of individual patients; and the knowledge graph contribution of our model. The patient history view also allows users to conduct *what-if* analysis to understand how a specific factor will cause the variance of the prediction result. We present the robustness and performance of our model by comparing our model with both traditional machine-learning methods and recent deep learning approaches for heart failure risk prediction tasks. Finally, we demonstrate the effectiveness of our system through a case study using a real-world data set with a medical expert. In summary, the main contributions of this study are as follows:

- We present the clinical risk prediction framework DG-RNN, which can incorporate medical domain knowledge with a graph-based attention mechanism.
- We introduce a global pooling operation to DG-RNN, which makes our prediction model interpretable. The model can output the medical events that cause the final clinical outcome.
- We designed and developed a visual analytics system, DG-Viz, which enables the exploration and interpretation of clinical risk prediction tasks by integrating our deep learning model with the design of visualizations and interactions.
- We validated the robustness and effectiveness of our system by conducting both quantitative experiments and a case study with medical experts. We summarized the insights from the feedback.

Note that DG-RNN was introduced in our previous conference paper [17]. The key differences between this paper and the prior conference paper are as follows:

- We discuss with clinicians and summarize four main themes of visual design requirements.
- We developed a new visual analytics system, DG-Viz, to display the DG-RNN prediction results and validated its effectiveness with a case study on a real-world data set.
- We provide two kinds of *what-if* operations to edit the input data (by removing medical codes and adding drugs) and compare the changes in predicted risks.

Figure 1. A screenshot of DG-Viz. (A) The patient distribution view shows an overview of all patients. (B) The demographic chart shows the demographics distribution of all patients. (C) The patient history view shows the contributions of all visits and medical codes of a single patient. The line chart presents the prediction results among time. (D) Knowledge graph view shows the whole network structure. v1-v28: different visits.



Methods

Analytical Tasks

To integrate the proposed interpretable model DG-RNN [17] with a visual analytics system, we conducted weekly meetings among all coauthors from this paper, who are experts in visualization, deep learning, and medical domains, to distill the requirements of the desired visual analytics system. The 2-month discussion elicited the following 4 main themes of the visual design requirements:

- R1: Provide an overview of all patients and their demographic information. It is a fundamental requirement for experts to provide an overview of the patients in the data set. In particular, they are interested in the following questions:
  - R1.1: What are the distributions of all patients? For example, can we find different subtypes of patients within the data set?
  - R1.2: What are the distributions of patients' demographic information? (eg, gender ratio and range of ages)
- R2: Present the medical history and prediction results of a single patient. This requirement enables users to explore a patient's history; the system should especially be able to do the following:
  - R2.1: Show all visits and medical codes for a single patient.
  - R2.2: Reveal the temporal time interval between different visits. The temporal interval information is important for experts to analyze patients' medical history.
  - R2.3: Visualize how the prediction results evolved with time. Users are curious about the prediction results up to a certain visit.
- R3: Enable the model interpretation. In addition to presenting the prediction results from the model, it is crucial to understand how the prediction results are made; to this end, we include the following goals:
  - R3.1: Demonstrate the contribution of patient visits and medical codes to the final prediction scores. Users should be able to identify the key factors affecting the prediction result.
  - R3.2: Reveal the contribution of the knowledge graph to the prediction results. In particular, users want to know what the whole knowledge graph looks like and how the contribution of a specific medical code is affected by its neighbors in the knowledge graph.
- R4: Provide the what-if analysis on the prediction model. Users are curious about how changes in medical codes will affect the outcome. In particular, the system should enable users to add or remove specific medical codes and observe how these updates will affect the final prediction results.

Deep Learning Model: DG-RNN

In this section, we provide a brief introduction on the basic ideas and important concepts of our proposed DG-RNN model. For details, please refer to the study by Yin et al [17].

Data Structure of EHRs

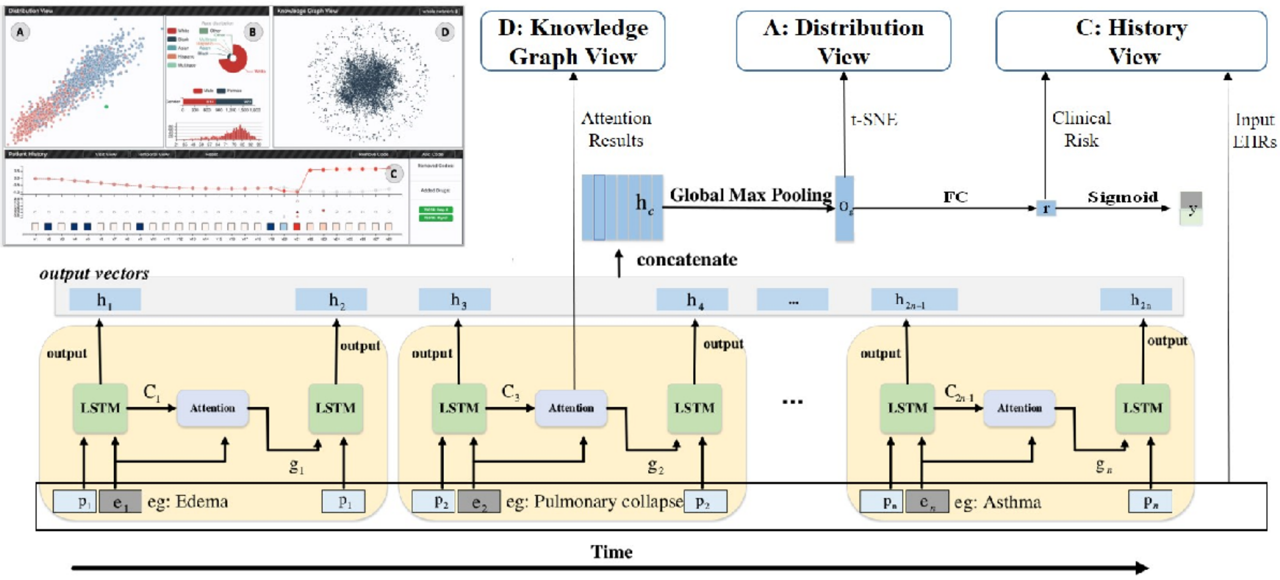
There is a sequence of visits in each patient's EHR history, where each visit consists of several medical codes. Following previous studies (such as the study by Zhu et al [18]), the medical events are ordered according to their time of occurrence. The codes in both the knowledge graph and EHR data are projected to the same embedding space. The EHR sequence of

the patient i is denoted as and represents the ground truth. After medical code embedding, the patient's medical events are represented as , where .

DG-RNN Model

As shown in Figure 2, DG-RNN takes both the medical events and the corresponding occurring time as inputs. For example, at the t^{th} DG-RNN step, the event embedding vector e_t and its time encoding vector p_t are inputs to the long short-term memory network (LSTM) [19], which produces a hidden state C_{2t-1} and an output vector h_{2t-1}. Then, the subgraph adjacent to the t^{th} event and C_{2t-1} is sent to the graph attention module, which computes the attention result g_t, which is sent to the LSTM again and another output vector h_{2t} is generated. Note that the unit of our model generates 2 output vectors for 1 input event, which can help to compute the contribution rates of the initial medical event and the potential information from the medical knowledge graph. Next, we concatenate all the output vectors and leverage a global pooling layer to generate a fixed-size vector o_g. Finally, a fully connected layer is adopted to predict the clinical risk. The model is trained by minimizing the cross-entropy loss between the ground truth and the predicted risk y_i for each patient i as follows:

Figure 2. Framework of domain-knowledge–guided recurrent neural network (DG-RNN), which takes the medical event embeddings and the corresponding time encoding vectors as inputs. For each event input, DG-RNN generates two output vectors. After all the input codes input to DG-RNN, we concatenate the output vectors and leverage a global max pooling and a fully connected layer (FC) to predict the clinical risk. We adopt t-distributed stochastic neighbor embedding (t-SNE) to map the global pooling layer's output vectors to a 2D space (the Distribution View A is DG-Viz), where the distance between patient represents their similarity. The attention results are displayed in the knowledge graph view D to show the knowledge graph's contribution in DG-RNN. The input medical codes and the output clinical risks are displayed in the History View C in DG-Viz, which shows the patient's risk changing trend. LSTM: long short-term memory; FC: fully connected layers; t-SNE: t-distributed stochastic neighbor embedding.



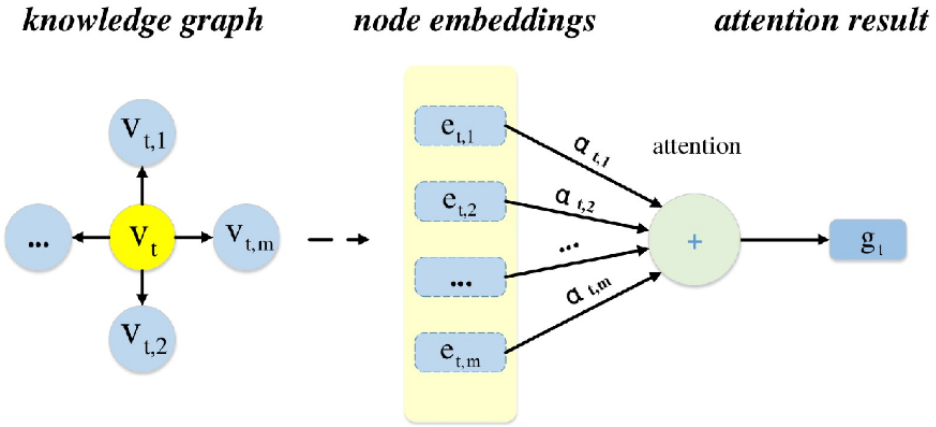
Knowledge Graph Attention Mechanism

To incorporate the medical domain knowledge, we propose a dynamical graph attention mechanism.

The relations (eg, *causes* and *is-caused-by*) and entities (eg, *diseases*) of the knowledge graph are projected into a *d-dimension* space. Given the t^{th} input event v_t as head entity, which has many relation edges in the knowledge graph, denoted as , the proposed attention mechanism is able to automatically attend to useful related tail entities in the knowledge graph. Formally, it takes the hidden state C_{2t-1} of the LSTM and the related relations R_t as inputs and then calculates the attention weights as follows:

where are the relation and tail entity embeddings, and are learnable parameters. Following the study by Zhou et al [20], our attention mechanism takes the related head node v_t, relation edge $r_{t,m}$, and tail node $e_{t,m}$ into account. Given the attention weights, we leverage soft attention to generate the attention result vector g_t, as shown in Figure 3. Following this, g_t is input to the LSTM, as shown in Figure 2.

Figure 3. Attention mechanism. In the knowledge graph, the yellow node means the current input medical event and the other nodes are its adjacent nodes. Our attention mechanism takes as inputs the embeddings of the adjacent nodes and generates the graph attention vector.



Global Max Pooling Operation

RNN-based models are sometimes inefficient because of their long-term dependency. It is possible for RNN models to forget the earlier data if the input sequences are too long. Therefore, we propose to concatenate the output vectors of the RNN and introduce a global max-pooling operation to DG-RNN, which shortens the distance between the earlier input's medical events and the final output risks. To the best of our knowledge, this is the first time that a max-pooling operation is leveraged in RNN-based models. As shown in Figure 2, the LSTM output vectors are concatenated, followed by a global pooling operation. The output o_g is sent to a fully connected layer to predict the clinical risk for the patient i, which is defined as

where  and  are the learnable parameters, z_i and y_i denote the clinical risk score and probability, respectively. The global pooling operation is helpful in calculating the contribution rates of various medical events to the final output clinical risks. Note that the details of how to compute medical events' contributions can be found in our conference paper [17].

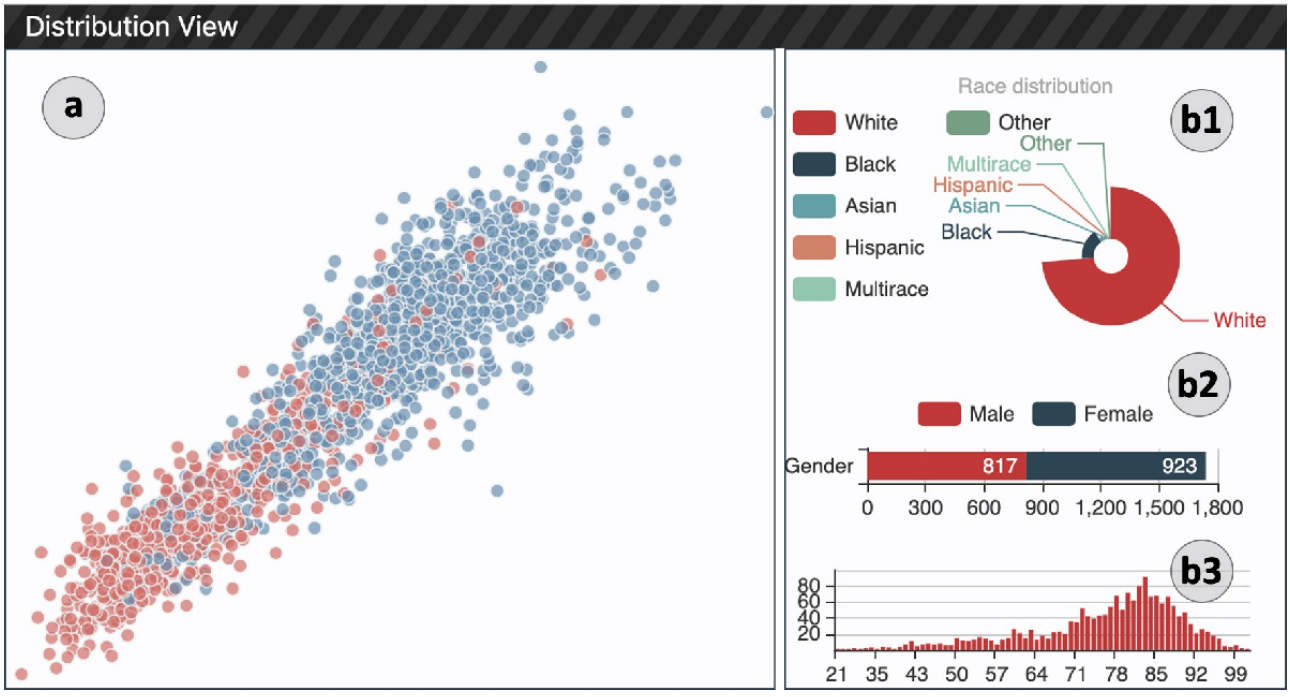
Visual Analytics System: DG-Viz

In this section, we explain the visual interface and the design rationale of the 3 components of DG-Viz.

Distribution View

The distribution view (Figure 4) provides an overview of the entire data set, including the overall distribution of patients (R1.1) and their demographic information (R1.2). It contains 2 components: (1) a projection chart showing the distribution of patients and (2) the right-side panel with demographic information.

Figure 4. Distribution view: (a) the projection scatter plot of all patients in the test data set, (b1) the race distribution chart, (b2) the gender distribution chart, and (b3) the age distribution histogram.



Projection View

The objective of the projection view is to position the patients in a two-dimensional (2D) space, and their relative similarities are reflected through their distance to help users discover clusters. For this purpose, we created a vectorized representation to encode the medical history information of each patient. In particular, for a given patient p, we use DG-RNN to predict the risk of their heart failure. The global max-pooling layer output vector, that is o_g in equation (4), represents the patient's features. Then, we adopt the t-distributed stochastic neighbor embedding (t-SNE) algorithm [21] to project all patients into a 2D space. As a result, the patients are positioned in such a way that similar patients are placed nearby, whereas dissimilar patients are placed far away. To differentiate patients' diagnosis results, we show patients with positive and negative heart failure outcomes in different colors (red: positive, blue: negative). The selected patient of interest is highlighted in green. Section (a) in Figure 4 shows that all patients were divided into 2 groups. A zooming interaction is also provided to explore and select target patients who are placed together.

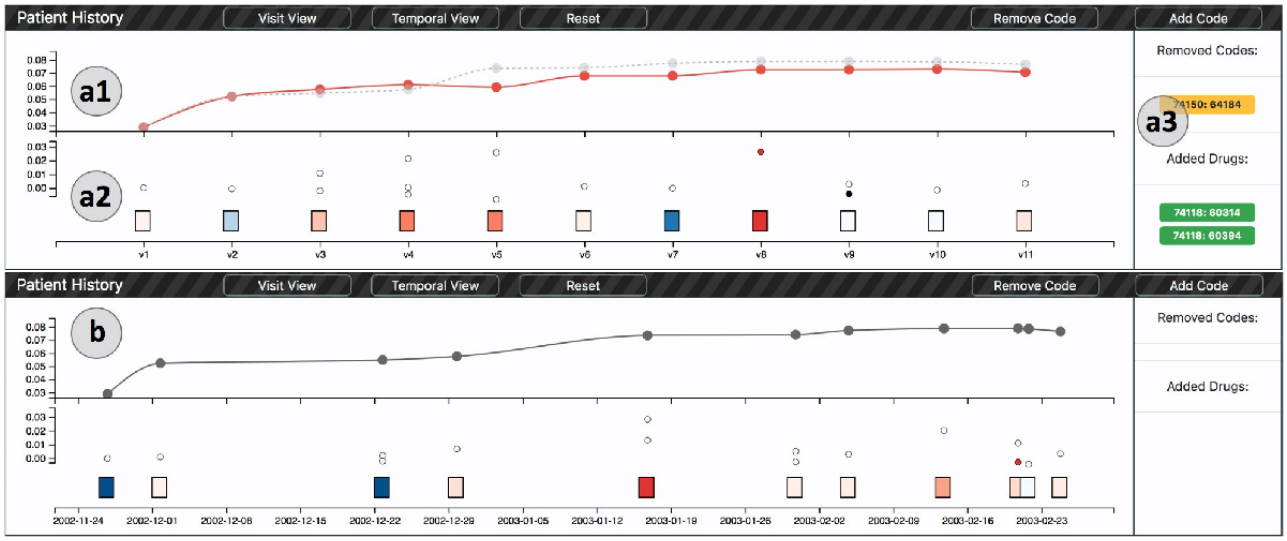
Demographic Panel

The demographic panel, section (b) in Figure 4 shows 3 different charts, which visualize the distributions of patients' race, gender, and age. The race distribution is presented with a Nightingale Rose Chart, where the area of each slice represents the number of patients belonging to the corresponding race. For example, in section (b1) in Figure 4, we can observe that most of the patients in the data set are White. The stacked horizontal bar chart presents the gender ratio of all patients. Users can also see the distributions of age in the bottom age histogram.

Patient History and Prediction View

After selecting the patient of interest in the distribution view, users can further investigate the patient's history information (R2.1, R2.2), see the prediction results (R2.3), and understand how the prediction results change by updating the input data (R3.1, R4). In the patient history view, there are 2 charts vertically shown from top to bottom, as shown in section (a) in Figure 5. The chart at the bottom (section a2 in Figure 5) is used to present the visit and the medical codes, and the top chart presents the prediction results (a1 in Figure 5).

Figure 5. Patient history view. Top: the visit view that arranges all visit records with the same distance. (a1): the prediction results involved with time, (a2): the visits and medical codes of the patient, (a3): added or removed medical codes. Bottom: (b) the temporal view that arranges all visit records based on their time intervals.



Visits and Medical Codes View

We sort the time stamps of all visit records. Next, we visualize these records using the rectangular boxes and arrange them from left to right in a chronological order, as shown in section (a1) in Figure 5. Just as in previous studies on visualizing the sequence models [4,22], the color of the visit box (from blue to white to red) represents the corresponding contribution risk (from negative to 0 to positive).

To provide an overview of all visit records while preventing the clutter visual layout, we position the visit box in a uniform manner, that is, the distances between all visit boxes are the same. However, the temporal interval information serves as an important indicator in clinical analysis. We also provide a temporal view (section b in Figure 5) with different placement of visit boxes. In the temporal view, users can observe the overall distribution of the visit time and zoom in the x-axis to explore the overlapping visits. For example, in section (b) in Figure 5, v9 and v10 are two of the closest visits.

To identify the key medical codes that contribute to the prediction results (R3.1), we allow users to compare the importance of medical codes from 2 contributions: (1) the total contribution of the medical code and (2) the contribution caused by the neighboring codes from the knowledge graph. We introduce a bi-encoding (position-color) method to encode these 2 contributions. First, the horizontal positions of these codes are aligned with their corresponding visit, whereas their vertical positions represent their total contribution risks. Users are able to scale the y-axis to observe the codes that appear together owing to a similar value. In terms of the knowledge graph contribution, we map the weight of the knowledge graph (from positive to 0 to negative) to a diverging color map (from red to white to blue). This design enables users to easily identify the key code with the highest contribution to the results of the

prediction, and the code that is impacted by the knowledge graph the most as well. For example, in section (a2) in Figure 5, we can observe that the medical code appearing in v8 presents the highest contribution risk. The color of this code also indicates that it is strongly influenced by the knowledge graph.

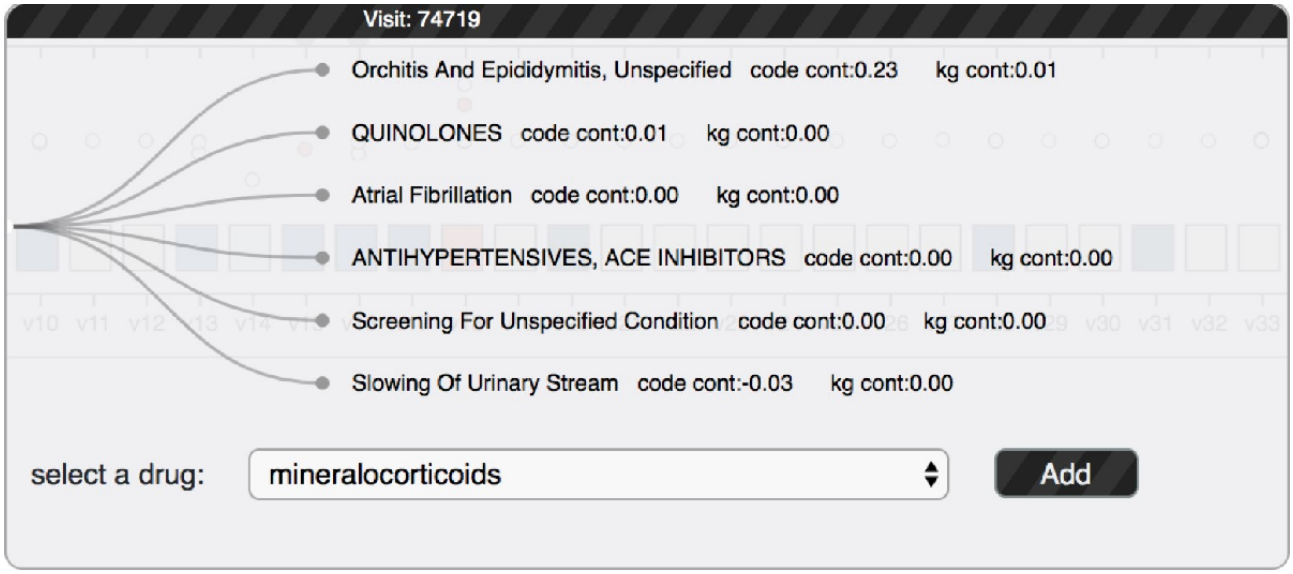
Prediction Results View

We show the prediction results involved with the time using a line chart (section a1 in Figure 5), which is an intuitive and straightforward approach to visualize time-series data [23,24]. In this chart, the horizontal axis is used to represent the visit time, and each node in the line chart is synchronized with the corresponding visit records. The vertical axis indicates the prediction score obtained up to a certain visit. For example, the predicted score at v3 is computed from the model with the input visits of v1, v2, and v3.

What-If Analysis View

We also provide a set of interactions to allow the users to conduct a what-if analysis. We provide 2 ways to edit the input data: removing medical codes (R3.1) or adding specific drugs. As shown in Figure 6, users can select multiple target medical codes from a pop-up panel by clicking the code circle in the code chart. Once the removing button is clicked, the line chart will show an extra red line to indicate the updated prediction results. The original prediction results will be visualized using a gray dashed line, which enables users to observe their difference effectively. The visit and code views will also be updated accordingly.

Figure 6. Code edit panel: users can see all the medical codes within a specific visit and add drug to this visit. kg: knowledge graph; cont: contribution.



To provide a what-if analysis by adding specific drugs, we identified 9 drugs for heart failure treatment through the literature [16]. These drugs have been discovered for more than 400,000 times in our data set. Users can select the visit box corresponding to the time they want to add the test drugs. As shown in section (a1) in Figure 5, users can choose the drug and obtain updated prediction results by clicking the add code button. The added drugs will be shown on the right side of the code view (section a3 in Figure 5).

Patients with high heart failure risks usually take more drugs than healthy patients. It is easy for DG-RNN to learn incorrect knowledge that drugs may cause higher risks. Thus, we resampled the data set when training the proposed model. First, we built a new data set by removing all the drugs and trained a logistic regression (LR) model to predict heart failure risks. Given the predicted risks without drugs, in each batch data, we selected equal numbers of case patients (who take drugs at least once) and control patients (who never take drugs) with similar heart failure probability. Finally, the DG-RNN was trained with the resampled batch data.

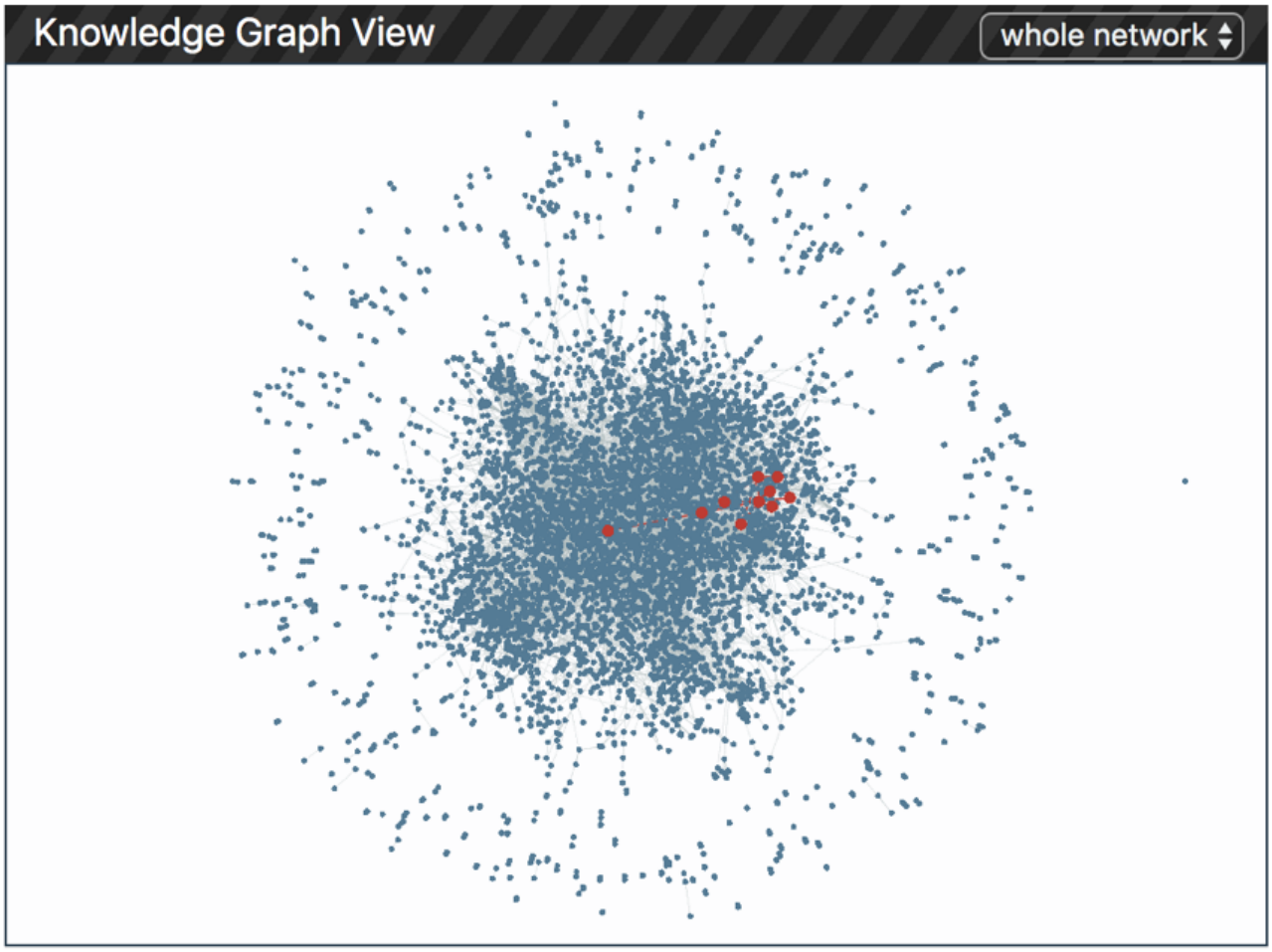
Knowledge Graph View

The knowledge graph view aims to reveal the whole structure of the knowledge graph used in the model and highlight the subgraph activated by a particular visit or medical code in the prediction. It also allows users to identify how the contribution of a specific medical code is affected by its neighbors in the knowledge graph (R3.2). It contains two subviews: (1) an overview of the whole knowledge graph network structure and (2) a local code network showing the local relationships between medical codes. Users can easily switch between them by clicking on the toggle on the top.

Whole Network

To visualize the whole knowledge graph structure, we use all the disease and drug entities and their relations to construct a network that includes 7273 nodes and 20,491 edges. We present this network using a force-directed graph, as shown in Figure 7. In this graph, each node is presented with a navy blue color, and edges are presented using gray lines. Clicking on a specific visit box or medical code point in patient history view will highlight the corresponding nodes and their connected neighbors in the whole network. For example, in Figure 7, we can observe that the medical codes used in the prediction are located in the center area of the whole network.

Figure 7. The whole knowledge graph in Knowledge Graph View.

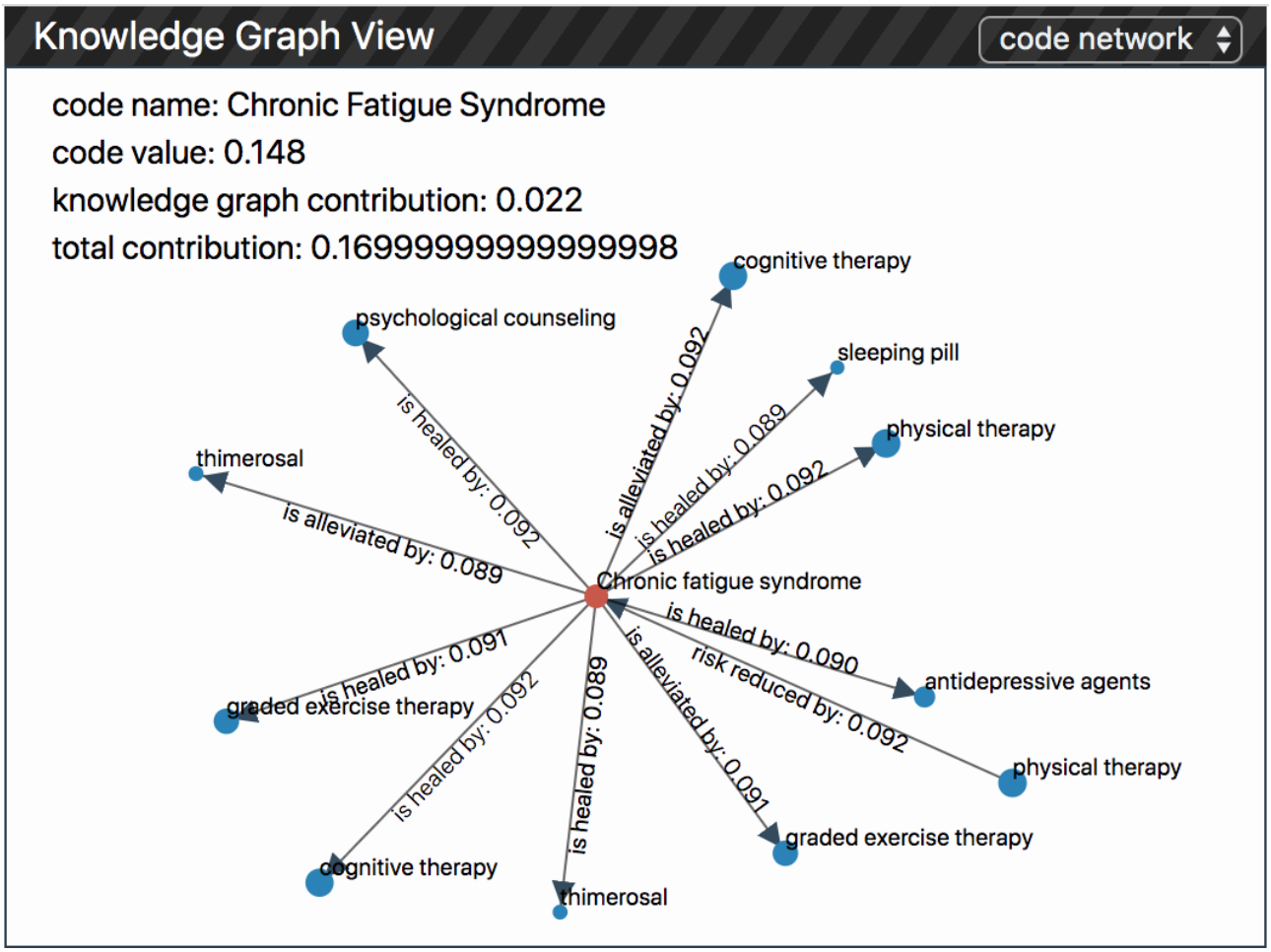


Code Network

To reveal the relationship between the selected medical codes and their neighbors in the knowledge graph, we place these codes in a force-directed graph. The red node in the center denotes the target node (ie, the selected medical node in the patient history view), and the blue dots around it represent the neighboring nodes in the knowledge graph. We encode the contribution of these neighbors using size, and a large dot represents an important node that contributes to the target node. The edges in the graph represent the relationship between the target nodes' neighbors. For example, in Figure 8, we can find that the contribution of *chronic fatigue syndrome* is affected by or affects 11 other nodes in the knowledge graph. Among these neighbors, cognitive therapy is the most important node.

In Multimedia Appendix 1, we provide a demo video of DG-Viz. It can also be found at YouTube [25].

Figure 8. The local network of a specific medical code and its neighbors in Knowledge Graph View.



Results

This section reports the results from 3 forms of evaluation: (1) quantitative experiment on heart failure risk prediction tasks to compare our model with the state-of-the-art models, (2) a case study with a medical physician, and (3) the feedback from the physician.

Data Sets

We conducted heart failure prediction experiments on a real-world longitudinal EHR database, which includes 218,680 patients for over 4 years. Patients with a diagnosis of heart failure were selected as case patients. For each case, we selected 3 control patients with the same *age* and *sex*. Each case patient's heart failure confirmation date is set as their operation criterion date. The control patients' criterion dates are the same as that of their corresponding case patients. Finally, we trace back from the operation criterion date and hold off the EHRs in a prediction window. Six different hold-off windows (ie, 7, 14, 30, 60, 90, and 120 days) were used in our experiments. The medical codes appearing less than 10 times were removed. Table 1 lists the statistics of the selected data sets.

Table 1. Statistics of data sets.

Characteristics	EHR[a]-120	EHR-90	EHR-60	EHR-30	EHR-14	EHR-7
Number of case patients	442	462	494	517	536	554
Number of control patients	1326	1386	1482	1551	1608	1662
Number of events in the data set	134,666	140,984	152,389	160,584	169,636	176,460
Number of unique events	967	974	978	983	989	995
Average of EHRs' length	76.17	76.29	77.11	77.65	79.12	79.62
Average number of events per visit	2.17	2.36	2.29	2.41	2.35	2.39

[a]EHR: electronic health record.

In addition to the initial EHR data, DG-RNN also takes medical knowledge graphs as inputs. A publicly available knowledge graph KnowLife [16] is leveraged in our experiments. KnowLife has millions of entities (eg, diseases and medications) and dozens of relations (eg, *causes* and *is-healed-by*). We initialize

the entities' and relations' embeddings with TransE [26] and fine-tune the embeddings when training the model.

Baselines

To validate the performance of the proposed DG-RNN, we compare DG-RNN with the following baselines, including 3 traditional machine-learning methods (ie, random forest [RF], LR, and support vector machine [SVM]) and 5 deep learning methods (ie, gated recurrent unit [GRU]) [27], LSTM [19], RETAIN [1], graph-based attention model (GRAM) [2], and knowledge-based attention mode (KAME) [3]). Moreover, we implement three versions of DG-RNN to validate the effectiveness of the knowledge graph attention module and the global pooling operation. DG-RNN is the main version of our model. DG-RNN-nk does not use the medical domain knowledge by removing the graph attention module. DG-RNN-np predicts the heart failure risk based on the last hidden state of the LSTM, without the global pooling operation.

Implementation Details

The traditional methods and deep learning models are implemented with scikit-learn and PyTorch 0.4.1, respectively. We adopted a grid search to find the best parameter for traditional methods. For a fair comparison between DG-RNN and knowledge-incorporated baselines (ie, GRAM and KAME), KnowLife [16] is used as the domain knowledge for both GRAM and KAME. Note that all the medical codes in EHRs and KnowLife are represented as ICD-9 codes, and all the models only accept the structured ICD-9 codes as inputs. When training deep learning–based models, we used the Adam optimizer with a mini-batch of 64 patients and trained using 1 graphics processing unit (TITAN XP GPU) for 50 epochs, with a learning rate of 0.0001. The outputs of DG-RNN include the risk probabilities and events' contribution risks. Patients' risk probabilities are used to train DG-RNN, whereas the events' contribution risks are only visualized in our DG-Viz system. Further implementation details can be found in our conference paper [17] and on github [28,29].

Results of Risk Prediction

The experimental results in Tables 2, 3, and 4 show that the proposed model outperforms the baselines, which demonstrates the effectiveness of DG-RNN. To better measure the difference in performance between the proposed DG-RNN and the baselines, following the study by Tang et al [30], we performed statistical testing and calculated the *P* value of area under a receiver operating characteristic (AUROC) score between the proposed DG-RNN and various baseline models using statistical *t* testing. For all the baselines, the *P* value results are very small ($P<.001$), which demonstrates that the risk prediction performance difference between DG-RNN and baselines is significant.

The performance of deep learning methods is much better than that of the 3 traditional machine-learning methods. The possible reason may be that deep learning approaches take the embedding of medical codes as inputs, which can capture the medical codes' clinical meaning, whereas the traditional approaches use high-dimensional one-hot representations, which have a semantic gap. Moreover, RNN-based methods are better for modeling patients' health status and consider the order of EHR sequences (temporal information). Among the 5 deep learning baselines, with the help of the attention mechanism, RETAIN performs better than GRU and LSTM. Considering the medical knowledge graph, KAME and GRAM outperform RETAIN, which demonstrates that medical domain knowledge does help to improve the performance in clinical applications.

Among the proposed model's 3 versions, our main version DG-RNN achieves the best performance. After removing the medical knowledge graph, there is about 2% AUROC decline for the version DG-RNN-nk, which demonstrates that medical domain knowledge from KnowLife is very helpful. Without the global pooling layer, DG-RNN-np also achieves worse performance than DG-RNN by 2%, which demonstrates the effectiveness of the introduced global pooling operation. The global pooling operation can shorten the distance between early occurring medical events and the final outputs, which makes the training process more efficient.

Table 2. Area under a receiver operating characteristic of the heart failure prediction task.

Model	EHR[a]-120	EHR-90	EHR-60	EHR-30	EHR-14	EHR-7
LR[b]	0.6883	0.6956	0.6932	0.7139	0.7347	0.7386
RF[c]	0.6726	0.6913	0.6965	0.7212	0.7217	0.7336
SVM[d]	0.6173	0.6339	0.6213	0.6258	0.6323	0.6372
GRU[e]	0.6504	0.6670	0.6939	0.7178	0.7438	0.7638
LSTM[f]	0.6628	0.6792	0.6982	0.7282	0.7459	0.7631
RETAIN[g]	0.6962	0.7115	0.7318	0.7437	0.7561	0.7683
GRAM[h]	0.7081	0.7292	0.7378	0.7525	0.7648	0.7656
KAME[i]	0.7168	0.7319	0.7392	0.7573	0.7662	0.7717
DG-RNN[j]-nk	0.7158	0.7310	0.7368	0.7486	0.7583	0.7663
DG-RNN-np	0.6995	0.7075	0.7182	0.7425	0.7596	0.7723
DG-RNN	0.7288	0.7437	0.7510	0.7663	0.7789	0.7863

[a]EHR: electronic health record.
[b]LR: logistic regression.
[c]RF: random forest.
[d]SVM: support vector machine.
[e]GRU: gated recurrent unit.
[f]LSTM: long short-term memory.
[g]RETAIN: reverse time attention model.
[h]GRAM: graph-based attention model.
[i]KAME: knowledge-based attention model.
[j]DG-RNN: domain-knowledge–guided recurrent neural network.

Table 3. Sensitivity of the heart failure prediction task.

Model	EHR[a]-120	EHR-90	EHR-60	EHR-30	EHR-14	EHR-7
LR[b]	0.6262	0.6441	0.6452	0.6512	0.6522	0.6684
RF[c]	0.6235	0.6456	0.6549	0.6612	0.6636	0.6723
SVM[d]	0.5689	0.5835	0.5732	0.5769	0.5822	0.5862
GRU[e]	0.6120	0.6227	0.6348	0.6524	0.6837	0.7001
LSTM[f]	0.6322	0.6407	0.6564	0.6869	0.6874	0.7006
RETAIN[g]	0.6556	0.6612	0.6719	0.6916	0.6938	0.7018
GRAM[h]	0.6614	0.6627	0.6718	0.6914	0.7030	0.7046
KAME[i]	0.6645	0.6714	0.6759	0.6828	0.6991	0.7036
DG-RNN[j]-nk	0.6634	0.6712	0.6790	0.6817	0.6926	0.7132
DG-RNN-np	0.6513	0.6569	0.6727	0.6801	0.6997	0.7101
DG-RNN	0.6754	0.6816	0.6856	0.7012	0.7145	0.7206

[a]EHR: electronic health record.
[b]LR: logistic regression.
[c]RF: random forest.
[d]SVM: support vector machine.
[e]GRU: gated recurrent unit.
[f]LSTM: long short-term memory.
[g]RETAIN: reverse time attention model.
[h]GRAM: graph-based attention model.
[i]KAME: knowledge-based attention model.
[j]DG-RNN: domain-knowledge–guided recurrent neural network.

Table 4. Specificity of the heart failure prediction task.

Model	EHR[a]-120	EHR-90	EHR-60	EHR-30	EHR-14	EHR-7
LR[b]	0.6402	0.6437	0.6429	0.6528	0.6727	0.6887
RF[c]	0.6301	0.6414	0.6484	0.6674	0.6720	0.6802
SVM[d]	0.5897	0.5904	0.5948	0.6041	0.6062	0.6079
GRU[e]	0.6231	0.6458	0.6510	0.6718	0.6947	0.7020
LSTM[f]	0.6106	0.6252	0.6293	0.6427	0.6563	0.6595
RETAIN[g]	0.6602	0.6619	0.6755	0.7016	0.7041	0.7165
GRAM[h]	0.6673	0.6835	0.6901	0.7014	0.7108	0.7114
KAME[i]	0.6720	0.6806	0.6842	0.6951	0.7119	0.7131
DG-RNN[j]-nk	0.6773	0.6819	0.6893	0.6924	0.7158	0.7190
DG-RNN-np	0.6707	0.6769	0.6791	0.7037	0.7078	0.7166
DG-RNN	0.6862	0.6976	0.7022	0.7128	0.7254	0.7273

[a]EHR: electronic health record.

[b]LR: logistic regression.

[c]RF: random forest.

[d]SVM: support vector machine.

[e]GRU: gated recurrent unit.

[f]LSTM: long short-term memory.

[g]RETAIN: reverse time attention model.

[h]GRAM: graph-based attention model.

[i]KAME: knowledge-based attention model.

[j]DG-RNN: domain-knowledge–guided recurrent neural network.

Case Study

To illustrate how a physician can explore the EHR data and interpret the prediction results, we provide a case study. In particular, we worked with the same medical expert in the design state of DG-Viz using the same EHR data set mentioned previously. We first introduced the functions and interaction methods of the DG-Viz system to the doctor. After becoming familiar with DG-Viz, the doctor was asked to perform a set of tasks, such as observing the patient overview, interpreting the prediction results, and testing their hypotheses. The doctor was free to ask any questions about the system during the study.

Figure 4 shows an overview of 1740 patients in our data set. As the expert pointed out, the initial overview of the patient cohort showed that patients clustered in the bottom left quadrant had a positive heart failure risk score ranging between 1 and 2. The center of the graph comprised patients with heart failure risk between 0 and a negative one. In the upper right quadrant, heart failure risk scores varied between 0 and −3. Heart failure risks were similar for 2 patients in close proximity to the distribution view. He mentioned that the overview would be particularly useful for the physician, at a quick glance, to see which patients are at a higher risk for a given disease.

Next, the expert was interested in identifying the medical codes that correlated with heart failure. To do this, he selected multiple patients with heart failure for further inspection. According to the visualization results, he mentioned that atrial fibrillation and cardiac dysrhythmia are often shown to contribute to the risk of heart failure. Arrhythmias are common and have a known association with heart failure, either as a cause or as a sequela, and increased his confidence in the heart failure prediction. Less frequently, shortness of breath, edema, cardiomegaly, and aortic valve disorders were shown to greatly contribute to the risk of heart failure. Shortness of breath and edema are common symptoms affecting patients with heart failure. Cardiomegaly is a finding either on physical examination or on diagnostic testing that is associated with heart failure. Aortic valve disorders, which include aortic regurgitation and aortic stenosis, are one of many causes of heart failure. All these nodes were consistent with the current medical understanding of heart failure.

The doctor was also interested in checking whether the prediction results of the system meet their expectations. For 1 patient with a heart failure risk of 3, the physician added an angiotensin receptor antagonist, a medication typically prescribed in heart failure, but only a slight decrease in heart failure risk was observed. However, adding additional antihypertensive medications (calcium channel blockers) lowered the risk of heart failure by a greater amount. This may indicate that the model agrees with the known causation between hypertension and heart failure. Not all patients showed this behavior, possibly indicating that their medical history did not include hypertension. For some patients, adding a loop diuretic increased the risk of heart failure. Loop diuretics are often

prescribed as symptomatic treatment for heart failure but are not known to decrease mortality or prevent the onset of disease. The need for a loop diuretic prescription in the absence of a heart failure diagnosis may indicate the early stages of the disease and is a good alert for clinicians.

Finally, the expert was asked to check the local knowledge graph structure and verify the correctness of the prediction results. He mentioned that a frequent prediction result node with high-risk contribution and high knowledge graph contributions was atrial fibrillation. The neighbor nodes displayed in the knowledge graph view included definitions of atrial fibrillation (*cardiac dysrhythmia*), as well as potential treatment options (*pacerone* and *cardiac ablation*). They were all related to heart failure. However, in another case of edema, he found that related diseases shown in the knowledge graph, including ulcerative colitis, cerebral abscess, and reparative closure, did not have a strong relationship with heart failure.

Overall, the doctor believed that DG-Viz is a great tool and an interesting way of *proving* to the physician that the predicted risk is valid. In particular, the interface provides the ability to explore the consequences of prescribing medication. The knowledge graph allows the physician to see contributing diagnoses to the overall risk. However, he also felt that the current interface "while intuitive, does contain a large amount of medical information and requires a substantial explanation of each item before the information can be synthesized." The color and position used to encode the contribution of visit and code might actually increase the cognitive load of the doctor.

Discussion

Principal Results

In this study, we present DG-Viz, an interactive clinical prediction system, which brings together the power of deep learning (ie, a DG-RNN–based model) and visual analytics to predict clinical risks and visually interpret the EHR prediction results. Experimental results and a case study on heart failure risk prediction tasks show that our system not only outperforms the state-of-the-art deep learning–based risk prediction models but also associates the intuitive visualization design, thus paving the way for interactive, interpretable, and accurate clinical risk predictions. This study can be regarded as an initial step, and there are many research opportunities to be further explored and pursued. The following subsections provide an in-depth discussion of our study in terms of technical challenges and future research.

Issues With EHR Prediction and Visualization

Predicting the risk of certain diseases and interpreting the results are still open questions in the health care community. One major challenge is the false prediction made by the deep learning model. In our case study, the domain expert was surprised that common causes such as coronary artery disease, hypertension, and diabetes related to heart failure were not seen. This might be because the data set we used did not contain many of these factors. In terms of interpreting deep learning models, uncertainty is becoming an important concern [31,32]. As a result, visualizing the uncertainty in the prediction model can be highly valuable. Even if the model is proven to be accurate, the visualization should address the false cases and present the uncertainty to the medical doctors. For example, revealing how the model fails to predict a specific case would deepen the doctors' understanding of the prediction model's intrinsic mechanisms.

Visualizing Patient Distribution

The projection view aims to provide an overview of patient distribution in the data set by mapping high-dimensional patient data into 2D space. In the present visualization results, we can observe that the patients diagnosed as positive and negative are well separated in the space. However, as mentioned in the feedback from our domain expert, determining the subtypes of the patients is also important in analyzing the patient distribution.

What-If Analysis

One important functionality of DG-Viz is to enable domain experts to test their hypotheses on patients through what-if analysis. In particular, we provide what-if analysis by allowing domain experts to add or remove specific medical codes and compare the changes. However, this interaction still suffers from some drawbacks such as the interaction cost. For example, when experts want to know when and what drugs are added to cause a significant difference in predictions, they must select all the drugs in sequence and add them to different dates to obtain the final result. To address the huge interaction cost, one solution is to develop tools such as interactive lenses [33] to present the results of each combination. In detail, users can obtain the results by binding specific drugs with lenses and covering the lens on a specific date. Moreover, automatically recommending the desirable prediction results (ie, computing all possible combinations of drugs and dates and only preserving the significant results) can also help users to obtain the what-if analysis results efficiently.

Generalization

DG-Viz is capable of visualizing several other EHR data sets such as MIMIC-III [34] and HCUP [35]. It can also be converted and used for similar RNN-based prediction models such as RETAIN [1]. However, as a preliminary prototype, it is not readily applicable to all EHR data sets and prediction models. When we design, implement, and evaluate DG-Viz, we encounter several limitations and challenges, which motivates us to generalize DG-Viz from 2 directions in the future. The first is to introduce adaptive mechanisms that allow the system to accommodate different EHR data sets. Among different EHR data sets, there is a great variety of data distributions, including the number of visits, the number of medical codes in each visit, and temporal intervals. For example, most patients in the data set used in our study had 20 to 50 visits with 1 to 5 medical codes per visit. In MIMIC-III, most patients only have 1 to 3 visits with 20 to 40 medical codes per visit. One way to address this issue is to compute the space layout of these visual elements automatically based on the data distribution. When the number of visits and medical code is too large, extra work such as aggregating and filtering (eg, only show medical code with the highest contribution) can also be adopted. In addition,

integrating the domain knowledge from experts with the prediction model through visualization and interaction is an important and interesting direction for us to investigate in the future.

Conclusions

In this work, we present DG-Viz, an interactive clinical prediction system, which brings together the power of deep learning (ie, a domain knowledge–guided RNN-based model) and visual analytics to predict clinical risks and visually interpret the EHR prediction results. We presented a graph attention module to dynamically attend to a subgraph of the whole medical knowledge graph, which can provide more domain information and thus significantly improve DG-RNN's performance. We introduced a global max-pooling operation to DG-RNN to make our prediction model more accurate. We designed, implemented, and evaluated a visual analytics tool to present the EHR data, revealing the knowledge graph network, and interpret the prediction results. Experimental results and a case study on heart failure risk prediction tasks show that our system not only outperforms the state-of-the-art deep-learning–based risk prediction models but also associates the intuitive visualization design, thus paving the way for interactive, interpretable, and accurate clinical risk predictions.

Acknowledgments

This project was funded in part under a grant with Lyntek Medical Technologies Inc. The authors would like to thank Dr Jian Chen for her insightful discussions and suggestions on our visualization design. The authors also thank the reviewers for their thorough review and constructive comments on our manuscript.

Authors' Contributions

PZ and BQ conceived and supervised the project. CY and PZ developed the deep learning model. RL, CY, and SY developed the visual analytics system. RL and CY conducted the experiments. RL, CY, and PZ analyzed experimental results. SY provided medical expertise and conducted the case study. RL, CY, and PZ wrote the first draft of the manuscript. All authors read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A demo video of DG-Viz.
[MP4 File (MP4 Video), 6596 KB - jmir_v22i9e20645_app1.mp4]

References

1. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart WF. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. ArXiv 2016:- epub ahead of print [FREE Full text]
2. Choi E, Bahadori MT, Song L, Stewart WF, Sun J. GRAM: Graph-Based Attention Model for Healthcare Representation Learning. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: 10.1145/3097983.3098126]
3. Ma F, You Q, Xiao H, Chitta R, Zhou J, Gao J. Kame: Knowledge-Based Attention Model for Diagnosis Prediction in Healthcare. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018 Presented at: CIKM'18; October 22-26, 2018; Turin, Italy. [doi: 10.1145/3269206.3271701]
4. Kwon BC, Choi M, Kim JT, Choi E, Kim YB, Kwon S, et al. Retainvis: visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. IEEE Trans Vis Comput Graph 2018 Aug 20:299-309. [doi: 10.1109/TVCG.2018.2865027] [Medline: 30136973]
5. Che Z, Kale D, Li W, Bahadori MT, Liu Y. Deep Computational Phenotyping. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: KDD'15; August 10-13, 2015; Sydney, Australia. [doi: 10.1145/2783258.2783365]
6. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
7. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: Diagnosis Prediction in Healthcare via Attention-Based Bidirectional Recurrent Neural Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: 10.1145/3097983.3098088]
8. Kim J, Park YR, Lee JH, Lee J, Kim Y, Huh JW. Development of a real-time risk prediction model for in-hospital cardiac arrest in critically ill patients using deep learning: retrospective study. JMIR Med Inform 2020 Mar 18;8(3):e16349 [FREE Full text] [doi: 10.2196/16349] [Medline: 32186517]

9. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2016 Presented at: PAKDD'16; April 19-22, 2016; Auckland, New Zealand. [doi: 10.1007/978-3-319-31750-2_3]
10. Baytas IM, Xiao C, Zhang X, Wang F, Jian AK, Zhou J. Patient Subtyping via Time-Aware LSTM Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: 10.1145/3097983.3097997]
11. Munkhdalai T, Liu F, Yu H. Clinical Relation Extraction Toward Drug Safety Surveillance Using Electronic Health Record Narratives: Classical Learning Versus Deep Learning. JMIR Public Health Surveill 2018 Apr 25;4(2):e29 [FREE Full text] [doi: 10.2196/publichealth.9361] [Medline: 29695376]
12. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. Sci Rep 2018 Apr 17;8(1):6085 [FREE Full text] [doi: 10.1038/s41598-018-24271-9] [Medline: 29666385]
13. Lipton ZC, Kale D, Wetzel R. Directly Modeling Missing Data in Sequences with RNNs: Improved Classification of Clinical Time Series. In: Proceedings of the 1st Machine Learning for Healthcare Conference. 2016 Presented at: PMLR'16; August 19-20, 2016; Los Angeles, CA, USA URL: http://proceedings.mlr.press/v56/Lipton16.html
14. Rais A, Viana A. Operations research in healthcare: a survey. Int T Oper Res 2011;18(1):1-31. [doi: 10.1111/j.1475-3995.2010.00767.x]
15. Guo S, Jin Z, Chen Q, Gotz D, Zhang H, Cao N. Visual Anomaly Detection in Event Sequence Data. In: IEEE International Conference on Big Data. 2019 Presented at: Big Data'19; December 9-12, 2019; Los Angeles, CA, USA p. 1125-1130. [doi: 10.1109/bigdata47090.2019.9005687]
16. Ernst P, Siu A, Weikum G. KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences. BMC Bioinformatics 2015 May 14;16:157 [FREE Full text] [doi: 10.1186/s12859-015-0549-5] [Medline: 25971816]
17. Yin C, Zhao R, Qian B, Lv X, Zhang P. Domain Knowledge Guided Deep Learning with Electronic Health Records. In: International Conference on Data Mining. 2019 Presented at: ICDM'19; November 8-11, 2019; Beijing, China. [doi: 10.1109/icdm.2019.00084]
18. Zhu Z, Yin C, Qian B, Cheng Y, Wei J, Wang F. Measuring Patient Similarities via a Deep Architecture with Medical Concept Embedding. In: 16th International Conference on Data Mining. 2016 Presented at: ICDM'16; December 12-15, 2016; Barcelona, Spain. [doi: 10.1109/icdm.2016.0086]
19. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Dec 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
20. Zhou H, Young T, Huang M, Zhao H, Xu J, Zhu X. Commonsense knowledge aware conversation generation with graph attention. Int Joint Conf Artif Intell 2018:4623-4629. [doi: 10.24963/ijcai.2018/643]
21. Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008:2579-2605 [FREE Full text]
22. Kwon BC, Verma J, Perer A. Peekquence: Visual Analytics for Event Sequence Data. In: Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics. 2013 Presented at: IDEA'16; August 14th, 2016; San Francisco, CA, USA. [doi: 10.1145/2501511.2501518]
23. Hochheiser H, Shneiderman B. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. Inf Vis 2016 Jul 21;3(1):1-18. [doi: 10.1057/palgrave.ivs.9500061]
24. Javed W, McDonnel B, Elmqvist N. Graphical perception of multiple time series. IEEE Trans Vis Comput Graph 2010;16(6):927-934. [doi: 10.1109/TVCG.2010.162] [Medline: 20975129]
25. DGViz. YouTube. URL: https://www.youtube.com/watch?v=H3aJ_HplqJ8&feature=youtu.be [accessed 2020-09-09]
26. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating Embeddings for Modeling Multi-Relational Data. In: Advances in Neural Information Processing Systems. 2013 Presented at: NIPS'13; December 5-8, 2013; Lake Tahoe, Nevada, United States p. 2787-2795.
27. Cho K, Van MB, Bahdanau D, Bengio Y. On the properties of neural machine translationncoder-decoder approaches. arXiv 2014:1259 epub ahead of print. [doi: 10.3115/v1/w14-4012]
28. yinchangchang / DG-RNN. GitHub. URL: https://github.com/yinchangchang/DG-RNN [accessed 2020-09-09]
29. DGViz. GitHub. URL: https://github.com/DGViz/DGViz.github.io [accessed 2020-09-09]
30. Tang F, Xiao C, Wang F, Zhou J. Predictive modeling in urgent care: a comparative study of machine learning approaches. JAMIA Open 2018 Jul;1(1):87-98 [FREE Full text] [doi: 10.1093/jamiaopen/ooy011] [Medline: 31984321]
31. Gal Y. Uncertainty in Deep Learning. Cambridge Machine Learning Group. 2016. URL: http://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf [accessed 2020-09-09]
32. Leibig C, Allken V, Ayhan MS, Berens P, Wahl S. Leveraging uncertainty information from deep neural networks for disease detection. Sci Rep 2017 Dec 19;7(1):17816 [FREE Full text] [doi: 10.1038/s41598-017-17876-z] [Medline: 29259224]
33. Tominski C, Gladisch S, Kister U. A Survey on Interactive Lenses in Visualization. In: 16th Eurographics Conference on Visualization. 2014 Presented at: EuroVis'14; June 9-13, 2014; Swansea, UK. [doi: 10.2312/eurovisstar.20141172]
34. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]

35. Healthcare Cost and Utilization Project (HCUP). The Agency for Healthcare Research and Quality. 2012. URL: https://www.ahrq.gov/data/hcup/index.html [accessed 2020-09-09]

Abbreviations

2D: two-dimensional
AUROC: area under a receiver operating characteristic
DG-RNN: domain-knowledge–guided recurrent neural network
EHR: electronic health record
GRAM: graph-based attention model
GRU: gated recurrent unit
ICD-9: International Classification of Diseases, Ninth Revision
KAME: knowledge-based attention model
LR: logistic regression
LSTM: long short-term memory
RETAIN: reverse time attention model
RF: random forest
RNN: recurrent neural network

Edited by G Eysenbach; submitted 29.05.20; peer-reviewed by F Ma, EH Kim; comments to author 23.06.20; revised version received 07.07.20; accepted 26.07.20; published 28.09.20.

Please cite as:
Li R, Yin C, Yang S, Qian B, Zhang P
Marrying Medical Domain Knowledge With Deep Learning on Electronic Health Records: A Deep Visual Analytics Approach
J Med Internet Res 2020;22(9):e20645
URL: http://www.jmir.org/2020/9/e20645/
doi:10.2196/20645
PMID:32985996

Viewpoint

Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint

Jingwen Zhang[1,2], PhD; Yoo Jung Oh[1], MA; Patrick Lange[3], MSc; Zhou Yu[3], PhD; Yoshimi Fukuoka[4], PhD, RN, FAAN

[1]Department of Communication, University of California, Davis, Davis, CA, United States

[2]Department of Public Health Sciences, University of California, Davis, Davis, CA, United States

[3]Department of Computer Science, University of California, Davis, Davis, CA, United States

[4]Department of Physiological Nursing, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Jingwen Zhang, PhD
Department of Communication
University of California, Davis
One Shields Avenue
Davis, CA, 95616
United States
Phone: 1 530 754 1472
Email: jwzzhang@ucdavis.edu

Abstract

Background: Chatbots empowered by artificial intelligence (AI) can increasingly engage in natural conversations and build relationships with users. Applying AI chatbots to lifestyle modification programs is one of the promising areas to develop cost-effective and feasible behavior interventions to promote physical activity and a healthy diet.

Objective: The purposes of this perspective paper are to present a brief literature review of chatbot use in promoting physical activity and a healthy diet, describe the AI chatbot behavior change model our research team developed based on extensive interdisciplinary research, and discuss ethical principles and considerations.

Methods: We conducted a preliminary search of studies reporting chatbots for improving physical activity and/or diet in four databases in July 2020. We summarized the characteristics of the chatbot studies and reviewed recent developments in human-AI communication research and innovations in natural language processing. Based on the identified gaps and opportunities, as well as our own clinical and research experience and findings, we propose an AI chatbot behavior change model.

Results: Our review found a lack of understanding around theoretical guidance and practical recommendations on designing AI chatbots for lifestyle modification programs. The proposed AI chatbot behavior change model consists of the following four components to provide such guidance: (1) designing chatbot characteristics and understanding user background; (2) building relational capacity; (3) building persuasive conversational capacity; and (4) evaluating mechanisms and outcomes. The rationale and evidence supporting the design and evaluation choices for this model are presented in this paper.

Conclusions: As AI chatbots become increasingly integrated into various digital communications, our proposed theoretical framework is the first step to conceptualize the scope of utilization in health behavior change domains and to synthesize all possible dimensions of chatbot features to inform intervention design and evaluation. There is a need for more interdisciplinary work to continue developing AI techniques to improve a chatbot's relational and persuasive capacities to change physical activity and diet behaviors with strong ethical principles.

(J Med Internet Res 2020;22(9):e22845) doi:10.2196/22845

KEYWORDS

chatbot; conversational agent; artificial intelligence; physical activity; diet; intervention; behavior change; natural language processing; communication

Introduction

Background

Physical inactivity and an unhealthy diet continue to be some of the leading risk factors for noncommunicable diseases (NCDs), such as cardiovascular disease, diabetes, and obesity [1,2], and death worldwide [3]. NCDs account for seven out of 10 deaths worldwide [3] and pose a substantial economic burden [4]. The prevalence of physical inactivity and an unhealthy diet varies considerably within and across countries. The United States is one of the countries experiencing a rapid rise in these risks. Nearly 80% of American adults do not meet the guidelines for both aerobic and muscle-strengthening activities [5], and the prevalence of overweight or obesity reached 71.6% in 2016 [6]. Therefore, developing cost-effective and feasible lifestyle interventions is urgently needed to reduce the prevalence [7].

Lifestyle modification programs have consistently evolved with emerging digital and communication technologies [8-13]. In the past two decades, there has been a large number of published studies using internet and mobile-based behavior interventions to support the effectiveness of using digital technologies to deliver intervention materials to diverse populations [8,14]. In recent years, the use of artificial intelligence (AI) and associated computational techniques has become the new frontier in expanding the landscape of health care and interventions [15].

Definition and Applications of an AI Chatbot

AI chatbots, also called conversational agents, employ dialog systems to enable natural language conversations with users by means of speech, text, or both [16]. Conceptually, the core technical capacity of AI chatbots is different from that of embodied virtual conversational agents or avatars that emphasize on synthesizing multimodal signals (eg, images, videos, and sounds) to simulate human face-to-face communication. In this paper, we focused on developing the AI chatbot's core feature of natural language conversation to facilitate more flexible information exchange between humans and the chatbot. The conversational capacity can range from constrained conversation (ie, users can only respond by selecting predefined conversational lines) to unconstrained conversation (ie, users can respond freely by inputting natural language conversational lines).

AI chatbots can be deployed in the form of mobile apps on smartphones, thus making programs available 24/7. AI chatbots have been rapidly transforming multiple fields, including business [17], governance [18], education [19], and health care [16,20]. As the top platforms supporting chatbot development, Amazon Alexa had more than 100,000 programs and Facebook Messenger had more than 300,000 active chatbots as of 2019, many of which are for health care and wellbeing. For instance, in April 2020, the World Health Organization launched a chatbot on Facebook Messenger to combat misinformation and to offer instant and accurate information about COVID-19 [21].

As chatbots increasingly become a convenient digital communication channel, they open up many opportunities for delivering personalized behavior change programs for disease prevention and health promotion on a large scale. Beyond connectivity and feasibility, the advantages of AI chatbot programs lie essentially in the computational power to develop and deliver personalized interventions [22-24]. Such interventions have the potential to overcome several limitations in the traditional paradigm of nonpersonalized interventions, as they are designed based on understanding individual characteristics and behavior trajectories and can incrementally adapt intervention strategies based on contextual conditions and personal cognitive and emotional states over time. In other words, chatbot technologies have the potential to "understand" individuals through natural human conversations, persuade individuals to change, and build sustaining supportive relationships for maintaining healthy behaviors.

AI Chatbots for Health Care and Lifestyle Modification Programs

Chatbots for promoting physical activity and a healthy diet are designed to achieve behavior change goals, such as walking for certain times and/or distances and following healthy meal plans [25-29]. Although no systematic review of chatbots for lifestyle modification programs has been published, there are several reviews on chatbots covering health care issues ranging from mental health support and smoking cessation to disease diagnosis [16,30]. Owing to the different natures of targeted behaviors, some chatbots were mainly designed to provide information and knowledge [31], whereas others were developed based on established mental health intervention programs such as cognitive behavioral therapy [32]. One relevant review [33] focused on discussing the development of embodied conversational agents for a healthy lifestyle, and pointed out that the interpretation and application of behavior change theories were usually not reported.

Most previous chatbot research relied on either finite-state (ie, dialog consisting of a sequence of predetermined steps or states) or frame-based systems (ie, dialog is not predetermined but dependent on the content of the user's input and the information that the system has to elicit) [34-36]. Such systems are restrained in their ability to allow free conversations, primarily due to the lack of large training data sets on human-to-human conversations in domains involving behavior changes.

The recent success of large pretrained language models, such as Bidirectional Encoder Representations from Transformers (BERT) developed by Google [37] and Generative Pre-Training-2 (GPT2) developed by Open AI [38], provides promising opportunities to incorporate language priors to down-stream natural language processing (NLP) tasks. For instance, several papers have shown that pretrained models can be tailored for task-oriented dialog generation, such as for conversations about restaurant recommendations and donation persuasion [39,40]. BERT and GPT2 are giant neural network models trained with large text data sets using self-supervised task objectives, such as recovering masked tokens and predicting the next word. As these models operate on representation space and do not have access to symbolic common-sense information, they produce outputs that are difficult for humans to interpret and can make errors that violate common senses in specific domains. One general direction to advance this field is to build systems that incorporate pretrained models to facilitate building

dialogs that are specific for communicating and persuading users to adopt regular physical activity and a healthy diet.

To advance the science of developing effective and ethical AI chatbots for health behavior changes, especially within the context of improving physical activity and healthy eating behaviors, we provide a theoretical perspective and a model to guide the development and evaluation of AI chatbots for behavior changes. The aims of this perspective paper are threefold as follows: (1) to briefly summarize the current state of applications of AI chatbots in promoting physical activity and a healthy diet; (2) to propose the AI chatbot behavior change model developed by our research team; and (3) to address ethical considerations and principles.

Methods

Preliminary Review of AI Chatbot–Based Physical Activity and Diet Interventions

To provide a background of the current state of chatbot-based behavior interventions for physical activity and diet, we conducted a rapid preliminary literature review using four electronic databases (PubMed, EMBASE, Web of Science, and ACM Digital Library) on August 24, 2020. We used a combination of keywords to identify peer-reviewed studies related to AI chatbots for physical activity or diet (ie, ["chatbot" OR "conversational agent" OR "conversational system" OR "dialog system" OR "dialogue system" OR "relational agent"] AND ["physical activity" OR "exercise" OR "diet" OR "nutrition"]). We included only full-length articles that reported chatbot-based physical activity or diet interventions and were written in English. One researcher initially screened study titles and abstracts to determine eligibility for inclusion. Thereafter, two researchers reviewed the full texts of the included studies to further determine their relevance and coded study features. The two researchers discussed their disagreements throughout the coding process and agreed upon the final results.

In total, the search returned 108 articles from the four databases, with 15 published articles in 2020, 26 in 2019, 15 in 2018, 14 in 2017, five in 2016, and the remaining 33 from 2015 or before. After the screening, 101 (93.5%) articles were excluded for the following reasons: commentary or opinion pieces, scoping reviews, or empirical studies that addressed health domains other than physical activity and diet (eg, chatbots assisting diagnostic tasks or offering mental health interventions or treatment).

Characteristics of AI Chatbot Interventions

We identified seven articles reporting six unique chatbots to increase physical activity and/or adoption of a healthy diet (Multimedia Appendix 1). Two papers reported on the same chatbot called Assistant to Lift your Level of activitY (Ally). One protocol [41] described the study design and one reported the actual optimization randomized controlled trial (RCT) (n=274) [42] for evaluating the effects of Ally in helping users to reach personalized daily step goals. The results showed that the intervention component of daily cash incentives delivered by Ally increased step-goal achievement. However, 30% of participants stopped using the app over the course of the study, presenting a challenge for the chatbot's ability to engage participants. In contrast, another study reported the results of an RCT (n=106) [43] to evaluate the Healthy Lifestyle Coaching chatbot. The findings demonstrated that this chatbot was effective in increasing physical activity after 12 weeks of the intervention among office workers. The remaining four studies employed pretest-posttest designs. One feasibility study (n=23) [44] tested Tess, a behavioral coaching chatbot, in assisting adolescent patients to cope with weight management and prediabetes symptoms. Patients actively engaged with the chatbot, reported experiencing positive progress toward their goals, and deemed the chatbot helpful. One proof-of-concept study [45] reported on the Paola chatbot, which provided educational messages on physical activity and diet, weekly check-ins, and answers to user questions. The results showed that participants reported relevant weight loss and improved diet. Another validation study [46] reported on the CoachAI chatbot, which provided social and tailored health coaching support, and found this chatbot to be effective, especially among users with high engagement levels. Lastly, a chatbot named Reflection Companion delivered daily adaptive mini-dialogs and activity graphs to promote self-reflections. The conversations successfully triggered self-reflections that led to increased motivation, empowerment, and adoption of physical activity behaviors (eg, walking to a grocery store instead of taking a car) [26].

The above-reviewed chatbots showed preliminary evidence supporting the efficacy of using chatbots to deliver physical activity and diet interventions. It is worth noting that four out of seven (57.1%) studies reported chatbots as the only intervention used to deliver behavior change strategies [26,43,44,46], whereas the other three articles reported chatbots as an auxiliary component complementing other intervention approaches such as messages and conversations delivered by human facilitators [41,42,45] (Multimedia Appendix 1). The reviewed chatbots were designed with different theoretical components and varied in their abilities to engage in natural language conversations, relationship building, and emotional understanding. Overall, owing to a lack of reporting on the details of the theoretical framework and a limited number of RCT evaluations, it is difficult to systematically evaluate how different design theories and factors contribute to intervention efficacy. Based on this preliminary review, we identified a lack of systematic thinking in the development of AI chatbots for lifestyle behavior changes.

None of the studies reported in detail how they developed the chatbot program and none discussed ethical considerations regarding issues such as transparency, privacy, and potential algorithmic biases. Consequently, it remains unclear how to evaluate a chatbot's efficacy, the theoretical mechanisms through which chatbot conversations influence users, and potential ethical problems. To address these gaps, in the next section, we present our theoretical framework that delineates design considerations, core theoretical components supporting a chatbot's conversational capacity, multiple dimensions for usability and outcome evaluations, and ethical principles that need to be emphasized to guide development in this emerging field.

Results

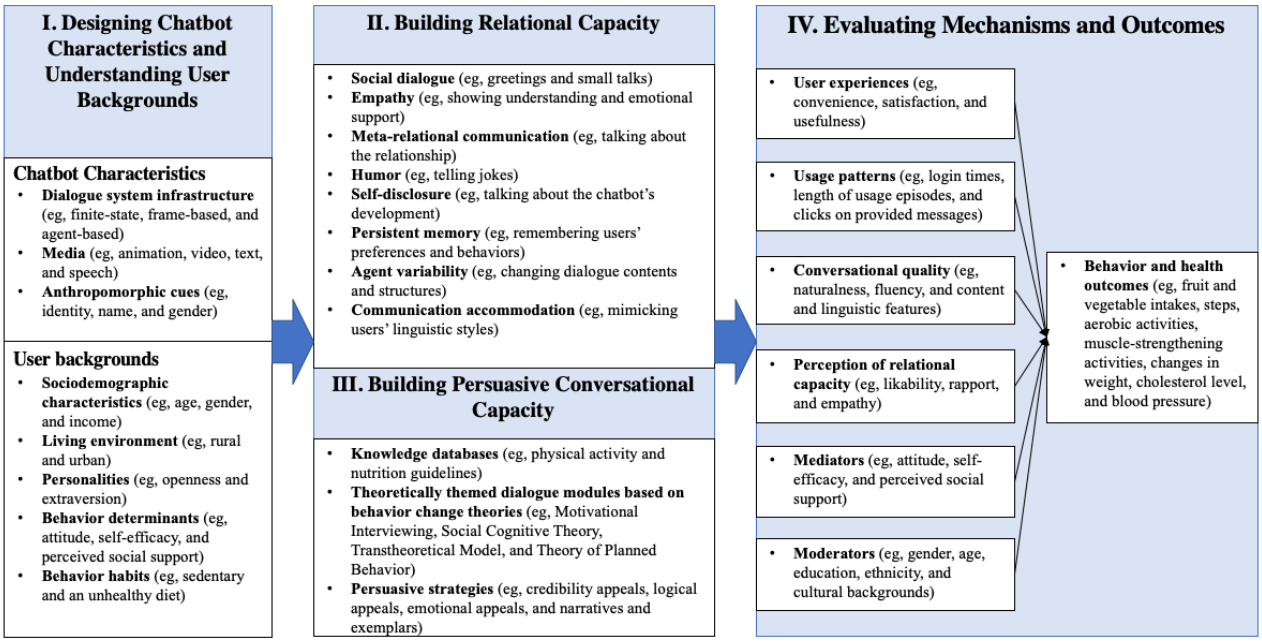
AI Chatbots as Persuasive Technology

We conceptualize behavior change chatbots as a type of persuasive technology [14], which is more complicated than designing a social chatbot to engage in general conversations (eg, talking about movies or weather) [47]. Persuasive technology broadly refers to computer systems that are designed to change the attitudes and behaviors of users [48]. Behavior change chatbots thus aim to change users' specific behaviors through engaging in conversations and delivering information and persuasive messages. In this regard, we propose that the chatbot dialog system needs to encompass two core capacities, including the relational capacity to establish and maintain a professional relationship with the user and the persuasive conversational capacity to change behaviors. Below, we describe a theoretical framework that elaborates on these two capacities and guides the design of AI chatbots for promoting physical activity and a healthy diet.

Theoretical Framework: The AI Chatbot Behavior Change Model

Figure 1 shows the theoretical framework for improving physical activity and diet using AI chatbots. We named this framework the AI chatbot behavior change model, which includes the following four major components: (1) designing chatbot characteristics and understanding user backgrounds; (2) building relational capacity; (3) building persuasive conversational capacity; and (4) evaluating mechanisms and outcomes. The four high-level components are specified in sequence to guide the design and evaluation of chatbots. This proposed model is based on reviewing relevant chatbot studies, recent developments in human-AI communication research, and innovations in NLP, as well as our own clinical and research experience and findings [23,49-54].

Figure 1. The artificial intelligence chatbot behavior change model.



Designing Chatbot Characteristics and Understanding User Background

Chatbots are set up to mimic the characteristics of human-human conversations. Designing a chatbot requires both system-related and agent-related considerations. Upon choosing a system infrastructure (eg, finite-state, frame-based, and agent-based infrastructure) and media (eg, animation, video, text, and speech), the characteristics of a chatbot (eg, identity, name, and gender) can be specified. In the past, researchers have experimented with using a robot [32], animal [55], or human identity, ranging in degrees of applying anthropomorphic cues [56].

The computers are social actors (CASA) paradigm [57] and the uncanny valley effect (UVE) [58,59] are the most widely used theoretical frameworks for studying human-computer interactions. While the CASA paradigm assumes that humans can develop positive social relations with computer systems as the human familiarity of the system increases, the UVE argues that too much human familiarity would bring feelings of eeriness and discomfort. To increase a chatbot's social presence, some studies framed chatbots as peers and gave them gendered names (eg, Anna for female [27]). Deciding what name to call the chatbot and whether to frame it as a human peer or as a transparent bot system requires careful consideration. Our recent work [52] suggests that as AI chatbots are quickly adopting human conversational capacities, the perceived identity of a chatbot has significant effects on the persuasion outcome and interpersonal perceptions. Furthermore, our study findings suggest that users respond better if the chatbot's identity is clearly presented. This may be because users can develop more agency and control if they know how to respond to the

conversational partner by applying different communication norms. For instance, if a chatbot is presented with a human identity and tries to imitate human inquiries by asking personal questions, the UVE can be elicited and make people feel uncomfortable [52]. However, contrary findings have also been identified as some studies show evidence that people respond well and disclose more personal information if the chatbot is presented as a bot and can also display emotions [60,61]. Identifying the boundary conditions for chatbot identity and disclosures in various application contexts requires more research to provide empirical findings.

Designing a personalized chatbot system requires the understanding of each individual user's background (eg, sociodemographic characteristics, living environment, and personality), behavior determinants, and habits [62-65]. The assumption is that a personalized intervention is more effective as it tailors both behavior change strategies and persuasive messaging to each user's unique background and needs to achieve personally optimized outcomes [63]. In general, the first component serves to set up the chatbot characteristics and collect useful user background information to inform the development of algorithms supporting the second component and the third component. Theoretically, user background information can be incorporated as contextual information to develop algorithms to generate personalized relational messages and persuasive messages. Which characteristics can be used to tailor which messages depends largely on the target population's needs and preferences [66,67]. Past literature has examined a number of useful characteristics for personalized influences, such as using different persuasive strategies to appeal to different personality traits [53,68] or setting personalized change goals based on behavior habits [42]. In the realm of physical activity chatbot interventions, the Ally chatbot system by Kramer et al was able to welcome each participant using personalized messages and track individual physical activity using the smartphone's built-in accelerometer [42]. The system specifically set a personalized activity goal slightly above the participant's current average activity level. Along this line, the application of control systems engineering in modeling individuals' behavior states and adapting personalized goals over time is a promising approach [22].

Building Relational Capacity

In order to use an AI chatbot as a social conversational agent, we emphasize designing the system's relational capacity in chatbot and user interactions [29,69-72]. Bickmore et al provided extensive discussions on the principles of building relational capacity in behavior change agents, such as using social dialog, empathy, meta-relational communication (talk about the relationship), humor, self-disclosure, persistent memory, and agent variability [70]. One of their studies showed that when compared to a nonrelational agent, a relational agent was more respected, liked, and trusted, which led to more positive behavior changes [29].

It is worth noting that most of the reported relational agents are embodied virtual agents, taking on specific anthropomorphic cues and nonverbal behaviors but using restricted scripted dialog designs. It remains less clear what relational capacity a nonembodied chatbot can achieve just through natural language conversations. Recent endeavors to accelerate natural conversations in everyday social companion chatbots have yielded promising results. One study reported that users of a companion chatbot (called "Replika") perceived the chatbot to be human-like, intelligent, supportive, and able to facilitate social connection. However, UVEs also emerged as some users felt that the chatbot's conversation was too natural and thus "creepy" [73]. In another case study that analyzed user reviews of the Amazon chatbot device, researchers found that over half of the reviewers referred to the chatbot using the personified name "Alexa," and as users' social interactions with the device increased, a greater level of personification occurred, which was associated with increased product satisfaction [74]. This suggests that people tend to personify Alexa, which is in line with the CASA paradigm. As a chatbot's natural conversational abilities continue to rapidly improve, it is likely that relational capacity building can lead to better user engagement and retainment, despite other technological limitations.

To scale up the relational capacity in chatbots, conversational norms and relational strategies need to be built into the system. One approach can be through extracting patterns from longitudinal human-human conversations and drawing on theories from interpersonal communication and the latest human-AI communication research [75,76]. For example, the integrated model of advice giving [77,78] and the communication accommodation theory [79,80], combined with the chatbot's capacity of persistent memory (eg, storing conversation history) and variability (eg, changing conversation content and structure), can provide useful insights in guiding the structure of conversations and specific choices in linguistic, semantic, and sentence styles.

Building Persuasive Conversational Capacity

Programs delivered by chatbots need to possess the core knowledge structures and intervention messages used in traditional approaches. Building behavior change messages into chatbot conversations first requires curating knowledge databases regarding physical activity and dietary guidelines. Thereafter, relevant behavior change theories need to be applied to generate themed dialog modules (eg, goal setting, motivating, and proving social support). Commonly used behavior change theories include motivational interviewing [81], the social cognitive theory [56], the transtheoretical model [82], and the theory of planned behavior [83]. One approach is to design human-human conversation episodes based on addressing each of the theoretical concepts (eg, a human interventionist providing social support to a participant) and to develop dialog modules that mimic such conversations.

In addition to delivering theory-based intervention messages, chatbots' efficacy in eliciting behavior changes can be augmented by employing persuasive messaging strategies [84]. This thinking stems from the line of work in public health communication that aims to integrate behavior change theories and message effect theories (ie, theories that direct the selection of specific persuasive appeals and message features to enhance the effectiveness of communication) [85]. Persuasive strategies are designed to motivate behavior changes and are nuanced

messaging choices to enhance attention, trust, and engagement, or to influence cognitive and emotional reactions. Persuasive strategies are important in shaping, changing, and reinforcing people's attitudes and behaviors. Previous research has shown that even simply asking questions about a behavior can lead to changes in the behavior, known as the "question-behavior" effect. For instance, one study found that asking people questions about exercise led to an increase in self-reported exercise [86]. Although this effect was small and based on survey reports, it suggests that questions can function as a reminder or cue to action. Thus, one task of chatbots can be to ask questions to allow users to reflect and then get motivated for behavior change. More persuasive strategies can be embedded into theoretically themed dialog modules, such as using classical rhetorical appeals [53,68], including credibility appeals (eg, showing messages from sources that the target audiences trust), logical appeals (eg, providing reasoning and evidence for benefits of physical activity and a healthy diet), and emotional appeals (eg, using fear, guilt, or hope appeals for motivation). In addition, specific persuasive messaging strategies, such as using narratives and exemplars (eg, telling stories to enhance self-efficacy), can also enhance personal involvement and engagement. For example, to augment the approach of motivational interviewing, we can consider using credibility appeal to strengthen user's trust in the chatbot, so that they become more comfortable in disclosing thoughts. In addition, to augment the approach of social cognitive theory, we can consider constructing narrative exemplars in terms of talking about relevant peers' successful experiences to boost participants' self-efficacy.

One common limitation of traditional programs is the static nature of persuasive messages, because of infrequent measurements of behaviors and users' behavior change stages. Chatbots deployed on smartphones can address this limitation by utilizing ecological momentary assessment methods, in-built accelerometers, GPS, and other sensors, in addition to collecting user-reported data from convenient short surveys through the smartphone. For instance, research has shown that an accelerometer installed on smartphones is accurate for tracking step count [9] and that GPS signals can be used to estimate activity levels [87]. By objectively tracking and modeling activity patterns, developing machine learning models to update personalized goals and persuasive messages becomes feasible. Our work has shown that by using steps and physical activity intensity records, models can predict an individual's probability of disengagement from the intervention [88]. Further, by using NLP and cluster analysis, we could differentiate individuals' motivation levels as communicated in the conversation to tailor intervention maintenance programs [23]. These results indicate that AI chatbots can adapt not only behavior change goals and techniques, but also conversational styles (eg, emotional tones) based on learning from a user's natural language inputs to enhance the engagement and effectiveness of messages.

Furthermore, rapid progress in mobile health technologies and functions has enabled the design of just-in-time adaptive interventions (JITAIs) [24]. JITAI designs in combination with real-time data from ecological momentary assessment, in-built accelerometers, GPS, and/or other sensors will allow chatbots to customize the timing, amount, content, and frequency of the intervention, by adapting each individual's internal and external changes over time. However, a recent scoping review of health care chatbots showed that the use of JITAIs in designing and evaluating chatbots in health care in general and promoting physical activity and a healthy diet in particular is sparse, suggesting that future research needs to consider using more of these adaptive approaches [89].

Evaluating Mechanisms and Outcomes

Figure 1 shows the proposed dimensions for evaluating AI chatbot programs, including user experiences, usage patterns, conversational quality, perception of relational capacity, mediators, moderators, and behavior outcomes. All dimensions can be considered to improve the chatbot design and to understand theoretical mechanisms for how chatbot programs change behaviors.

User experiences concern users' subjective evaluations of the overall interaction with the system. Many scales have been developed to assess a program's convenience, satisfaction, usefulness, helpfulness, etc [90]. Usage patterns document objectively logged data regarding users' interactions with the system, including records such as login times, length of usage episodes, and clicks on provided messages [91]. Conversational quality can be measured from users' subjective evaluation of the conversation's coherence, naturalness, and fluency. In addition, objective content and linguistic analyses of conversations can be used to assess specific dimensions of conversations such as the length of conversations and amount of information exchanged. Perception of relational capacity evaluates users' perception of the chatbot identity and its relational capacity. Some studies have assessed the extent to which users deem a chatbot as a friend and its likability, as well as its capacity to achieve rapport, relate to human emotions, and show empathy [92-94]. Mediators refer to factors that help to explain why and how chatbot interventions are effective in promoting physical activity and a healthy diet. Chatbots can lead people to change their perceptions of themselves (eg, attitude, self-efficacy, and perceived social support) and help people to shape and form new behavior choices and patterns. These intermediate changes are important to explain the mechanisms of chatbot interventions and to design more effective interventions in the future. Moderators often refer to user characteristics such as gender, age, education, ethnicity, and cultural backgrounds, and these subgroups (eg, men vs women) may respond to a chatbot intervention differently. Advances in digital technologies can unintentionally reinforce or increase existing health disparities [95]. Thus, evaluating moderation effects is crucial in documenting a potential digital divide or lack thereof. Lastly, behavior outcomes denote actual changes in behavior and health, including diet (eg, fruit and vegetable intake five times per day [96]) and physical activity changes (eg, daily steps, aerobic activities, and muscle-strengthening activities [97]), and subsequent effects on health outcomes such as weight and blood pressure.

Ethical Considerations

General ethical principles and guidelines for AI's integration in health care need to be adopted in designing chatbots for

lifestyle modification programs [15,98-100]. Key ethical considerations include having transparency and user trust, protecting user privacy, and minimizing biases. To gain the trust of users, credibility and transparency have to be established and communicated. A brief introduction of the intention and expertise of the research team behind the chatbot may enhance its credibility. Similarly, providing users with high-level explanations on the machine learning algorithms and data processing can help increase transparency. Protection of user privacy faces multiple challenges. There is emerging research showing that multiple sets of anonymized data can be modeled to reidentify individuals [101,102]. In the context of chatbot interventions, high standards of confidentiality and data anonymization, such as differential privacy [103], need to be adopted to decrease the risks of reidentification.

Within the context of persuasive health technology, beyond considering the general ethical principles in AI described above, another central framework that needs to be incorporated is the bioethics framework [104] consisting of (1) nonmaleficence, (2) beneficence, (3) respect for autonomy, and (4) justice. *Nonmaleficence* means the obligation to not inflict any harm or incur the least harm possible to reach a beneficial outcome. *Beneficence* denotes a moral obligation to act for others' benefits. Building a commitment to nonmaleficence and beneficence means the chatbot's intent is to benefit users with information, knowledge, care, and guidance, as well as to take positive steps to prevent and remove harm from the user. For example, chatbots need to be designed to understand expressions from users that indicate they may be undergoing difficult situations requiring human moderators' help. Specifically, it is important to foresee and preemptively plan for the possibility that technical and algorithmic errors can occur, and it is pivotal to have human moderators in place to monitor user engagement regularly and be able to connect with users when challenging situations arise. *Respect for autonomy* means that the user has the capacity to act intentionally with understanding and without being controlled or manipulated by the chatbot. This specifies that users should be provided with full transparency about the intervention's goals, methods, and potential risks. Given the complexity in AI and technological designs, researchers need to strive to provide comprehensible explanations that users can understand and then take decisions for themselves [105]. In addition, users should be fully informed in the consent process and consent form as to how their data will be used to improve the chatbot overtime during or even after the intervention and should be given the opportunity to opt out of having their data used in this manner. *Commitment to justice* requires researchers to consider the technology's equity access and benefits to different populations, especially the consideration of high-needs users who are lower in socioeconomic status and digital literacy, or users with disabilities that could impact their interaction with chatbots. It is thus recommended that underserved populations, especially racial and ethnic minority groups, be represented and involved in all stages of the design and implementation of chatbot interventions to ensure health equity and social justice. Specifically, researchers need to consider applying debiasing strategies in building the dialog system [106,107] and socially aware algorithm design [108]. Given that the research field of using chatbots for behavior changes is still in its nascent phase, ensuring adherence to ethical principles and incorporating corresponding evaluative metrics is necessary for the field to move forward.

Discussion

In this paper, we reviewed and synthesized literature involving lifestyle modification program studies, theories and studies from behavior science and communication research, and technical advancements in AI and NLP, and proposed the *AI chatbot behavior change model*. The strength of the proposed model is that it considers a wide range of chatbot-related components, including chatbot/user characteristics, relational capacity, and persuasive conversational capacity, and points out potential mediating and moderating factors to be evaluated to establish the efficacy of chatbots in changing physical activity and diet behaviors, as well as health outcomes.

To our knowledge, this is the first theoretical framework to provide a guideline to design and evaluate chatbot-based physical activity and diet behavior interventions. We contextualize the framework in the domains of physical activity and diet behaviors because these two are frequent daily behaviors that need continued engagement and monitoring. Chatbots as a convenient conversational tool can connect with people in real time to optimize behavior change interventions.

Moving science forward, systematic approaches and interdisciplinary collaborations are needed to design effective AI-based chatbot physical activity and healthy eating programs. Our proposed theoretical framework is the first step to conceptualize the scope of the work and to synthesize all possible dimensions of chatbot features to inform intervention design. However, when applied in specific contexts, researchers and practitioners can prioritize certain features that are mostly relevant to the target population, according to initial formative research conducted with the target population [54]. In essence, we encourage researchers to select and design chatbot features through working with the target communities using stakeholder-inclusive and participatory design approaches [109,110]. We think such inclusive approaches are much needed and can be more effective in bringing benefits while minimizing unexpected inconvenience and potential harms to the community. In this regard, we do not mean that every new chatbot program has to be developed from scratch. Previously established effective programs and their highlighted features can be incorporated and translated to a chatbot program and pilot tested with the target population. From there, the above-mentioned JITAI approach can be studied to test how different features can be adaptively applied to different individuals over time.

In summary, our study calls for more interdisciplinary work to continue enriching the conceptualization of a chatbot as a relational and persuasive agent and to develop approaches to leverage AI techniques to improve a chatbot's relational and persuasive capacities with strong ethical principles. We call for future research to continue expanding and modifying this framework and to conduct empirical studies to evaluate its applicability in the actual design and assessment of interventions.

Acknowledgments

This project was supported by a grant (K24NR015812) from the National Institute of Nursing Research; a grant (RAP Team Science Award) from the University of California, San Francisco; a grant (Artificial Intelligence in Healthcare Innovation Award) from the University of California, Davis; and a research gift from Intel. The study sponsors had no role in the study design; collection, analysis, or interpretation of data; writing of the report; or decision to submit the report for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of chatbot-based physical activity and diet interventions.
[DOCX File , 30 KB - jmir_v22i9e22845_app1.docx]

References

1. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. The Lancet 2010 Nov 20;376(9754):1775-1784. [doi: 10.1016/S0140-6736(10)61514-0] [Medline: 21074255]
2. Wagner K, Brath H. A global view on the development of non communicable diseases. Prev Med 2012 May;54 Suppl:S38-S41. [doi: 10.1016/j.ypmed.2011.11.012] [Medline: 22178469]
3. Bennett JE, Stevens GA, Mathers CD, Bonita R, Rehm J, Kruk ME, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. The Lancet 2018 Sep;392(10152):1072-1088. [doi: 10.1016/S0140-6736(18)31992-5]
4. Bloom DE, Cafiero ET, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S, et al. The Global Economic Burden of Noncommunicable Diseases. World Economic Forum. 2011 Sep. URL: http://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf [accessed 2020-08-30]
5. Clarke TC, Norris T, Schiller JS. Early Release of Selected Estimates Based on Data From the 2018 National Health Interview Survey. National Center for Health Statistics. 2019. URL: https://www.cdc.gov/nchs/nhis/releases/released201905.htm [accessed 2020-08-28]
6. Hales C, Carroll M, Fryar C, Ogden C. Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017–2018. NCHS Data Brief. 2020. URL: https://www.cdc.gov/nchs/data/databriefs/db360-h.pdf [accessed 2020-08-29]
7. Bertram M, Banatvala N, Kulikov A, Belausteguigoitia I, Sandoval R, Hennis A, et al. Using economic evidence to support policy decisions to fund interventions for non-communicable diseases. BMJ 2019 May 20;365:l1648. [doi: 10.1136/bmj.l1648] [Medline: 31110025]
8. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
9. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
10. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
11. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
12. Zhang J, Brackbill D, Yang S, Centola D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. Prev Med Rep 2015;2:651-657 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.005] [Medline: 26844132]
13. Zhang J, Calabrese C, Ding J, Liu M, Zhang B. Advantages and challenges in using mobile apps for field experiments: A systematic review and a case study. 2017 Sep 05;6(2):179-196. [doi: 10.1177/2050157917725550]
14. IJsselsteijn W, de Kort Y, Midden C, Eggen B, van den Hoven E. Persuasive technology for human well-being: setting the scene. In: IJsselsteijn WA, de Kort YA, Midden C, Eggen B, van den Hoven E, editors. Persuasive Technology. PERSUASIVE 2006. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2006:1-5.
15. Walsh C, Chaudhry B, Dua P, Goodman KW, Kaplan B, Kavuluru R, et al. Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. JAMIA Open 2020 Apr;3(1):9-15 [FREE Full text] [doi: 10.1093/jamiaopen/ooz054] [Medline: 32607482]

16. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
17. Adam M, Wessel M, Benlian A. AI-based chatbots in customer service and their effects on user compliance. Electron Markets 2020 Mar 17:1-19. [doi: 10.1007/s12525-020-00414-7]
18. Androutsopoulou A, Karacapilidis N, Loukis E, Charalabidis Y. Transforming the communication between citizens and government through AI-guided chatbots. Government Information Quarterly 2019 Apr;36(2):358-367. [doi: 10.1016/j.giq.2018.10.001]
19. Georgescu A. Chatbots for education - trends, benefits and challenges. 2018 Presented at: eLearning & Software for Education; April 20, 2018; Bucharest p. 195-200. [doi: 10.12753/2066-026x-18-097]
20. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
21. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020 May 4;3(1):65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
22. Conroy DE, Hojjatinia S, Lagoa CM, Yang C, Lanza ST, Smyth JM. Personalized models of physical activity responses to text message micro-interventions: A proof-of-concept application of control systems engineering methods. Psychol Sport Exerc 2019 Mar;41:172-180 [FREE Full text] [doi: 10.1016/j.psychsport.2018.06.011] [Medline: 30853855]
23. Fukuoka Y, Lindgren TG, Mintz YD, Hooper J, Aswani A. Applying Natural Language Processing to Understand Motivational Profiles for Maintaining Physical Activity After a Mobile App and Accelerometer-Based Intervention: The mPED Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Jun 20;6(6):e10042 [FREE Full text] [doi: 10.2196/10042] [Medline: 29925491]
24. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
25. Gabrielli S, Marie K, Corte C. SLOWBot (Chatbot) lifestyle assistant. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth '18; May 24, 2018; New York, United States p. 367-370. [doi: 10.1145/3240925.3240953]
26. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection Companion. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Jul 05;2(2):1-26. [doi: 10.1145/3214273]
27. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. In: Maedche A, vom Brocke J, Hevner A, editors. Designing the Digital Transformation. DESRIST 2017. Lecture Notes in Computer Science. Cham, Switzerland: Springer; 2017.
28. Stein N, Brooks K. A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
29. Sillice MA, Morokoff PJ, Ferszt G, Bickmore T, Bock BC, Lantini R, et al. Using Relational Agents to Promote Exercise and Sun Protection: Assessment of Participants' Experiences With Two Interventions. J Med Internet Res 2018 Feb 07;20(2):e48 [FREE Full text] [doi: 10.2196/jmir.7640] [Medline: 29415873]
30. Pereira J, Díaz Ó. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
31. Brixey J, Hoegen R, Lan W, Rusow J, Singla K, Yin X, et al. SHIHbot: A Facebook chatbot for sexual health information on HIV/AIDS. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. 2017 Presented at: The 18th Annual SIGdial Meeting on Discourse and Dialogue; 2017; Saarbrücken, Germany p. 370-373. [doi: 10.18653/v1/w17-5544]
32. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
33. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]
34. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ Couns 2013 Aug;92(2):160-166 [FREE Full text] [doi: 10.1016/j.pec.2013.05.007] [Medline: 23809167]
35. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. Front. Robot. AI 2017 Oct 12;4:1-9. [doi: 10.3389/frobt.2017.00051]
36. Philip P, Micoulaud-Franchi JA, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7(12):42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]

37. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of NAACL-HLT 2019. 2019 Presented at: The 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019); June 7, 2019; Minneapolis, Minnesota.
38. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI Blog. 2019. URL: https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf [accessed 2020-08-29]
39. Wu Q, Zhang Y, Li Y, Yu Z. Alternating roles dialog model with large-scale pre-trained language models. arXiv. 2019. URL: https://arxiv.org/abs/1910.03756 [accessed 2020-08-29]
40. Gu J, Wu Q, Wu C, Shi W, Yu Z. A tailored pre-training model for task-oriented dialog generation. arXiv. 2020. URL: https://arxiv.org/pdf/2004.13835.pdf [accessed 2020-08-28]
41. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Protocol of a Microrandomized Trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
42. Kramer J, Künzler F, Mishra V, Smith S, Kotz D, Scholz U, et al. Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
43. Piao M, Ryu H, Lee H, Kim J. Use of the Healthy Lifestyle Coaching Chatbot App to Promote Stair-Climbing Habits Among Office Workers: Exploratory Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 May 19;8(5):e15085 [FREE Full text] [doi: 10.2196/15085] [Medline: 32427114]
44. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
45. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
46. Fadhil A, Wang Y, Reiterer H. Assistive Conversational Agent for Health Coaching: A Validation Study. Methods Inf Med 2019 Jun;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
47. Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers Inf Technol Electronic Eng 2018 Jan 8;19(1):10-26. [doi: 10.1631/fitee.1700826]
48. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco, CA: Morgan Kaufmann; 2002.
49. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A Novel Diabetes Prevention Intervention Using a Mobile App: A Randomized Controlled Trial With Overweight Adults at Risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
50. Fukuoka Y, Haskell W, Lin F, Vittinghoff E. Short- and Long-term Effects of a Mobile Phone App in Conjunction With Brief In-Person Counseling on Physical Activity Among Physically Inactive Women: The mPED Randomized Clinical Trial. JAMA Netw Open 2019 May 03;2(5):e194281 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.4281] [Medline: 31125101]
51. Fukuoka Y, Vittinghoff E, Hooper J. A weight loss intervention using a commercial mobile application in Latino Americans-Adelgaza Trial. Transl Behav Med 2018 Sep 08;8(5):714-723 [FREE Full text] [doi: 10.1093/tbm/ibx039] [Medline: 29474702]
52. Shi W, Wang X, Oh Y, Zhang J, Sahay S, Yu Z. Effects of persuasive dialogues: testing bot identities and inquiry strategies. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: The 2020 CHI Conference on Human Factors in Computing Systems; April 30, 2020; Oahu, Hawaii p. 1-13. [doi: 10.1145/3313831.3376843]
53. Wang X, Shi W, Kim R, Oh Y, Yang S, Zhang J. Persuasion for good: towards a personalized persuasive dialogue system for social good. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: Annual Meeting of the Association for Computational Linguistics; August 2, 2019; Florence, Italy p. 5635-5649. [doi: 10.18653/v1/p19-1566]
54. Zhang J, Jemmott Iii JB. Mobile App-Based Small-Group Physical Activity Intervention for Young African American Women: a Pilot Randomized Controlled Trial. Prev Sci 2019 Aug;20(6):863-872. [doi: 10.1007/s11121-019-01006-4] [Medline: 30788692]
55. Oviatt S, Darves C, Coulston R. Toward adaptive conversational interfaces. ACM Trans Comput Hum Interact 2004 Sep;11(3):300-328. [doi: 10.1145/1017494.1017498]
56. Phillips E, Zhao X, Ullman D, Malle B. What is human-like?: decomposing robots' human-like appearance using the Anthropomorphic roBOT (ABOT) database. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. 2018 Presented at: ACM/IEEE International Conference on Human-Robot Interaction; March 8, 2018; Chicago, Illinois p. 105-113. [doi: 10.1145/3171221.3171268]

57. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge, UK: Cambridge University Press; 1996.
58. Mori M, MacDorman K, Kageki N. The Uncanny Valley [From the Field]. IEEE Robot. Automat. Mag 2012 Jun;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]
59. Kätsyri J, Förger K, Mäkäräinen M, Takala T. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Front Psychol 2015 Dec;6(12):390 [FREE Full text] [doi: 10.3389/fpsyg.2015.00390] [Medline: 25914661]
60. Ho A, Hancock J, Miner A. Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. J Commun 2018 Aug;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
61. Lee Y, Yamashita N, Huang Y, Fu W. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: The 2020 CHI Conference on Human Factors in Computing Systems; April 30, 2020; Oahu, Hawaii p. 1-12. [doi: 10.1145/3313831.3376175]
62. Vandelanotte C, De Bourdeaudhuij I, Brug J. Two-year follow-up of sequential and simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. ann. behav. med 2007 Jun 21;33(2):213-219. [doi: 10.1007/Bf02879903] [Medline: 5637250]
63. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010 Sep;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
64. Kaptein M, Markopoulos P, de Ruyter B, Aarts E. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. International Journal of Human-Computer Studies 2015 May;77:38-51. [doi: 10.1016/j.ijhcs.2015.01.004]
65. Sakai R, Van Peteghem S, van de Sande L, Banach P, Kaptein M. Personalized Persuasion in Ambient Intelligence: The APStairs System. In: Keyson DV, editor. Ambient Intelligence. AmI 2011. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2011:205-209.
66. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]
67. Bull FC, Kreuter MW, Scharff DP. Effects of tailored, personalized and general health messages on physical activity. Patient Education and Counseling 1999 Feb;36(2):181-192. [doi: 10.1016/s0738-3991(98)00134-7] [Medline: 10223022]
68. Lukin S, Anand P, Walker M, Whittaker S. Argument strength is in the eye of the beholder: audience effects in persuasion. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017 Presented at: The 15th Conference of the European Chapter of the Association for Computational Linguistics; 2017; Valencia, Spain. [doi: 10.18653/v1/e17-1070]
69. Velicer W, Redding CA, Blissmer B, Babbin SF, Paiva A, Bickmore T, et al. Using relational agents in tailored interventions for multiple risk factors: preliminary 12 month results. Ann Behav Med 2013 Mar;45:S180.
70. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. 'It's just like you talk to a friend' relational agents for older adults. Interacting with Computers 2005 Dec;17(6):711-735. [doi: 10.1016/j.intcom.2005.09.002]
71. Bickmore T, Schulman D, Yin L, Ring L. Reaching minority populations with relational agents. Ann Behav Med 2013;45:S60.
72. Bickmore T, Schulman D, Pfeifer L, Yin L. Relational agents for promoting intervention engagement with underserved populations. Ann Behav Med 2011;41(Symposium 30A):3097.
73. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. J Med Internet Res 2020 Mar 06;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
74. Purington A, Taft J, Sannon S, Bazarova N, Taylor S. "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017 Presented at: The CHI Conference Extended Abstracts on Human Factors in Computing Systems; 2017; Denver, Colorado p. 2853-2859. [doi: 10.1145/3027063.3053246]
75. Mou Y, Xu K. The media inequality: Comparing the initial human-human and human-AI social interactions. Computers in Human Behavior 2017 Jul;72:432-440. [doi: 10.1016/j.chb.2017.02.067]
76. Hancock J, Naaman M, Levy K. AI-mediated communication: definition, research agenda, and ethical considerations. J Comput Mediat Commun 2020;25(1):89-100. [doi: 10.1093/jcmc/zmz022]
77. Feng B. Testing an Integrated Model of Advice Giving in supportive interactions. Hum Commun Res 2009;35(1):115-129. [doi: 10.1111/j.1468-2958.2008.01340.x]
78. MacGeorge EL, Guntzviller LM, Hanasono LK, Feng B. Testing Advice Response Theory in Interactions With Friends. Communication Research 2013 Nov 12;43(2):211-231. [doi: 10.1177/0093650213510938]
79. Gallois C, Gasiorek J, Giles H, Soliz J. Communication Accommodation Theory: Integrations and New Framework Developments. In: Giles H, editor. Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts. Cambridge, UK: Cambridge University Press; 2016:192-210.
80. Pretorius M. Communication accommodation theory analysis of nurse–patient interaction: Implications for course design. Int J Appl Linguist 2017 Aug 24;28(1):71-85. [doi: 10.1111/ijal.12184]
81. Miller W, Rollnick S. Motivational interviewing: helping people change. New York, NY: Guilford Press; 2012.

82. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Mar 21;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
83. Fishbein M, Ajzen I. Predicting and changing behavior: The reasoned action approach. New York, NY: Psychology Press; 2011.
84. Dillard J, Pfau M. The persuasion handbook: Developments in theory and practice. Thousand Oaks, CA: Sage Publications; 2002.
85. Cappella J. Integrating message effects and behavior change theories: organizing comments and unanswered questions. J Commun 2006;56:S265-SS79. [doi: 10.1111/j.1460-2466.2006.00293.x]
86. Williams P, Block LG, Fitzsimons GJ. Simply asking questions about health behaviors increases both healthy and unhealthy behaviors. Social Influence 2006 Jun;1(2):117-127. [doi: 10.1080/15534510600630850]
87. Ellis K, Godbole S, Marshall S, Lanckriet G, Staudenmayer J, Kerr J. Identifying Active Travel Behaviors in Challenging Environments Using GPS, Accelerometers, and Machine Learning Algorithms. Front Public Health 2014 Dec;2(12):36 [FREE Full text] [doi: 10.3389/fpubh.2014.00036] [Medline: 24795875]
88. Zhou M, Fukuoka Y, Goldberg K, Vittinghoff E, Aswani A. Applying machine learning to predict future adherence to physical activity programs. BMC Med Inform Decis Mak 2019 Aug 22;19(1):169 [FREE Full text] [doi: 10.1186/s12911-019-0890-0] [Medline: 31438926]
89. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review. J Med Internet Res 2020 Jun 05;22(6):e18301 [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
90. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
91. Deng T, Kanthawala S, Meng J, Peng W, Kononova A, Hao Q, et al. Measuring smartphone usage and task switching with log tracking and self-reports. 2018 Apr 25;7(1):3-23. [doi: 10.1177/2050157918761491]
92. Liu B, Sundar SS. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
93. Lopatovska I, Williams H. Personification of the Amazon Alexa: BFF or a mindless companion. 2018 Presented at: The 2018 Conference on Human Information Interaction & Retrieval; March 15, 2018; New Brunswick, New Jersey p. 265-268. [doi: 10.1145/3176349.3176868]
94. Bickmore T, Cassell J. Social dialogue with embodied conversational agents. In: van Kuppevelt JC, Dybkjær L, Bernsen NO, editors. Advances in Natural Multimodal Dialogue Systems. Dordrecht, Netherlands: Springer; 2005:23-54.
95. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
96. Kim DJ, Holowaty EJ. Brief, validated survey instruments for the measurement of fruit and vegetable intakes in adults: a review. Preventive Medicine 2003 Apr;36(4):440-447. [doi: 10.1016/s0091-7435(02)00040-3]
97. Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY. Objective measurement of physical activity and sedentary behaviour: review with new data. Arch Dis Child 2008 Jul;93(7):614-619. [doi: 10.1136/adc.2007.133272] [Medline: 18305072]
98. Torresen J. A Review of Future and Ethical Perspectives of Robotics and AI. Front. Robot. AI 2018 Jan 15;4:1-10. [doi: 10.3389/frobt.2017.00075]
99. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds Mach (Dordr) 2018 Nov 26;28(4):689-707 [FREE Full text] [doi: 10.1007/s11023-018-9482-5] [Medline: 30930541]
100. Kizilcec R. How much information? Effects of transparency on trust in an algorithmic interface. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: The 2016 CHI Conference on Human Factors in Computing Systems; May 12, 2016; San Jose, CA p. 2390-2395. [doi: 10.1145/2858036.2858402]
101. Na L, Yang C, Lo C, Zhao F, Fukuoka Y, Aswani A. Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. JAMA Netw Open 2018 Dec 07;1(8):e186040 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.6040] [Medline: 30646312]
102. McCoy TH, Hughes MC. Preserving Patient Confidentiality as Data Grow: Implications of the Ability to Reidentify Physical Activity Data. JAMA Netw Open 2018 Dec 07;1(8):e186029 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.6029] [Medline: 30646303]
103. Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy. FNT in Theoretical Computer Science 2014;9(3-4):211-407. [doi: 10.1561/0400000042]
104. Engelhardt H. The foundations of bioethics. New York, NY: Oxford University Press; 1996.
105. Smids J. The voluntariness of persuasive technology. In: Bang M, Ragnemalm EL, editors. Persuasive Technology. Design for Health and Safety. PERSUASIVE 2012. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2012:123-132.
106. Liu H, Dacon J, Fan W, Liu H, Liu Z, Tang J. Does gender matter? Towards fairness in dialogue systems. arXiv. 2020. URL: https://arxiv.org/abs/1910.10486 [accessed 2020-08-29]

107. Henderson P, Sinha K, Angelard-Gontier N, Ke N, Fried G, Lowe R, et al. Ethical challenges in data-driven dialogue systems. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. 2018 Presented at: The 2018 AAAI/ACM Conference on AI, Ethics, and Society; February 3, 2018; New Orleans, Louisiana p. 123-129. [doi: 10.1145/3278721.3278777]
108. Kearns M, Roth A. The ethical algorithm: the science of socially aware algorithm design. New York, NY: Oxford University Press; 2019.
109. Sjöberg C, Timpka T. Participatory design of information systems in health care. J Am Med Inform Assoc 1998 Mar 01;5(2):177-183 [FREE Full text] [doi: 10.1136/jamia.1998.0050177] [Medline: 9524350]
110. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: Using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
CASA: computers are social actors
GPT2: Generative Pre-Training-2
JITAI: just-in-time adaptive intervention
NCD: noncommunicable disease
NLP: natural language processing
RCT: randomized controlled trial
UVE: uncanny valley effect

Edited by G Eysenbach; submitted 27.07.20; peer-reviewed by J Xue, Y Liao, E Sezgin, K Schneider; comments to author 19.08.20; revised version received 03.09.20; accepted 17.09.20; published 30.09.20.

Please cite as:
Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y
Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint
J Med Internet Res 2020;22(9):e22845
URL: https://www.jmir.org/2020/9/e22845
doi:10.2196/22845
PMID:32996892

Original Paper

Weight Reduction Through a Digital Nutrition and Food Purchasing Platform Among Users With Obesity: Longitudinal Study

Emily A Hu[1,2], PhD, MHS; Viet Nguyen[1], BA; Jason Langheier[1], MD, MPH; Dexter Shurney[1,3], MD, MPH

[1]Zipongo, Inc, DBA Foodsmart, San Francisco, CA, United States

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[3]American College of Lifestyle Medicine, Chesterfield, MO, United States

Corresponding Author:
Emily A Hu, PhD, MHS
Zipongo, Inc, DBA Foodsmart
600 California Street, 11th Floor
San Francisco, CA, 94108
United States
Phone: 1 415 604 4122
Email: emily.hu@zipongo.com

Abstract

Background: Digital nutrition apps that monitor or provide recommendations on diet have been found to be effective in behavior change and weight reduction among individuals with obesity. However, there is less evidence on how integration of personalized nutrition recommendations and changing the food purchasing environment through online meal planning and grocery delivery, meal kits, and grocery incentives impacts weight loss among individuals with obesity.

Objective: The objective of this observational longitudinal study was to examine weight loss and predictors of weight loss among individuals with obesity who are users of a digital nutrition platform that integrates tools to provide nutrition recommendations and changes in the food purchasing environment grounded in behavioral theory.

Methods: We included 8977 adults with obesity who used the digital Foodsmart platform, created by Zipongo, Inc, DBA Foodsmart between January 2013 and April 2020. We retrospectively analyzed user characteristics and their associations with weight loss. Participants reported age, gender, height, at least 2 measures of weight, and usual dietary intake. Healthy Diet Score, a score to measure overall diet quality, was calculated based on responses to a food frequency questionnaire. We used paired *t* tests to compare differences in baseline and final weights and baseline and final Healthy Diet Scores. We used univariate and multivariate logistic regression models to estimate odds ratios and 95% CI of achieving 5% weight loss by gender, age, baseline BMI, Healthy Diet Score, change in Healthy Diet Score, and duration of enrollment. We conducted stratified analyses to examine mean percent weight change by enrollment duration and gender, age, baseline BMI, and change in Healthy Diet Score.

Results: Over a median (IQR) of 9.9 (0.03-54.7) months of enrollment, 59% of participants lost weight. Of the participants who used the Foodsmart platform for at least 24 months, 33.3% achieved 5% weight loss. In the fully adjusted logistic regression model, we found that baseline BMI (OR 1.02, 95% CI 1.02-1.03; *P*<.001), baseline Healthy Diet Score (OR 1.06, 95% CI 1.05-1.08; *P*<.001), greater change in Healthy Diet Score (OR 1.12, 95% CI 1.11-1.14; *P*<.001), and enrollment length (OR 1.28, 95% CI 1.23-1.32; *P*<.001) were all significantly associated with higher odds of achieving at least 5% weight loss.

Conclusions: This study found that a digital app that provides personalized nutrition recommendations and change in one's food purchasing environment appears to be successful in meaningfully reducing weight among individuals with obesity.

(J Med Internet Res 2020;22(9):e19634) doi:10.2196/19634

KEYWORDS

digital; nutrition; meal planning; weight loss; obese; food environment; food ordering; food purchasing; behavioral economics; behavior change; eating behavior; mHealth; app

Introduction

The increasing prevalence of obesity worldwide is a critical public health problem [1,2]. In the United States, about 39.6% of adults 20 and older were considered obese in the years 2015-2016, and the prevalence is projected to increase [3]. Overweight and obesity pose serious health challenges as they are strong risk factors for cardiovascular disease, type 2 diabetes, chronic kidney disease, many cancers, and mortality [1,4,5].

The prevention and management of obesity are extremely necessary given the potential health and cost consequences [6]. For decades, there has been mounting evidence from large trials such as the Diabetes Prevention Program (DPP) showing that change in lifestyle, often related to weight reduction, can have dramatic effects on health and chronic disease [7-12]. However, interventions like DPP have failed to sustain weight loss more than 18-24 months and can be costly due to coaching time and the cycles of losing and regaining weight [13,14].

Digital health technologies that incorporate nutrition education and monitoring have gained increasing popularity to change and manage dietary choices [15-17]. Previous studies on mobile apps to improve nutrition are promising as their results indicate that digital nutrition interventions may be effective in changing dietary behavior to improve weight, glucose, and blood pressure among healthy individuals and people at risk of or with chronic disease [18-21]. While many of these apps provide general diet recommendations, few apps have a decision engine capable of providing personalized dietary advice, meal planning assistance, and online grocery delivery to users [16].

Meal planning and at-home cooking have been found to be associated with greater adherence to dietary guidelines, increased fruit and vegetable intake, and greater variety of foods consumed [22,23]. Meal planning behaviors, including frequency of planning meals ahead of time, grocery shopping and cooking, have been associated with lower likelihood of obesity in men and women [23].

To our knowledge, no studies with meaningful scale have examined the effect of a digital technology that provides personalized healthy meal plans and changes in the food purchasing environment (through online grocery shopping, purchase discounts, delivery, and meal kits) on health outcomes. There is a need for additional evidence on how digital technologies that alter behavioral economics, such as food purchasing, might play a role in improving diet and driving more cost-effective health outcomes for individuals with obesity.

Our goal was to conduct an observational longitudinal study leveraging existing data from a digital nutrition platform to investigate the effectiveness of a personalized nutrition, meal planning, and food purchasing program on weight loss among individuals with obesity.

Methods

Study Population

The current study is a longitudinal analysis of 8977 adults with obesity (aged 18 to 80 years, living in the United States) who enrolled in the Foodsmart platform. Of the 888,999 users who had enrolled up to April 2020, we excluded individuals who did not report weight (n=562,276), those who reported extreme values for height (<54 in or >78 in, ie, <1.37 m or >1.98 m) or weight (<60 lb or >400 lb, ie, <27.2 kg or >181.4 kg) (n=25,946), and those who were not obese (BMI<30 kg/m^2) at the time of enrollment (n=200,308). We further excluded those who reported BMI after more than 3 days from joining Foodsmart, those whose BMI changed more than 15 kg/m^2 in less than 10 months, and participants with greater than 16% weight change in less than 1 month (n=13,548). We additionally excluded those who did not report weight at least 2 times (n=9023) and participants who did not fill out the Nutriquiz survey twice (n=68,921).

Foodsmart Platform

Foodsmart (Zipongo Inc DBA Foodsmart) built a digital nutrition platform called Foodsmart that is designed to make healthier dietary choices simple and sustainable through personalization of nutrition and meal/recipe recommendations by creating a food purchasing environment that provides healthy options for all people, whether they enjoy cooking, prefer to use meal kits, or prefer prepared meals. Foodsmart is made up of 2 components (FoodSmart and FoodsMart) with self-directed tools that drive knowledge, motivation, and planning to make it easier and more affordable to prepare tasty, healthy food at home (Figure 1).

Figure 1. Components and tools of Foodsmart.

The platform was developed using Prochaska's Theory of Change model as the baseline theory supplemented with elements from the Behaviour Change Techniques Taxonomy [24,25] (Table 1). The tools from both FoodSmart and FoodsMart are designed to target all stages (pre-contemplation, contemplation, preparation, action, and maintenance) of behavior change in healthier eating. These tools encourage users to reflect and assess their dietary habits with Nutriquiz, helping create a specific plan for users to eat healthier daily, offering tools to purchase healthy foods, and providing incentives and communication to maintain healthy behaviors.

Table 1. Foodsmart platform components and tools linked with behavior change stages and techniques.

Foodsmart components and tools	Stages of change [24]	Behaviour Change Techniques [25]
FoodSmart		
Nutriquiz dietary assessment and re-assessment and dietary recommendations	• Pre-contemplation: Encourages user to think about dietary habits • Contemplation: Results encourage users to think of changes to make in diet • Preparation: Helps create a specific plan on which foods to change • Maintenance: Monitors progress by re-taking Nutriquiz	• Provide information about behavioral health link • Prompt intention formation • Prompt specific goal setting • Prompt self-monitoring of behavior • Prompt self-monitoring of performance • Provide feedback on performance • Provide opportunities for social comparison
Family meal planning (recipe recommendations for each meal through linkage to recipe database)	• Preparation: Assists the user in making a plan to cook • Action: Automatically loads recipe ingredients to grocery list	• Prompt barrier identification • Set graded tasks • Provide instruction • Stress management • Time management
Social liking and commenting of recipes	• Preparation: Prepares the user to cook by browsing and interacting with recipes; also builds social support to be successful	• Plan social support or social change
Enrollment and activation marketing (incentives, enrollment emails, newsletters)	• Pre-contemplation: Enrollment emails and newsletters create awareness of capabilities • Contemplation: Emails encourage people to activate certain features based on needs; incentives provide contingent awards for participating • Maintenance: Newsletters and emails to encourage people to keep using platform	• Provide general encouragement • Provide contingent awards
FoodsMart (advertising of unhealthy foods is filtered out)		
Online grocery list and food ordering (including prepared meals and meal kits)	• Preparation: Online grocery list helps identify barriers • Action: Online food ordering helps stress and time management • Maintenance: Once a user practices and demonstrates the behavior of creating a list online, more likely to maintain online food ordering	• Default behavioral economics • Prompt barrier identification • Prompt practice • Set graded tasks • Provide instruction • Model or demonstrate the behavior • Stress management • Time management
Food discounts and incentives	• Contemplation: Incentives provide contingent awards for participating • Preparation: Discounts allow for budgeting before grocery shopping • Action: Makes it feasible to buy healthy food that otherwise can't afford • Maintenance: Discounts and incentives encourage continual usage by helping with stress and time management	• Behavioral economics • Stress management • Time management

The first component is FoodSmart, which contains the in-app Nutriquiz, a dietary assessment (based on the National Cancer Institute's Diet History Questionnaire). Users can take Nutriquiz to report their dietary habits, which provides immediate and specific feedback on aspects of their diet to improve on as well as personalized meal and recipe planning based on the Nutriquiz results. Over time, users can retake the Nutriquiz assessment to monitor their own progress related to specific nutrients and food groups as well as their progress on health goals, like weight. The second component is FoodsMart, which helps reset one's default behavioral economics by altering the food purchasing environment. This is achieved through personalized

meal plan conversion to a grocery list and integrated online ordering and delivery of groceries, meal kits, and prepared foods, where food advertising paid for by food manufacturers is removed and replaced with nudges to make healthier substitutions that align with user preferences and their personalized meal planning. Customized grocery discounts on healthier options help the user save money and further nudge the user to make healthier choices. The Foodsmart platform has been in use since 2013; and 90% of users enrolled in 2017 or later, after most of the major content and design changes to the platform were made (Multimedia Appendix 1). The platform has evolved over time, with the most significant change being the addition of grocery and food ordering in the last few years. The product is available through certain health plans and employers who have signed up for Foodsmart, and they can provide this product as an option or benefit for their members/employees to enroll in. It is available to be used on the web, iOS, and Android operating systems.

Measurements

All data were self-reported through the Foodsmart app during the study period. When users created their account, they were prompted to fill out a survey created by Foodsmart called Nutriquiz, a 53-item food frequency questionnaire adapted from the National Cancer Institute Diet History Questionnaire, which has been previously validated [26]. The questionnaire ascertains biological sex, birth date, weight, and usual intake of food groups and nutrients. For example, it asks, "How often do you eat fruit?" Possible responses include "never," "monthly," "weekly," and "daily." Other food groups assessed included vegetables, whole grains, proteins, carbohydrates, fats, fiber, sodium, and water. Foodsmart's research team created a healthy diet score called Healthy Diet Score, which is based on the Alternative Healthy Eating Index-2010 (AHEI-2010) and the Commonwealth Scientific and Industrial Research Organization (CSIRO) Healthy Diet Score [27,28]. Similar to the AHEI-2010, the Healthy Diet Score includes fruits, vegetables, and sodium components; and each component is scored 1-10 using absolute cutoffs. In order to keep the score concise, the Healthy Diet Score combined macronutrient components in a similar fashion to the CSIRO Healthy Diet Score, which includes only 1 category each for protein, carbohydrates, and fats. Additionally, it has a component for fluids, which was modified to be hydration in the Healthy Diet Score since percent fluid intake is more relevant than total quantity of fluids. For calculation of the Healthy Diet Score, participants were assigned a score from 0-10 for 7 components: fruit, vegetable, protein ratio (white meat/vegetarian protein to red/processed meat), carbohydrate ratio (total fiber to total carbohydrate), fat ratio (polyunsaturated to saturated/trans fats), sodium, and hydration (percent of daily fluid goal). Higher scores indicated healthier habits. A total Healthy Diet Score to evaluate overall diet quality was calculated by summing the scores of the 7 components, with the total possible score ranging from 0 to 70. Change in Healthy Diet Score was calculated as the difference between the first Healthy Diet Score and the last Healthy Diet Score. We compared participants whose Healthy Diet Score decreased or was stable (no improvement in diet quality) with those participants whose Healthy Diet Score increased (improvement in diet quality) between the first and last report. We collapsed decreased and stable categories due to a low number of participants in the stable category.

Participants were asked to add weight and height data when they joined and could update their weight at any time during usage of the platform. Baseline BMI was calculated as first weight entry in kilograms divided by height in square meters (kg/m^2). We categorized participants by baseline obesity class. Class 1 obesity was defined as a BMI between 30 to 34.9 kg/m^2; class 2 was defined as a BMI of 35 to 39.9 kg/m^2; and class 3 was defined as a BMI of 40 kg/m^2 or higher. To calculate a change in weight, we subtracted the last reported weight from the first reported weight. Our primary outcome was 5% or greater weight loss, which has been found to be clinically significant and associated with improvements in cardiometabolic risk factors such as lipid profile and insulin sensitivity [9,12,29,30].

Duration of enrollment (in months) in Foodsmart was calculated as follows: the number of days between the date on which participants initially entered their weight and the date on which they entered their last follow-up weight was calculated and divided by a factor of 30.437 to convert to months. We classified participants into enrollment categories of 0 to 6 months, greater than 6 to 12 months, greater than 12 to 18 months, greater than 18 to 24 months, and 24 months or greater. For stratified analyses, we collapsed the greater than 18 to 24 months and 24 months or greater into one category of greater than 18 months.

Statistical Analysis

Descriptive statistics were used to examine baseline characteristics of the total study population and to compare whether participants lost at least 5% of their initial body weight. Categorical variables were reported as frequencies (%) and continuous variables were reported as mean (SD). Chi-square tests and analysis of variance (ANOVA) tests were used to test differences for categorical and continuous variables, respectively, between participants who achieved 5% weight loss and participants who did not.

We examined the change in weight and Healthy Diet Score by using paired *t* tests between baseline and final weights and Healthy Diet Scores of participants. We then used univariate logistic regression models to estimate odds ratios and 95% CI between achievement of 5% weight loss and independent variables: gender, age, baseline BMI, baseline Healthy Diet Score (per 2-point increase), change in Healthy Diet Score (per 2-point increase), and length of enrollment (per 6 months). Multivariate logistic regression models were adjusted for variables that were statistically significant to investigate independent associations with achievement of 5% weight loss.

Further, we conducted stratified analyses to examine differences in percent weight change by enrollment length and stratified by gender, age category, BMI class, and change in Healthy Diet Score. We used bar graphs to visualize differences and ANOVA tests to statistically test for differences, using a Bonferroni-corrected *P* value of .0031 to account for multiple comparisons.

We considered a *P* value smaller than .05 to be significant for all tests except for the ANOVA tests used for detecting differences in stratified groups. R studio version 1.2.5033 and Stata version 16 (StataCorp) were used for all analyses.

The study was declared exempt from institutional review board oversight by the Pearl Institutional Review Board given the retrospective design of the study and less than minimal risk to participants.

Results

Participant Characteristics

Baseline characteristics of the total study sample and stratified by whether participants achieved 5% weight loss are shown in Table 2. Categorical variables were reported as frequencies (%) and continuous variables were reported as mean (SD).

Table 2. Baseline characteristics of Foodsmart users.

	Total (N=8977)	Did not lose ≥5% of initial weight (n=6838)	Lost ≥5% of initial weight (n=2139)	*P* value[a]
Male, %	20.1	20.1	20.2	.9
Age, years, mean (SD)	46.6 (11.0)	46.3 (11.0)	47.3 (11.0)	.01
Height, m, mean (SD)	1.7 (0.1)	1.7 (0.1)	1.7 (0.1)	.1
Baseline weight, kg, mean (SD)	101.7 (18.3)	101.2 (18.1)	103.4 (18.9)	<.001
Baseline BMI, kg/m^2, mean (SD)	36.3 (5.6)	36.2 (5.6)	36.8 (5.8)	<.001
Obesity category				**<.001**
Obesity class 1 (30-34.9 kg/m^2), %	52.3	53.4	48.7	
Obesity class 2 (35-39.9 kg/m^2), %	26.7	26.6	27.0	
Obesity class 3 (≥40 kg/m^2), %	21.1	20.1	24.4	
Baseline Healthy Diet Score (0-70), mean (SD)	30.3 (8.6)	30.2 (8.6)	30.6 (8.6)	.1
Final Healthy Diet Score (0-70), mean (SD)	32.6 (8.5)	31.9 (8.5)	34.8 (8.3)	<.001
Change in Healthy Diet Score, mean (SD)	2.3 (7.5)	1.7 (7.3)	4.2 (7.9)	<.001
Enrollment length, months, mean (SD)	11.4 (8.3)	10.7 (8.1)	13.6 (8.3)	<.001
Weight change, %, mean (SD)	–1.5 (7.5)	1.5 (4.9)	–11.1 (6.4)	<.001
Weight change, kg, mean (SD)	–1.7 (7.9)	1.4 (5.0)	–11.6 (7.7)	<.001

[a]Chi-square tests and analysis of variances tests were used to test differences for categorical and continuous variables.

Using *t* tests, we found that final BMI was significantly lower than the baseline BMI (*P*<.001); and final Healthy Diet Score was significantly higher than the baseline Healthy Diet Score (*P*<.001). In total, 59% of participants reported losing weight, and 24% reported loss of at least 5% of their initial weight. Compared to participants who did not lose at least 5% of their initial weight, those who did were more likely to be slightly older, have a slightly higher baseline weight, be classified as obesity class 3, have a higher change in Healthy Diet Score, and be enrolled in the Foodsmart program longer (Table 2). Their baseline Healthy Diet Scores were comparable.

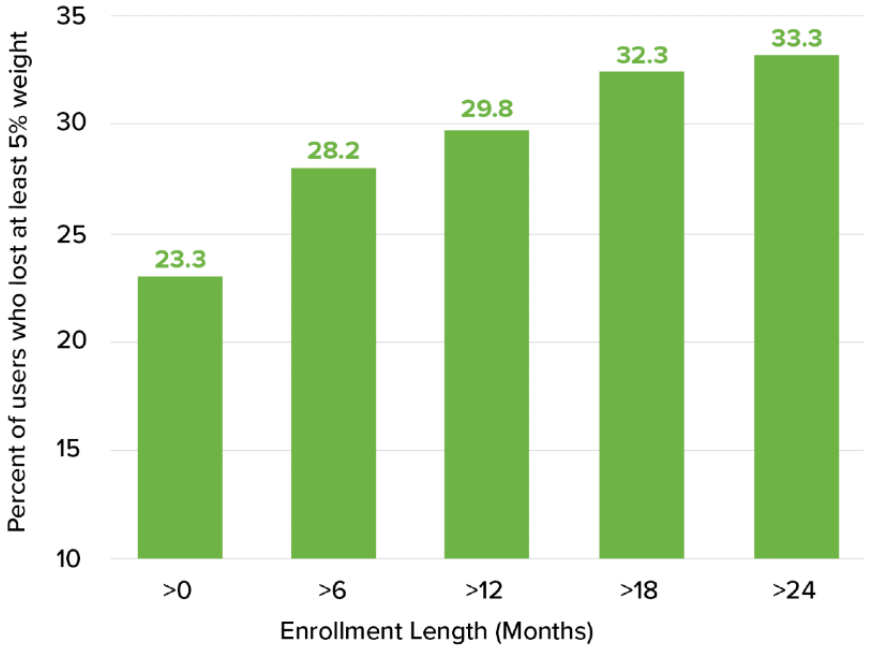
Predictors of At Least 5% Weight Loss

The percentage of users who lost at least 5% of their initial weight increased with longer enrollment duration (Figure 2).

Individual variables contributing to at least 5% weight loss were assessed using univariate logistic regression models (Table 3).

We collectively analyzed the variables and 5% weight loss in a multivariate logistic regression model. Baseline BMI, baseline Healthy Diet Score, greater change in Healthy Diet Score, and enrollment length were all directly associated with higher odds of achieving at least 5% weight loss.

Figure 2. Percent of users who lost at least 5% of initial weight by cumulative enrollment duration.



No. participants who lost 5%	2139	1749	1057	588	222
No. with follow-up reporting	8977	6212	3543	1789	667

Table 3. Factors contributing to at least 5% weight loss in univariate and multivariate logistic regression models.

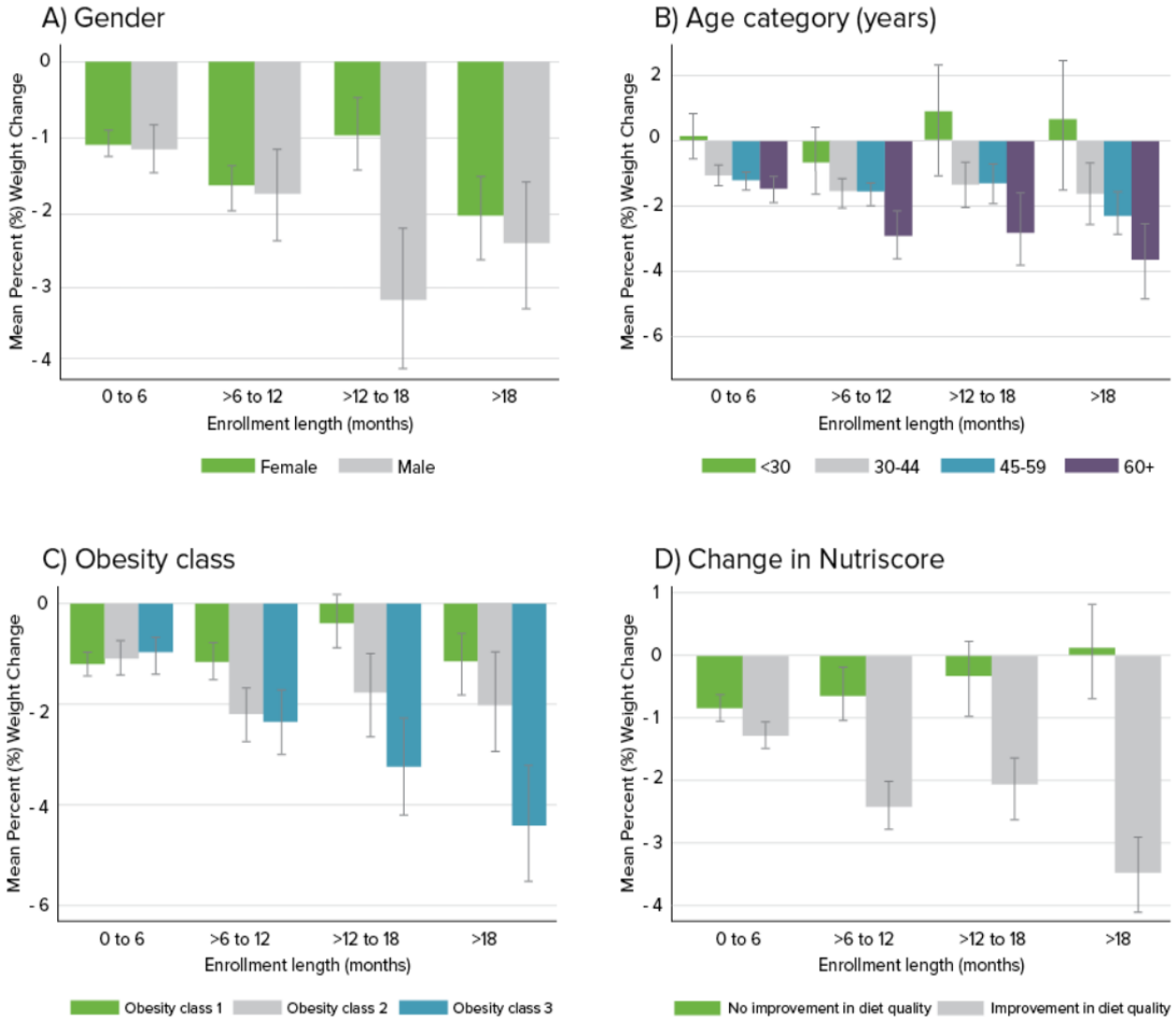
	Univariate OR (95% CI)	*P* value	Multivariate OR (95% CI)	*P* value
Gender (male)	1.01 (0.89-1.14)	.9	1.03 (0.91-1.17)	.7
Age, years	1.01 (1.00-1.01)	<.001	1.00 (1.00-1.01)	.4
Baseline BMI, kg/m^2	1.02 (1.01-1.03)	<.001	1.02 (1.02-1.03)	<.001
Baseline Healthy Diet Score, per 2-point increase	1.01 (1.00-1.02)	<.001	1.06 (1.05-1.08)	<.001
Change in Healthy Diet Score, per 2-point increase	1.09 (1.08-1.11)	<.001	1.12 (1.11-1.14)	<.001
Enrollment length, per 6 months	1.28 (1.24-1.32)	<.001	1.28 (1.23-1.32)	<.001

Stratified Analyses

Figure 3A-D shows the mean percent weight change stratified by enrollment length category (less than 6 months, greater than 6 months to 12 months, greater than 12 months to 18 months, greater than 18 months) and by gender, age category, BMI category, and change in Healthy Diet Score (increase vs stayed the same or decrease).

While male users experienced, on average, greater weight loss compared to female users, the difference was much more pronounced among participants who were enrolled for 12-18 months. We also observed that when stratified by age, participants who were older experienced greater weight loss compared with those who were younger in a dose-response relationship. Participants in the highest age category of 60 and older lost the most weight, and this association became more robust with longer enrollment duration. Similarly, when stratified by baseline obesity class, participants who were in obesity class 3 had the largest improvements in weight, followed by obesity class 2 and then obesity class 1. The associations strengthened with enrollment duration. Participants who increased their Healthy Diet Score between their first and last reports of dietary intake experienced greater weight loss compared with participants whose Healthy Diet Score stayed the same or decreased.

Figure 3. Mean (SD) percent weight change stratified by enrollment length and A) gender, B) age category, C) baseline obesity class, and D) change in Healthy Diet Score. Gray error bars indicate standard deviations of the mean; * indicates a statistically significant difference between groups assessed using ANOVA tests and a Bonferroni-corrected *P* value of .0031 to adjust for multiple comparisons.



Discussion

In the present study of 8977 Foodsmart platform users with obesity, we found that 59% of participants reported a decrease in weight, and 24% reported at least 5% weight loss of their initial weight while enrolled in the program; median (IQR) duration of follow-up was 9.9 months (0.03-54.7). Baseline BMI, baseline Healthy Diet Score, change in Healthy Diet Score, and longer duration of enrollment were all associated with higher odds of achieving 5% or greater weight loss. Age and gender were not associated with at least 5% weight loss. We found that the percentage of participants achieving 5% weight loss increased with enrollment duration. These findings suggest that Foodsmart platform users with obesity are likely to lose weight and that longer enrollment duration could potentially lead to greater weight loss. We believe these results to be clinically significant as 5% loss of initial weight has been linked to improved health outcomes among people who are obese, with prediabetes, or type 2 diabetes [9,12,29,30].

These results are in line with previous studies that found digital nutrition interventions to be successful in weight loss [31,32]. The majority of prior studies on digital apps has focused on nutrition monitoring and reporting or health coaching. However, the tools of the Foodsmart platform are unique in that, in addition to dietary recommendations, the app offers a personalized meal and recipe planning program and a unique food purchasing environment that addresses barriers to healthy eating by offering healthy options for everyone such as online ordering and delivery of groceries, meal kits, and prepared foods. Meal planning and cooking at home have been found to be associated with better diet quality and lower likelihood of obesity, primarily due to having control of ingredients, cooking methods, and portion sizes [22,23]. Although the use of commercial online grocery shopping, delivery, and meal kits has been increasing in recent years, few studies have examined the impact of these new purchasing behaviors on health outcomes. A study on medically tailored meal delivery for patients with diabetes and food insecurity found that home delivery of 10 meals per week for 12 weeks was associated with improvements in Healthy Eating Index Score, food insecurity,

and hypoglycemia [33]. However, this was a short-term study with direct food provisions, which do not precisely mirror the Foodsmart platform. Nonetheless, the study suggests that healthy food delivery may be a viable strategy in improving health outcomes. More research is warranted to evaluate the potential cost savings of these types of programs that change the food purchasing environment to create healthier eating.

The finding that change in Healthy Diet Score was the strongest predictor of achieving 5% weight loss is noteworthy. The Healthy Diet Score captures overall dietary quality and serves as a proxy for engagement with the Foodsmart platform since the program is designed to improve diet quality. This demonstrates that participants who used the Foodsmart program and improved their diet quality were more likely to lose weight. Furthermore, the association between change in Healthy Diet Score and weight loss was compounded by duration of the program (Figure 3D). This finding is in agreement with previous studies that have found weight loss from mobile health apps to be greater with longer enrollment duration [34]. However, these findings showed that among people who used the program for over 12 months, 5% weight loss was achieved by one-third of users. Previous studies have shown that long-term maintenance of weight loss is challenging as more than half of lost weight was regained within 2 years [35,36]. Although this study was not designed to examine whether weight loss was sustained, we found that longer enrolment duration was associated with greater weight loss. Additional research is needed to further examine the sustainability of this type of intervention by examining trends of multiple weight measurements over time.

We found that despite more females using the program compared to males, that on average, male users lost more weight compared to females when stratified by enrollment length and gender (Figure 3A). However, in the fully adjusted logistic regression model, we did not find a statistically significant association between gender and 5% weight loss. While other studies have also found that males lost more weight compared to women when using health apps [34], the reason for the greater percent change could be higher baseline weight in male users. It was also interesting that when mean percent weight change was stratified by enrollment duration and age category, participants who were older consistently lost more weight, and the effect compounded with longer duration of enrollment. For participants under age 30, on average, weight increased if they enrolled for longer than 12 months. This finding was contrary to what some might expect given high rates of technology use by younger adults [37]. It may be that meal planning and food purchasing interventions may be more successful among older adults, or they may be more focused on their health. Or, this may be due to the increasing trend of eating out and less cooking, leading to weight gain, among younger populations [38,39]. Additionally, younger users may have been more likely to disengage with the app and then re-engage after gaining weight.

There are several limitations of the present study. Due to the observational nature of this study, we cannot conclude any causal associations between change in diet quality and weight loss. Since we did not have a control group, it is difficult to attribute a weight loss to the Foodsmart platform itself. This study serves as an exploration in which factors are associated with weight loss among Foodsmart users. Since we did not have exact dates for leaving the program, we used the last entry of weight as a proximal end date. Because we are using real-world data rather than the settings of a controlled study, participants were free to start and stop usage of the app as they wished. Therefore, it is challenging to draw firm conclusions on how duration of usage was associated with weight loss. Another limitation is that measures of height, weight, and diet were self-reported by participants. However, prior studies suggest that there is moderate to high agreement between online self-reported and measured anthropometric data [40]. Unfortunately, we did not have information on other factors that may be important predictors of weight loss such as total energy intake, race, or socioeconomic status. We did not assess engagement level or usage among participants since our goal was to examine the overall Foodsmart program. All users in this analysis took and retook the Nutriquiz and weight change assessments, which may, in and of themselves, have driven an impact due to their ability to motivate and drive self-insight and knowledge about nutrition and to track progress.

This study also had several strengths. With almost 9000 participants, this study included a large number of participants with obesity that provided us with sufficient power to examine percent weight change stratified by 2 variables. We also had a broad range of enrollment lengths, allowing us to examine weight change and maintenance in a time span of more than 2 years. Few studies, especially randomized clinical trials, on digital apps have follow-up data for weight change after more than 2 years.

In conclusion, this was one of the first studies of this scale and time length to examine weight loss among individuals with obesity who were users of a digital nutrition platform with personalized dietary recommendations and online meal planning, food ordering, grocery discounts, and incentives. Future studies are warranted to determine the sustainability and cost-effectiveness of weight loss through a digital nutrition intervention with these features vs other alternatives. Randomized clinical trials are needed to tease out causal associations.

Acknowledgments

The authors would like to thank Christine Vaughan for performing initial analyses for this study. The study was funded by Foodsmart.

Authors' Contributions

EAH analyzed data, interpreted results, and wrote the manuscript. VN acquired data. JL and DS interpreted results. All the authors reviewed and approved the final version of the manuscript and take responsibility for the manuscript.

Conflicts of Interest

EAH, VN, JL, and DS are employees of and own stocks of Zipongo, Inc DBA Foodsmart.

Multimedia Appendix 1
Major changes to the Foodsmart platform since 2013.
[DOCX File , 13 KB - jmir_v22i9e19634_app1.docx]

References

1. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. N Engl J Med 2017 Dec 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
2. GBD 2013 Risk Factors Collaborators, Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, et al. . [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]
3. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016. JAMA 2018 Apr 24;319(16):1723-1725 [FREE Full text] [doi: 10.1001/jama.2018.3060] [Medline: 29570750]
4. Seidell JC, Halberstadt J. The global burden of obesity and the challenges of prevention. Ann Nutr Metab 2015;66 Suppl 2:7-12 [FREE Full text] [doi: 10.1159/000375143] [Medline: 26045323]
5. Kopelman PG. Obesity as a medical problem. Nature 2000 Apr 06;404(6778):635-643. [doi: 10.1038/35007508] [Medline: 10766250]
6. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. Lancet 2011 Aug 27;378(9793):815-825. [doi: 10.1016/S0140-6736(11)60814-3] [Medline: 21872750]
7. Johnston BC, Kanters S, Bandayrel K, Wu P, Naji F, Siemieniuk RA, et al. Comparison of weight loss among named diet programs in overweight and obese adults: a meta-analysis. JAMA 2014 Sep 03;312(9):923-933. [doi: 10.1001/jama.2014.10397] [Medline: 25182101]
8. Kouvelioti R, Vagenas G, Langley-Evans S. Effects of exercise and diet on weight loss maintenance in overweight and obese adults: a systematic review. J Sports Med Phys Fitness 2014 Aug;54(4):456-474. [Medline: 24739257]
9. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 07;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
10. Wadden TA, Volger S, Sarwer DB, Vetter ML, Tsai AG, Berkowitz RI, et al. A two-year randomized trial of obesity treatment in primary care practice. N Engl J Med 2011 Nov 24;365(21):1969-1979 [FREE Full text] [doi: 10.1056/NEJMoa1109220] [Medline: 22082239]
11. Goodpaster BH, Delany JP, Otto AD, Kuller L, Vockley J, South-Paul JE, et al. Effects of diet and physical activity interventions on weight loss and cardiometabolic risk factors in severely obese adults: a randomized trial. JAMA 2010 Oct 27;304(16):1795-1802 [FREE Full text] [doi: 10.1001/jama.2010.1505] [Medline: 20935337]
12. Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care 2011 Jul;34(7):1481-1486 [FREE Full text] [doi: 10.2337/dc10-2415] [Medline: 21593294]
13. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: Innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015 Jan;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]
14. Rhee E. Weight Cycling and Its Cardiometabolic Impact. J Obes Metab Syndr 2017 Dec 30;26(4):237-242 [FREE Full text] [doi: 10.7570/jomes.2017.26.4.237] [Medline: 31089525]
15. Azar KMJ, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
16. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular Nutrition-Related Mobile Apps: A Feature Assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
17. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]

18. Kankanhalli A, Shin J, Oh H. Mobile-Based Interventions for Dietary Behavior Change and Health Outcomes: Scoping Review. JMIR Mhealth Uhealth 2019 Jan 21;7(1):e11312 [FREE Full text] [doi: 10.2196/11312] [Medline: 30664461]
19. Martin CK, Miller AC, Thomas DM, Champagne CM, Han H, Church T. Efficacy of SmartLoss, a smartphone-based weight loss intervention: results from a randomized controlled trial. Obesity (Silver Spring) 2015 May;23(5):935-942. [doi: 10.1002/oby.21063] [Medline: 25919921]
20. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes Prevention and Weight Loss with a Fully Automated Behavioral Intervention by Email, Web, and Mobile Phone: A Randomized Controlled Trial Among Persons with Prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
21. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
22. Monsivais P, Aggarwal A, Drewnowski A. Time spent on home food preparation and indicators of healthy eating. Am J Prev Med 2014 Dec;47(6):796-802 [FREE Full text] [doi: 10.1016/j.amepre.2014.07.033] [Medline: 25245799]
23. Ducrot P, Méjean C, Aroumougame V, Ibanez G, Allès B, Kesse-Guyot E, et al. Meal planning is associated with food variety, diet quality and body weight status in a large sample of French adults. Int J Behav Nutr Phys Act 2017 Feb 02;14(1):12 [FREE Full text] [doi: 10.1186/s12966-017-0461-7] [Medline: 28153017]
24. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
25. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
26. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study. Am J Epidemiol 2001 Dec 15;154(12):1089-1099. [doi: 10.1093/aje/154.12.1089] [Medline: 11744511]
27. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, et al. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr 2012 Jun;142(6):1009-1018 [FREE Full text] [doi: 10.3945/jn.111.157222] [Medline: 22513989]
28. Hendrie GA, Baird D, Golley RK, Noakes M. The CSIRO Healthy Diet Score: An Online Survey to Estimate Compliance with the Australian Dietary Guidelines. Nutrients 2017 Jan 09;9(1):47 [FREE Full text] [doi: 10.3390/nu9010047] [Medline: 28075355]
29. Douketis JD, Macie C, Thabane L, Williamson DF. Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. Int J Obes (Lond) 2005 Oct;29(10):1153-1167. [doi: 10.1038/sj.ijo.0802982] [Medline: 15997250]
30. Blackburn G. Effect of degree of weight loss on health benefits. Obes Res 1995 Sep;3 Suppl 2:211s-216s [FREE Full text] [doi: 10.1002/j.1550-8528.1995.tb00466.x] [Medline: 8581779]
31. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing Self-Monitoring Strategies for Weight Loss in a Smartphone App: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
32. Senecal C, Collazo-Clavell M, Larrabee BR, de Andrade M, Lin W, Chen B, et al. A digital health weight-loss intervention in severe obesity. Digit Health 2020;6:2055207620910279 [FREE Full text] [doi: 10.1177/2055207620910279] [Medline: 32180992]
33. Berkowitz SA, Delahanty LM, Terranova J, Steiner B, Ruazol MP, Singh R, et al. Medically Tailored Meal Delivery for Diabetes Patients with Food Insecurity: a Randomized Cross-over Trial. J Gen Intern Med 2019 Mar;34(3):396-404 [FREE Full text] [doi: 10.1007/s11606-018-4716-z] [Medline: 30421335]
34. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
35. Anderson JW, Konz EC, Frederich RC, Wood CL. Long-term weight-loss maintenance: a meta-analysis of US studies. Am J Clin Nutr 2001 Nov;74(5):579-584. [doi: 10.1093/ajcn/74.5.579] [Medline: 11684524]
36. Hall KD, Kahan S. Maintenance of Lost Weight and Long-Term Management of Obesity. Med Clin North Am 2018 Jan;102(1):183-197 [FREE Full text] [doi: 10.1016/j.mcna.2017.08.012] [Medline: 29156185]
37. van der Vaart R, van Driel D, Pronk K, Paulussen S, Te Boekhorst S, Rosmalen JGM, et al. The Role of Age, Education, and Digital Health Literacy in the Usability of Internet-Based Cognitive Behavioral Therapy for Chronic Pain: Mixed Methods Study. JMIR Form Res 2019 Nov 21;3(4):e12883 [FREE Full text] [doi: 10.2196/12883] [Medline: 31750839]
38. Hartmann C, Dohle S, Siegrist M. Importance of cooking skills for balanced food choices. Appetite 2013 Jun;65:125-131. [doi: 10.1016/j.appet.2013.01.016] [Medline: 23402717]
39. Larson NI, Perry CL, Story M, Neumark-Sztainer D. Food preparation by young adults is associated with better diet quality. J Am Diet Assoc 2006 Dec;106(12):2001-2007. [doi: 10.1016/j.jada.2006.09.008] [Medline: 17126631]

40. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self-reported height, weight, and body mass index in young adults? J Med Internet Res 2014 Jan 07;16(1):e4 [FREE Full text] [doi: 10.2196/jmir.2909] [Medline: 24398335]

Abbreviations

AHEI-2010: Alternative Healthy Eating Index-2010
ANOVA: analysis of variance
CSIRO: Commonwealth Scientific and Industrial Research Organization
DPP: Diabetes Prevention Program

Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by L König, C Mauch, CA Almenara; comments to author 20.05.20; revised version received 17.06.20; accepted 13.08.20; published 02.09.20.

Please cite as:
Hu EA, Nguyen V, Langheier J, Shurney D
Weight Reduction Through a Digital Nutrition and Food Purchasing Platform Among Users With Obesity: Longitudinal Study
J Med Internet Res 2020;22(9):e19634
URL: http://www.jmir.org/2020/9/e19634/
doi:10.2196/19634
PMID:32792332

Original Paper

A Personalized Voice-Based Diet Assistant for Caregivers of Alzheimer Disease and Related Dementias: System Development and Validation

Juan Li[1], PhD; Bikesh Maharjan[1], BSc; Bo Xie[2], PhD; Cui Tao[3], PhD

[1]North Dakota State University, Fargo, ND, United States

[2]University of Texas, Austin, TX, United States

[3]University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Cui Tao, PhD
University of Texas Health Science Center at Houston
7000 Fannin Street Suite 600
Houston, TX, 77030
United States
Phone: 1 7135003981
Email: cui.tao@uth.tmc.edu

Abstract

Background: The world's aging population is increasing, with an expected increase in the prevalence of Alzheimer disease and related dementias (ADRD). Proper nutrition and good eating behavior show promise for preventing and slowing the progression of ADRD and consequently improving patients with ADRD's health status and quality of life. Most ADRD care is provided by informal caregivers, so assisting caregivers to manage patients with ADRD's diet is important.

Objective: This study aims to design, develop, and test an artificial intelligence–powered voice assistant to help informal caregivers manage the daily diet of patients with ADRD and learn food and nutrition-related knowledge.

Methods: The voice assistant is being implemented in several steps: construction of a comprehensive knowledge base with ontologies that define ADRD diet care and user profiles, and is extended with external knowledge graphs; management of conversation between users and the voice assistant; personalized ADRD diet services provided through a semantics-based knowledge graph search and reasoning engine; and system evaluation in use cases with additional qualitative evaluations.

Results: A prototype voice assistant was evaluated in the lab using various use cases. Preliminary qualitative test results demonstrate reasonable rates of dialogue success and recommendation correctness.

Conclusions: The voice assistant provides a natural, interactive interface for users, and it does not require the user to have a technical background, which may facilitate senior caregivers' use in their daily care tasks. This study suggests the feasibility of using the intelligent voice assistant to help caregivers manage patients with ADRD's diet.

(J Med Internet Res 2020;22(9):e19897) doi:10.2196/19897

KEYWORDS

Alzheimer disease; dementia; diet; knowledge; ontology; voice assistant

Introduction

Problem

The progressive brain disorder of Alzheimer disease (AD) causes brain cells to degenerate and die slowly; degrades memory and thinking skills; and, eventually, destroys the ability to perform the simplest tasks of daily life [1]. AD is the most common form of dementia worldwide and the sixth leading cause of death in the United States [2]. In most people with AD, symptoms first appear in their mid-60s [3]. Because the number and proportion of older persons is increasing dramatically [4], estimates suggest that more than 5.5 million Americans, most of them 65 years or older, may have AD [2]. The current growth of the population 65 years and older is one of the most significant demographic trends in US history [2], prompted by the aging of the baby boomers, who in 2030 will be 66-84 years of age and will number 61 million people [5].

People with AD and related dementias (ADRD) are usually cared for by family members or friends, most commonly in their own homes. According to the Centers for Disease Control and Prevention, in 2019, more than 16 million Americans provided about 18.5 billion hours of unpaid care for family and friends with ADRD [6]. Approximately one-third of dementia caregivers are 65 years or older, and approximately one-quarter of dementia caregivers are "sandwich generation" caregivers, people who care not only for an aging parent but also for children younger than 18 years [6]. ADRD caregivers provide care for a longer duration than caregivers of people with other types of conditions [7]. Taking care of a family member with ADRD is overwhelming, and it is important to help caregivers care not only for patients but also for themselves.

Existing Digital Tools for Healthy Diet

So far, there is no cure for ADRD. Limited medical treatments are available for ADRD's symptoms, but good care from family caregivers can make the biggest difference in patients' quality of life [8]. Proper food and nutrition are one of the easiest, most effective, and potentially most enjoyable ways of preventing ADRD and slowing its development [9]. Poor nutrition can increase behavioral symptoms and cause weight loss, and there are many recommendations and guidelines for an ADRD diet [10-13]. However, given the overwhelming amount of available information, it is difficult for the busy caregiver to decide exactly what is best for the patient. In addition, it is not easy for caregivers to remember all the guidelines and constraints. Websites and mobile apps are available for nutrition coaching and meal planning [14,15], but they are not well accepted by ADRD caregivers for several reasons: these tools do not specifically target people with ADRD; their recommendations tend to be generic without considering patients' specific preferences, comorbidities, culture, or traditions; the tools do not consider caregivers' time limits for preparing meals or their economic ability to afford food items; and the complexity of such digital tools makes them intimidating for ADRD caregivers, especially older adults who struggle in using computers or smartphones due to inexperience with technology.

Voice-Enabled Technology

Voice-enabled technology, which is becoming popular worldwide, offers a potential solution. In the United States, the number of people who use voice assistants such as Amazon's Alexa or Apple's Siri is rapidly increasing; in 2019, more than 111 million people in the United States used a voice assistant at least monthly, equivalent to 39% of internet users and 34% of the total US population [16]. Amazon's Alexa is the most popular voice technology, with a 75% share of the market [17]. This voice interface is capable of performing various tasks and controlling various systems. Currently, interaction and communication with Alexa are available in English, German, French, Italian, Spanish, and Japanese [18].

Alexa's capabilities can be extended with "voice skills," apps developed by third-party vendors. As of March 23, 2020, there were more than 70,000 skills available. These "skills" enable users to interact with their Alexa-enabled devices using natural language in a variety of ways including gaming; shopping; and gathering information about news, weather, travel, health, and fitness.

As voice-activated technology has become increasingly popular, researchers and engineers have begun to develop it for use in health care domains, ranging from health news briefs, nutritional guides, and fitness trackers to programs for meditation and yoga. The Mayo Clinic, for example, has launched a "First Aid" Alexa voice skill that answers questions about how to treat common mishaps such as cuts and burns [19]. It offers self-care instructions for first aid, as well as guidance on when a person should seek emergency help.

Objectives

In this paper, we present the design, implementation, and preliminary testing of an artificial intelligence–powered voice assistant to provide personalized education and guidance for caregivers on food, nutrition, and cooking for a loved one with ADRD. This voice assistant gives the ADRD caregiver continuous access to useful tips about food, nutrition, and eating behaviors. It also recommends food and meals. The tips and recommendations are personalized, specific to the patients with ADRD's condition including the patient's ADRD stage, preferences, and medical conditions. Moreover, the assistant's recommendations are specific to the caregiver, based on the caregiver's time limits, financial ability, and education level. Because voice is our most common mode of communication, a voice assistant provides a natural way to engage with technology that requires minimal training. This is especially valuable for caregivers who are older adults, who account for more than one-third of caregivers, and who may find it difficult to use other forms of technology that require fine motor skills, hand-eye coordination, or good vision.

Methods

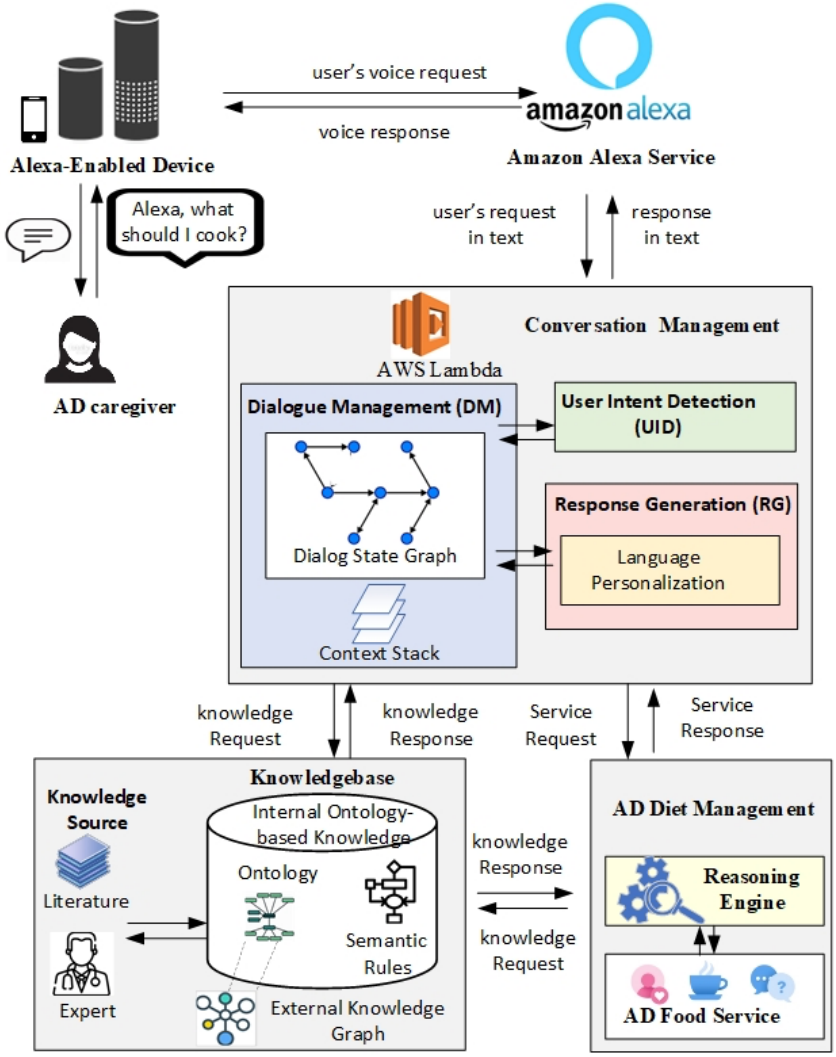
System Overview

The goal of our project is to design a voice-enabled intelligent assistant to support effective diet management for ADRD caregivers. The voice assistant is built on top of a comprehensive personalized ADRD diet care knowledge base, a semantics-based knowledge graph reasoning engine, and Amazon's cloud-based voice service Alexa.

Figure 1 illustrates the system architecture that enables a caregiver's access via Alexa. This architecture has three major components: (1) conversation management, which manages conversations with users; (2) a knowledge base, which provides knowledge for the system; and (3) ADRD diet management, which provides detailed personalized diet education, recommendations, and planning services to users. We have designed and implemented these three components as shown in the three boxes in Figure 1. At the same time, we use Alexa's existing services to recognize the user's speech, convert the user's speech to text, and transform the system's response to speech and respond to the user. The conversation management component includes three major modules: user intent detection (UID), dialogue management (DM), and response generation (RG). The knowledge base component includes internal and external knowledge. The internal knowledge consists of

ontologies and rules defined by ontological language. External knowledge includes related existing knowledge that can be reused in the system. ADRD diet management includes a semantic reasoning engine to support customized question answering, education, recommendations, and planning services for food and nutrition.

Figure 1. Architecture of the system. AD: Alzheimer disease; AWS: Amazon Web Services.



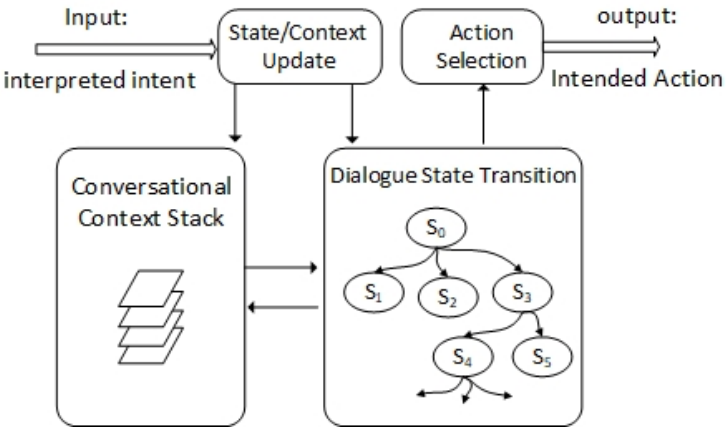
As illustrated in Figure 1, when the system receives a user's request, the request is sent to the conversation management component, where it goes through the three major modules of conversation management to produce a response. First, the UID module analyzes the user's request or command text and matches the text with predefined intent and the dialogue state. Second, the DM module examines the input, executes the dialogue policy, and updates the dialogue state. Based on the user's intent, the Amazon Web Services' (AWS's) Lambda service will call back-end ADRD diet management information corresponding to the intent. This service is supported mainly by a semantics-based reasoner, working on the user's input, facts, and rules defined in the knowledge base. The semantic reasoning links the dialogue with the user's (ie, the caregiver's and patient's) profile, context, and ADRD medical guidelines [10-12,20-23], enabling personalized services. Third, the RG module uses the speech act and content selected by the DM module to build a response. The response language is chosen based on the user's background such as education.

Conversation Management

This component allows the user to interact with the voice assistant. It receives the user's voice input from Amazon Alexa. Alexa converts the user's speech into text using automatic speech recognition and natural language understanding. The UID module matches the user's voice, converted to text, with a predefined intent. The intent corresponds to a system action that fulfills the user's spoken request. To effectively identify the user's intent, we extend users' utterances with synonyms and ontology concepts (ie, classes and instances) from our knowledge base. Because the system focuses on diet management, we have chosen top-level food-, nutrition-, and diet-related classes, and classes or instances that are connected to the chosen top-level classes by two major relationships of an ontology: the hierarchical specialization/generalization (or *IS-A*) relationship and the *type* relationship. This allows the DM module to better identify ADRD food management–relevant entities.

Figure 2 illustrates the information flow of the DM module. The interpreted user's intent is the input to the module. The intent corresponds to an action that fulfills the user's spoken request. Based on our requirement analysis, in our prototype system, we have predefined five categories of intents such as providing tips on proper diet and suggestions for healthy meals. When a user speaks to the voice assistant (asks a question, requests a service, or answers a question), the voice assistant elicits required information or clarifies unclear information. To determine what questions the virtual assistant should ask, in what order, and when, we use a finite-state approach to model the dialogue structure. A state transition graph is designed for every type of service request. This graph can model and track context, different dialogue states, and corresponding transitions to maintain the flow of the dialogue. It includes an initial state, S0, representing the start of the conversation, and a set of successor states for any particular state (eg, S4 and S5 are successors of S3) that may result from input messages. Incoming messages from the user will set the conversation to a new state and a corresponding response. The next state can be calculated based on the dialogue history, the input from the user, and the conversation's context. Dialogue history is recorded as the path of states from the initial state to the current state. The graph helps the system produce coherent responses within an ongoing conversation. It also guides the conversation's direction toward a predefined schema instead of letting the conversation's topic drift randomly or letting the conversation be controlled by the user.

Figure 2. Information flow of dialogue management module.



In a dialogue session, context information is integrated with dialogue states to determine the next potential state. A short-term context will be kept for a few dialogue states, but a long-term context may be kept permanently. For example, a patient with ADRD's age, ADRD stage, food preferences, and allergies constitute long-term context information; the meal a caregiver chooses for lunch can be short-term context information. Different context information may have different expiry times. The topic of the last conversation and the last conversation state in the dialogue state graph are also stored in the session context.

After an action is determined on the basis of the dialogue state graph, it is returned to the RG module. At each turn of conversation between the user and the voice assistant, the DM module performs a series of steps to collect all the required data (slots) from the user. It then tries to find the response that both addresses the user's request and meets the constraints of the conversation topic and context. The response requires knowledge support from the back-end AD diet management component. This component bridges the gap between high-level application logic and low-level dialogue specifications. It allows the dialogue to be dynamically adapted according to the current high-level context.

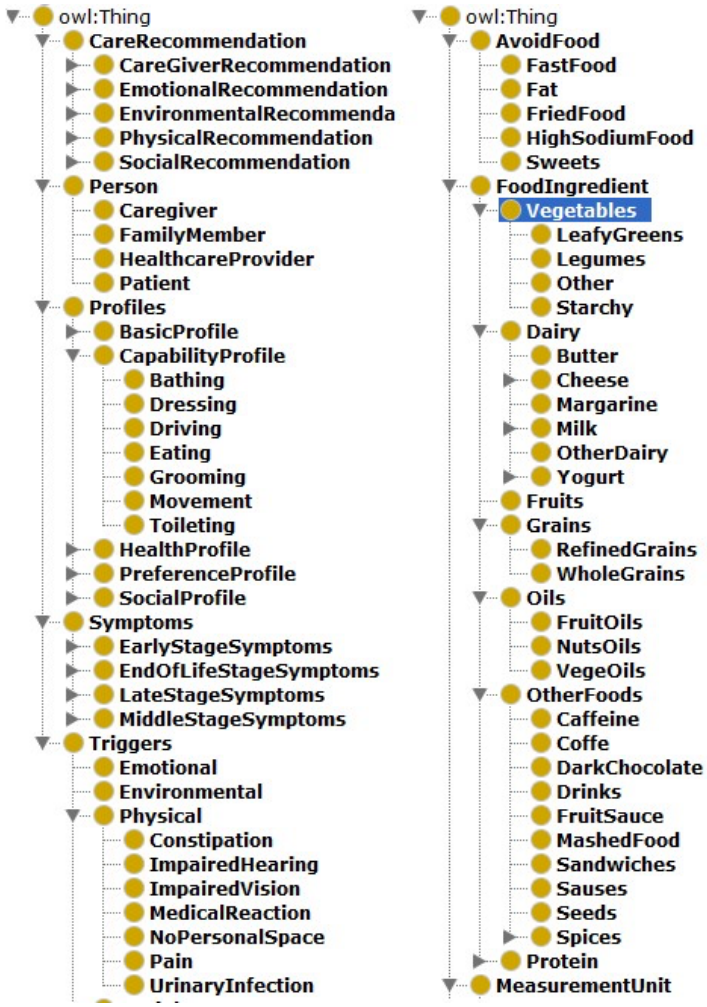
Knowledge Base

The voice assistant's "brain" is a comprehensive knowledge base, which stores the knowledge that informs the answers, recommendations, and tips that the voice assistant provides. To construct this knowledge base, we conducted an extensive review of the scientific literature on ADRD home care and on food and nutrition tips for ADRD care [10-12,20-23]. We used an ontology-based information model to fuse heterogeneous-related information and knowledge in constructing the knowledge base, thus extending Alexa's knowledge base by adding new customized knowledge. This ontology focuses on concepts, relations, constraints, and rules about general care for patients with ADRD, diet guidelines for ADRD, each patient's profile, and each caregiver's profile. We thus enrich and extend our knowledge with existing (semi)structured knowledge, for example, by adding knowledge related to disease (Unified Medical Language System [24], SNOMED Clinical Terms [25], International Statistical Classification of Diseases and Related Health Problems, 10th revision [26]), drugs (DrugBank [27]), and food and nutrition (United States Department of Agriculture's [USDA's] FoodData Central [28]). Currently, we have integrated only the USDA's FoodData Central for detailed food and nutrition knowledge within the system. This food and nutrition information, together with the ADRD diet guidelines that we have collected [10-13,20-23], should be able to address most of the proposed services required by the voice assistant. This is an ongoing project; we are working to enrich the voice assistant's functionality to support other aspects of ADRD care, which may benefit from other knowledge sources.

An ontology represents a domain of discourse explicitly. It defines the concepts and relationships used to describe and represent the domain [29]. It defines domain knowledge semantically in standard ways that allow comprehensive,

transportable machine understanding. The knowledge base mainly uses three ontologies: (1) a user profile that captures both the patient's and the caregiver's social, economic, cultural, and physical properties, and provides evidence for personalization; (2) ADRD diet support and education, which captures essential knowledge in a user-friendly language for healthy eating; and (3) food and nutrition knowledge extracted from other sources [30] but reorganized for our application's purpose. Figure 3 shows part of the system's high-level ontology.

Figure 3. Part of the classes defined in the ontology used in the knowledge base.



Besides concepts (classes) and relationships (properties) defined in the ontology, the knowledge base includes rules based on professional medical guidelines related to AD diet. We have collected AD general caring guidelines and food and nutrition guidelines such as from the National Institute on Aging [23], Alzheimer's Association [22], Mayo Clinic [31], alzheimers.net, the National Heart Foundation of Australia [20], and the USDA [32]. The collected rules were further verified by two experienced clinicians on our research team who specialize in the diagnosis and management of cognitive disorders including ADRD and were then converted to rules that computers can "understand." We use the Semantic Web Rule Language (SWRL) [33], an expressive World Wide Web Consortium standard Web Ontology Language (OWL)–based rule language to present the generated rules.

AD Diet Management

The back-end ADRD diet management component receives the user's request as input and calls the corresponding back-end service to generate a response that is sent back to the RG module in the conversation management component. The back-end services are built to enable context-aware personalized recommendations, education, and question answering. The foundation of this component is a description logic (DL) [34] query-answering and reasoning engine over the OWL ontology and its extended knowledge graph. In particular, we use DL to check the satisfiability of the whole knowledge base and answer complex queries (eg, unions of conjunctive queries) over the knowledge base. We separate the ontology into TBox (classes and properties) and ABox (instances) [35]. Therefore, reasoning can be performed on two different levels, the TBox level and ABox level. This improves reasoning performance by avoiding complex operations over a large number of ABox instances. Instead, most of the queries can be performed using the Structured Query Language engine over the ABox triple store.

A backward-chaining algorithm is used for automatic reasoning. Backward chaining [36] is a goal-driven reasoning process. To prove a goal, all of its subgoals need to be recursively proved. The reasoning engine uses data in the knowledge base to match the initial goal by matching rules for true consequent and then

assuming true antecedent. As an example, two class subsumption rules are given in equations 1 and 2.

Supposing a being of class D and A, the following backward chaining application in equation 3 illustrates how to prove a being of class B.

The reasoning engine takes the AD diet management rules in SWRL format [33], facts from food and nutrition ontology, and the patient with ADRD's and caregiver's profile information from the profile ontology in the knowledge base. When a user's query or request is passed to the back end from the DM module, the contextual information and auxiliary information such as intent and sentiment are fed into the query and reasoning engine together with the rules and facts to control the generation of responses. The goals (and subgoals) always match the affirmed versions of the consequents of implications, and their antecedents are then considered as new goals, which ultimately must match known facts (eg, in the previously mentioned example, a being of class D and A). The backward chaining–based reasoner links the dialogue between the user and voice assistant with the profile of the particular caregiver and patient, other context, and ADRD diet medical guidelines. In this way, the system can provide context-aware personalized recommendations.

Results

Prototype Implementation

We have implemented a prototype of the voice assistant as an Amazon skill "ADRD-DietAssistant." We have conducted a set of preliminary evaluation tests to assess the voice assistant's technical viability and effectiveness.

- Front-end implementation:
  - Alexa Skills Kit (ASK):
    - Dialogue design, intent definition, utterances to activate intents, slot customization
  - The Amazon Lambda service:
    - Runtime: Node.js 12.14.1
    - Software development kit (SDK) to communicate with Alexa: ASK-SDK
    - Logging service: Amazon CloudWatch Logs
- The back-end diet management server is configured as follows:
  - Processor: Intel Core i-7-4770 CPU @ 3.40 GHz (8 CPUs), ~ 3.4 GHz
  - RAM: 32 GB
  - Operating system: Windows 10 Home 64-bit (10.0, Build 17763)
  - Programming Language: Java 8, Java Development Kit 1.8
  - Semantic service: OWL Application Programming Interface and SparkJava
  - Database: MySQL: The back-end AD Diet Management

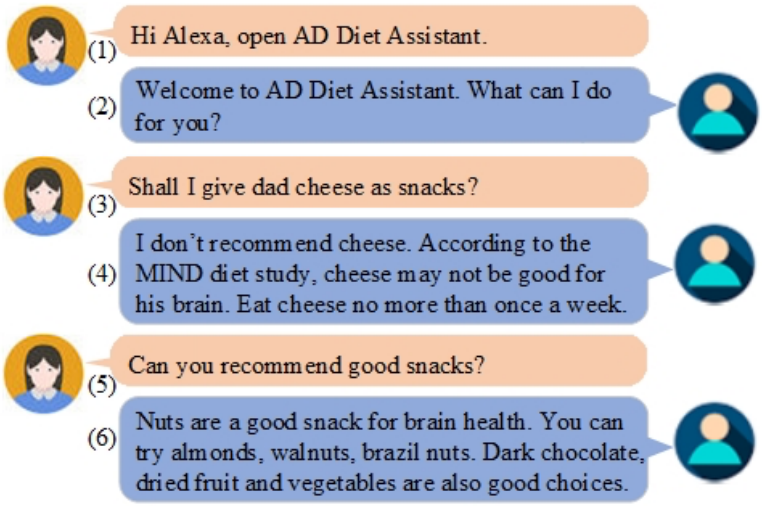
Evaluation Outcomes

Use Cases

The voice assistant can support a variety of food and nutrition recommendations, education, and planning services including tips on proper diet, handling challenges in eating, food item and nutrition explanations, meal suggestions, and recall of daily and weekly diet histories. In a conversation, different service categories can be integrated to serve the user's requests.

Figure 4 demonstrates one example of a conversation. The user is a female caregiver; she is taking care of her 65-year-old father, who is in early-stage ADRD with type 2 diabetes. In this conversation, the user asks the voice assistant about the suitability of food items and gets the right food recommendation.

Figure 4. An example of a conversation between the user and the voice assistant about food and nutrition education, and meal suggestion. AD: Alzheimer disease; MIND: Mediterranean–Dietary Approaches to Stop Hypertension Intervention for Neurodegenerative Delay.

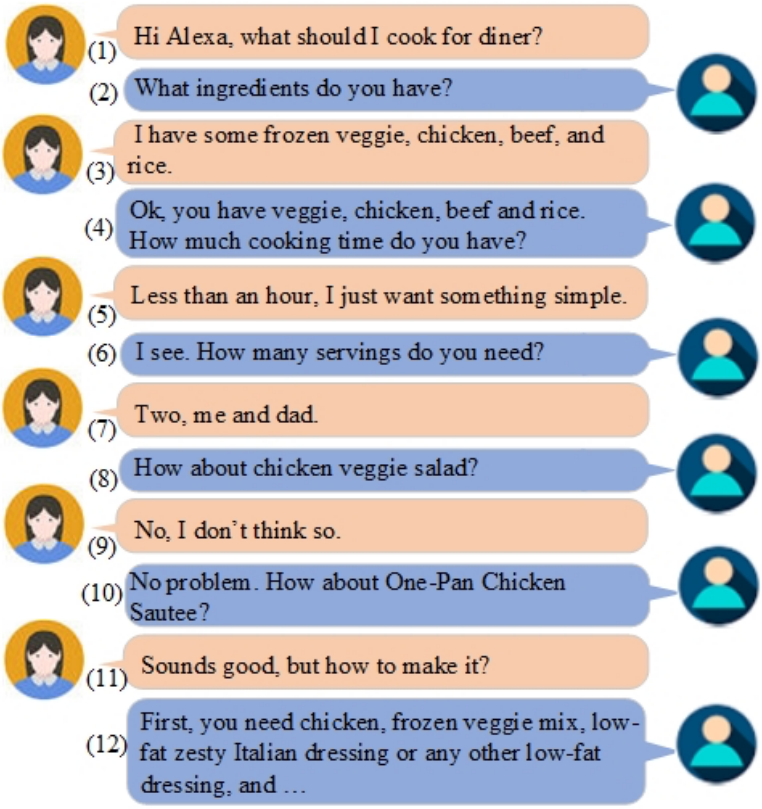


Utterance 1 is an invocation that activates the diet assistant voice skill. Once the skill is activated, the voice assistant greets the user (Utterance 2). A conversation then follows based on the user's requests. The voice assistant triggers a specific intent

designed to handle the user's specific request. In Utterance 3, the user asks the voice assistant whether a certain food is good for the patient. This triggers the back-end service to check whether the food item—cheese in this case—is good as a snack or not. The variables collected from this utterance are "cheese" and "snack," which are passed on to the system's back end. The system's back end has an intent handler that then checks to determine whether "cheese" is good as a "snack" for people with Alzheimer and diabetes, using facts and rules stored in the ontological knowledge base. It then prepares a speech to be sent back to the user as a response, as shown in Utterance 4. In Utterance 5, the user asks for a good snack suggestion. The back end then uses intent chaining to find a healthy "snack." The response is shown in Utterance 6.

Another sample conversation is shown in Figure 5. In this conversation, the voice assistant recommends a healthy recipe to the caregiver in response to her request, the context of the conversation, the patient's profile, and healthy eating guidelines. Utterances 2-7 show how the voice assistant acquires the information to make an appropriate recommendation so that it can prepare all the slots to call the back-end intent. Utterance 7 shows that a user may provide more information than was requested by the voice assistant. The voice assistant gives a recommendation in Utterance 8. The user rejects the suggestion in Utterance 9. Utterance 10 shows that the voice assistant acknowledges the rejection and provides an alternative suggestion. In Utterance 12, the voice assistant presents the recommendation.

Figure 5. An example fragment of a conversation between the user and the voice assistant about meal recipe suggestion.



To provide a satisfying response to a user's requests, the back-end diet management server will consider multiple factors that affect the patient's diet decisions, such as ADRD nutrition requirements, food availability, cooking time, servings, preferences, medical restrictions, and cultural and religious constraints. The system extracts all these constraints from the user's profile, context, and health guidelines.

For the recipe recommended in Figure 5, the system follows multiple guidelines in the knowledge base, including the following:

- The Mediterranean–Dietary Approaches to Stop Hypertension Intervention for Neurodegenerative Delay (MIND) diet [10] encourages the consumption of all kinds of vegetables, berries, nuts, olive oil, whole grains, fish, beans, poultry, and a moderate amount of wine.
- The MIND diet also encourages limiting one's consumption of butter and margarine, cheese, red meat, fried food, pastries, and sweets.
- The US Department of Health and Human Services Dietary Guidelines [32] state that if a person consumes three meals per day, one meal should contain 800-850 calories for a man.
- The American Diabetes Association Diabetes Guidelines [37] state that a person's dinner total calories should be 25% of the person's Estimated Energy Requirement, protein intake should be 20%-30% of meal energy, and sugar should be less than 10% of meal energy.

The voice assistant also checks for other factors affecting the user's diet decisions, such as the user's (patient's or caregiver's) preferences, allergies, religious constraints, and so forth.

Qualitative Evaluation

We have performed a set of qualitative evaluations to test the system's performance. The testers, who included one faculty member and four graduate students on our research team, performed 180 conversations with the voice assistant in our research lab. These conversations were invoked by the testers' requests. Each of the requests belonged to one of the following service categories: food item and nutrition explanations, meal and recipe suggestions, and tips for proper diet. Otherwise, testers had maximum freedom and flexibility in talking to the voice assistant.

In our testing, we defined a threshold, the maximum number of turns in the dialogue between the user and the voice assistant, to limit the dialogue's length. If a conversation exceeded this limit and the user did not receive a satisfying response, the conversation was noted as failed. In our evaluation, we set the threshold of maximal dialogue turns at 20. Given this threshold, the dialogue success rate is defined by equation 4:

The correctness of information provided by the voice assistant is defined in equation 5. Correctness is determined by manually checking whether the recommended content is correct on the basis of the medical guidelines and consistent with a user profile artificially generated for the test, that is, whether the recommendation satisfies the user's request and is consistent with the patient's and caregiver's profile, following the constraints and guidelines in the knowledge base:

Table 1 shows the test results. The dialogue success rate ranged from 80% to 100%. Failed dialogues occurred when the voice assistant could not match the user's intention with back-end services—when it could not relate the user's utterance with predefined intents (function) and the slots (parameters) required to call the intents. For recipe recommendations, most failed conversations occurred when the voice assistant could not provide a satisfying recipe at the beginning. Failure also occurred when matching knowledge was not found in the knowledge base. Although Alexa can effectively match what is said by the user against all possible combinations of defined utterances and slot samples, it has a limited capability in matching an undefined utterance to a slot.

Table 1. Dialogue performance test.

Metrics/purpose	Category 1	Category 2	Category 3
Average dialogue time (sec)	29.7	68.8	45.7
Time per turn (sec)	6.75	6.68	6.34
Number of turns per dialogue	4.4	10.3	7.2
Dialogue success rate, n/N	51/59	48/60	61/61
Recommendation correctness rate, n/N	51/51	48/48	61/61

Discussion

Principal Results

Healthy eating may help slow and even prevent the progression of ADRD and subsequently improve the health status and quality of life in older age [21]. To assist healthy eating, we have designed, developed, and evaluated a personalized, actionable, and engaging voice assistant to help ADRD caregivers manage patients' diet. Building on an ontology-based knowledge base extended with external knowledge graphs and a reasoning and query-answering engine, the voice assistant can answer user's questions and provide personalized and refined food and nutrition recommendations. It follows evidence-based clinical guidelines and takes into account health constraints and personal preferences of both caregivers and patients. The voice assistant has been evaluated using various types of cases. Preliminary test results have demonstrated reasonable rates of dialogue success and recommendation correctness. This study provides preliminary evidence that it is feasible and effective to implement a voice-based virtual assistant to provide diet-related services to ADRD caregivers.

Limitations and Future Work

This study has limitations. The qualitative evaluation was performed within a laboratory setting. Whether or not actual caregivers in a natural setting will use the assistant in ways similar to those of researchers in the laboratory remains for further research. We plan to deploy the voice assistant in a natural setting with ADRD caregivers and patients with early stage ADRD as users. More comprehensive user studies will allow us to evaluate the voice assistant's usability, users' satisfaction, and outcomes for improved health and quality of life. We will also conduct focus groups to collect users' feedback about their knowledge needs to improve our knowledge base.

To increase the dialogue success rate, we propose several approaches to improve the flexibility and accuracy of matching:

- The voice assistant can provide more oral guidance during conversations so that users have a better understanding of the voice assistant's expectations.
- Developers should define more combinations of utterances and synonyms for each intent and its slots to improve the matching rate.
- Although Amazon uses machine learning to match a user's intention with defined intent, this learning is limited, and matching performance largely depends on keywords and sentences provided at the design stage. We will design our machine learning algorithms based on domain knowledge and user feedback to improve accuracy of understanding.

- We will extend our knowledge base with more relevant knowledge.

In addition, we plan to capture conversations and use them for machine learning to improve conversation accuracy. If we can solve potential issues of privacy and security with users' approval, in the future, we will collect and use large amounts of conversation information for increased accuracy.

Conclusions

Voice is the most natural and powerful mode of human communication. Our voice assistant gives users a natural, interactive interface, and it does not require the technical background that might be a barrier to senior caregivers' use of new technologies for daily care tasks. This may enhance caregivers' engagement in ADRD diet care and reduce caregiver burden, improving the care of patients as well as caregivers' own health and well-being.

Acknowledgments

This study was supported, in part, by the National Science Foundation Division of Information and Intelligent Systems Award 1722913.

Conflicts of Interest

None declared.

References

1. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. Alzheimer's Dementia 2019 Mar 01;15(3):321-387. [doi: 10.1016/j.jalz.2019.01.010]
2. Mather M, Scommegna P, Kilduff L. Fact sheet: aging in the United States. Population Reference Bureau. 2019 Jul 15. URL: https://www.prb.org/aging-unitedstates-fact-sheet [accessed 2020-03-07]
3. Holtzman D, Morris J, Goate A. Alzheimer's disease: the challenge of the second century. Sci Transl Med 2011 Apr 06;3(77):77sr1 [FREE Full text] [doi: 10.1126/scitranslmed.3002369] [Medline: 21471435]
4. Ageing and health. World Health Organization. 2018 Feb 15. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2020-03-07]
5. Ortman JM, Velkoff VA, Hogan H. An aging nation: the older population in the United States. U.S. Department of Commerce, Economics and Statistics Administration. 2014 May. URL: https://www.census.gov/prod/2014pubs/p25-1140.pdf [accessed 2020-07-26]
6. Caregiving for a person with Alzheimer's disease or a related dementia. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/aging/caregiving/alzheimer.htm [accessed 2020-07-26]
7. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimers Dement 2016 Apr;12(4):459-509. [doi: 10.1016/j.jalz.2016.03.001] [Medline: 27570871]
8. Yiannopoulou K, Papageorgiou S. Current and future treatments for Alzheimer's disease. Ther Adv Neurol Disord 2013 Jan;6(1):19-33 [FREE Full text] [doi: 10.1177/1756285612461679] [Medline: 23277790]
9. Swaminathan A, Jicha GA. Nutrition and prevention of Alzheimer's dementia. Front Aging Neurosci 2014;6:282. [doi: 10.3389/fnagi.2014.00282] [Medline: 25368575]
10. Morris M, Tangney C, Wang Y, Sacks F, Bennett D, Aggarwal N. MIND diet associated with reduced incidence of Alzheimer's disease. Alzheimers Dement 2015 Sep;11(9):1007-1014 [FREE Full text] [doi: 10.1016/j.jalz.2014.11.009] [Medline: 25681666]
11. Otaegui-Arrazola A, Amiano P, Elbusto A, Urdaneta E, Martínez-Lage P. Diet, cognition, and Alzheimer's disease: food for thought. Eur J Nutr 2014 Feb;53(1):1-23. [doi: 10.1007/s00394-013-0561-3] [Medline: 23892520]
12. Lange K, Lange KM, Makulska-Gertruda E, Nakamura Y, Reissmann A, Kanaya S, et al. Ketogenic diets and Alzheimer's disease. Food Sci Hum Wellness 2017 Mar;6(1):1-9. [doi: 10.1016/j.fshw.2016.10.003]
13. Włodarek D. Role of ketogenic diets in neurodegenerative diseases (Alzheimer's disease and Parkinson's disease). Nutrients 2019 Jan 15;11(1):169 [FREE Full text] [doi: 10.3390/nu11010169] [Medline: 30650523]
14. Gabrielli S, Dianti M, Maimone R, Betta M, Filippi L, Ghezzi M, et al. Design of a mobile app for nutrition education (TreC-LifeStyle) and formative evaluation with families of overweight children. JMIR mHealth uHealth 2017 Apr 13;5(4):e48 [FREE Full text] [doi: 10.2196/mhealth.7080] [Medline: 28408361]
15. Mauch C, Wycherley T, Laws R, Johnson B, Bell L, Golley R. Mobile apps to support healthy family food provision: systematic assessment of popular, commercially available apps. JMIR mHealth uHealth 2018 Dec 21;6(12):e11867 [FREE Full text] [doi: 10.2196/11867] [Medline: 30578213]
16. Petrock V. US voice assistant users 2019: who, what, when, where and why. eMarketer. 2019 Jul 15. URL: https://www.emarketer.com/content/us-voice-assistant-users-2019 [accessed 2020-03-06]

17. Sterling G. Analyst: Amazon Echo owns 75 percent of global smart speaker market. Marketing Land. 2018 Nov 16. URL: https://marketingland.com/analyst-amazon-echo-owns-75-percent-of-global-smart-speaker-market-251963 [accessed 2020-07-24]
18. Amazon Alexa. Wikipedia. URL: https://en.wikipedia.org/wiki/Amazon_Alexa [accessed 2020-07-24]
19. Skills from Mayo Clinic. Mayo Clinic. URL: https://www.mayoclinic.org/voice/apps
20. Gabb G, Mangoni A, Anderson C, Cowley D, Dowden J, Golledge J, et al. Guideline for the diagnosis and management of hypertension in adults - 2016. Med J Aust 2016 Jul 18;205(2):85-89. [doi: 10.5694/mja16.00526] [Medline: 27456450]
21. Jones RW. Preventing Alzheimer's disease. In: Bairu M, Weiner MW, editors. Global Clinical Trials for Alzheimer's Disease. Cambridge, Massachusetts: Academic Press; 2014:33-46.
22. Alzheimer's Association. URL: https://act.alz.org/ [accessed 2020-07-24]
23. National Institute on Aging. URL: https://www.nia.nih.gov/ [accessed 2020-07-24]
24. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
25. Benson T, Grieve G. Principles of Health Interoperability SNOMED CT, HL7 and FHIR. London: Springer-Verlag; 2016.
26. Winters-Miner LA, Bolding PS, Hilbe JM, Goldstein M, Hill T, Nisbet R, et al. ICD-10. In: Practical Predictive Analytics and Decisioning Systems for Medicine. Cambridge, Massachusetts: Academic Press; 2015.
27. Wishart DS DS. DrugBank. In: Altman RB, Flockhart D, Goldstein DB, editors. Principles of Pharmacogenetics and Pharmacogenomics. Cambridge, England: Cambridge University Press; 2012:55-65.
28. FoodData Central. U.S. Department of Agriculture. URL: https://fdc.nal.usda.gov/ [accessed 2020-07-24]
29. Thomas C. Ontology in Information Science. London, UK: IntechOpen; Mar 08, 2018:30-53.
30. Li J, Alian S. Design and development of a biocultural ontology for personalized diabetes self-management of American Indians. In: IEEE. 2018 Nov 12 Presented at: 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom); 17-20 Sept. 2018; Ostrava, Czech Republic. [doi: 10.1109/healthcom.2018.8531107]
31. Mayo Clinic. URL: https://www.mayoclinic.org/ [accessed 2020-07-24]
32. Dietary guidelines for Americans: 2015-2020 eighth edition. U.S. Department of Agriculture. URL: https://health.gov/sites/default/files/2019-09/2015-2020_Dietary_Guidelines.pdf [accessed 2020-07-24]
33. Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M. SWRL: a semantic web rule language combining OWL and RuleML. W3C. 2004 May 21. URL: https://www.w3.org/Submission/SWRL/ [accessed 2020-07-24]
34. Krötzsch M, Simančík F, Horrocks I. A description logic primer. In: Völker J, Lehmann J, editors. Perspectives on Ontology Learning. Amsterdam: IOS Press; 2012:3-19.
35. Calvanese D, De Giacomo G, Lembo D, Lenzerini M, Rosati R. DL-Lite: Tractable Description Logics for Ontologies. In: AAAI. 2020 Apr 03 Presented at: Proceedings of the Twentieth National Conference on Artificial Intelligence. AAAI Press; July 9-13, 2005; Pittsburgh, Pennsylvania p. 602-607 URL: https://www.aaai.org/Papers/AAAI/2005/AAAI05-094.pdf
36. Al-Ajlan A. The comparison between forward and backward chaining. IJMLC 2015 Apr;5(2):106-113. [doi: 10.7763/ijmlc.2015.v5.492]
37. American Diabetes Association. Standards of Medical Care in Diabetes-2016. Diabetes Care 2016 Jan 01;39(Supplement 1):S1-S112 [FREE Full text]

Abbreviations

AD: Alzheimer disease
ADRD: Alzheimer disease and related dementias
ASK: Alexa Skills Kit
DL: description logic
DM: dialogue management
MIND: Mediterranean–Dietary Approaches to Stop Hypertension Intervention for Neurodegenerative Delay
OWL: Web Ontology Language
RG: response generation
SDK: software development kit
SWRL: Semantic Web Rule Language
UID: user intent detection
USDA: United States Department of Agriculture

Edited by G Eysenbach; submitted 05.05.20; peer-reviewed by J Xia, X Jing; comments to author 25.06.20; accepted 22.07.20; published 21.09.20.

Please cite as:
Li J, Maharjan B, Xie B, Tao C
A Personalized Voice-Based Diet Assistant for Caregivers of Alzheimer Disease and Related Dementias: System Development and Validation
J Med Internet Res 2020;22(9):e19897
URL: http://www.jmir.org/2020/9/e19897/
doi:10.2196/19897
PMID:32955452

Original Paper

Quality of Care Perceived by Older Patients and Caregivers in Integrated Care Pathways With Interviewing Assistance From a Social Robot: Noninferiority Randomized Controlled Trial

Roel Boumans[1], MSc; Fokke van Meulen[2,3], PhD; William van Aalst[1], MSc; Joyce Albers[4], MSc; Marèse Janssen[4], MSc; Marieke Peters-Kop[4], MSc; Getty Huisman - de Waal[5], PhD; Alexandra van de Poll[4], MD, MSc; Koen Hindriks[6], PhD; Mark Neerincx[7], PhD; Marcel Olde Rikkert[1], MD, PhD

[1]Geriatric Department, Radboud University Medical Center, Nijmegen, Netherlands

[2]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

[3]Center for Sleep Medicine, Kempenhaege Foundation, Heeze, Netherlands

[4]Geriatric Department, Canisius Wilhelmina Ziekenhuis, Nijmegen, Netherlands

[5]IQ healthcare, Radboud University Medical Center, Nijmegen, Netherlands

[6]Social AI Group, Vrije Universiteit, Amsterdam, Netherlands

[7]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands

Corresponding Author:
Roel Boumans, MSc
Geriatric Department
Radboud University Medical Center
Reinier Postlaan 4
Nijmegen, 6525 GC
Netherlands
Phone: 31 0622372708
Email: roel.boumans@radboudumc.nl

Abstract

Background: Society is facing a global shortage of 17 million health care workers, along with increasing health care demands from a growing number of older adults. Social robots are being considered as solutions to part of this problem.

Objective: Our objective is to evaluate the quality of care perceived by patients and caregivers for an integrated care pathway in an outpatient clinic using a social robot for patient-reported outcome measure (PROM) interviews versus the currently used professional interviews.

Methods: A multicenter, two-parallel-group, nonblinded, randomized controlled trial was used to test for noninferiority of the quality of care delivered through robot-assisted care. The randomization was performed using a computer-generated table. The setting consisted of two outpatient clinics, and the study took place from July to December 2019. Of 419 patients who visited the participating outpatient clinics, 110 older patients met the criteria for recruitment. Inclusion criteria were the ability to speak and read Dutch and being assisted by a participating health care professional. Exclusion criteria were serious hearing or vision problems, serious cognitive problems, and paranoia or similar psychiatric problems. The intervention consisted of a social robot conducting a 36-item PROM. As the main outcome measure, the customized Consumer Quality Index (CQI) was used, as reported by patients and caregivers for the outpatient pathway of care.

Results: In total, 75 intermediately frail older patients were included in the study, randomly assigned to the intervention and control groups, and processed: 36 female (48%) and 39 male (52%); mean age 77.4 years (SD 7.3), range 60-91 years. There was no significant difference in the total patient CQI scores between the patients included in the robot-assisted care pathway (mean 9.27, SD 0.65, n=37) and those in the control group (mean 9.00, SD 0.70, n=38): P=.08, 95% CI –0.04 to 0.58. There was no significant difference in the total CQI scores between caregivers in the intervention group (mean 9.21, SD 0.76, n=30) and those in the control group (mean 9.09, SD 0.60, n=35): P=.47, 95% CI –0.21 to 0.46. No harm or unintended effects occurred.

Conclusions: Geriatric patients and their informal caregivers valued robot-assisted and nonrobot-assisted care pathways equally.

Trial Registration: ClinicalTrials.gov NCT03857789; https://clinicaltrials.gov/ct2/show/NCT03857789

(J Med Internet Res 2020;22(9):e18787) doi:10.2196/18787

KEYWORDS

integrated care pathway; social robot; quality of care; noninferiority randomized controlled trial

Introduction

In 2019, society was facing a global shortage of 17 million health care workers [1], along with increasing health care demands from a growing number of older adults [2]. Social robots are being considered as solutions to part of this problem [3,4]. For example, social robots—humanoid robots that are capable of social interaction with humans [5]—might be able to support professionals in hospital-implemented integrated care pathways [6].

Such pathways are already being used to optimize workforce use and cost-effectiveness by delivering health care for a well-defined group of patients during a well-defined period [7]. The overall aim of a care pathway is to enhance the quality of care by improving patient outcomes, promoting patient safety, increasing patient satisfaction, and optimizing the use of resources [7]. Pathways also make it possible to standardize certain parts of communication with patients (eg, for information on the process of care and questionnaires needed to assess outcomes) [6]. A care pathway can be visualized in the form of a time diagram (see Multimedia Appendix 1, Figure MA1-1) depicting the aims of the pathway steps and the responsible health care professionals who interact with the patient. Although all of these dialogues are important, not all may require the actual presence of health care professionals. Some could be carried out by social robots, under the supervision of health care professionals.

Many studies have been conducted on assistive robots for health care professionals [8], as well as on the cost-effectiveness of care pathways. We focus on health care robots that perform a verbal health care–related interaction with patients. For example, Di Nuovo et al used the social robot Pepper to study the assessment of cognitive skills of university personnel with the Montreal Cognitive Assessment (MoCA) [9-11]. Bandera et al designed CLARC (CLinical Assistant Robot for Comprehensive geriatric assessment), a robot designed to perform a comprehensive geriatric assessment, but have not yet published results on its interviewing performance [12-14]. Broadbent et al used a robot to provide at-home assistance to people with chronic obstructive pulmonary disease. This robot spoke but could not listen; patients entered their responses on a touch screen [15]. D'Onofrio et al describe the MARIO (Managing active and healthy Aging with use of caRing servIce rObots) robot that was designed for the practical daily living support of people with dementia in nursing homes, focusing on differences in feasibility between the United Kingdom, Ireland, and Sweden [16]. An evaluation of a social robot conducting interviews using medical questions with community-dwelling older adults has been described in Boumans et al [17]. In a crossover study, 31 participants were subjected to a question-and-answer dialogue with the robot that included personalization and affective statements. Participants scored the robot's subjective usability, on average, as 80.1 (SD 11.6) on a scale from 0 to 100. Subsequently, they performed an ecological validation on the agreement of data collected by automated acquisition for three complete patient-reported outcome measures (PROMs), also among community-dwelling older adults. Data acquisition by a humanoid robot was compared to acquisition by a nurse in a crossover study. The conclusion was that a moderate-to-substantial agreement could be demonstrated between the frailty, well-being, and resilience scores [18]. The Lio robot (F&P Robotics) is appreciated as a support to older adults in care homes for functions such as handing over physical objects and support in performing exercises, but is not used for medical interviewing [19]. The same is true for the Care-O-bot 4 robot (Fraunhofer Institute) [20]; however, the development of this robot has been reported to be discontinued [19]. To our knowledge, however, no studies have been conducted on the quality of care, acceptance, and efficiency of social robots as an integrated part of care pathways in an outpatient clinic.

This study is, thus, the first to evaluate robot interaction with older patients within the outpatient clinic context. The older patient population was chosen, as their consultations often take more time and are more complex—due to sensory and cognitive impairments—than those for younger and less complex patient groups. Our target group thus allows substantial room for robot-assisted support.

Our hypothesis is that the quality of care perceived by patients and caregivers in a pathway that includes a social robot for a standardized part of health care professional–patient dialogue is not significantly lower than that perceived by the control group, whose pathway involves the continued presence of health care professionals; this is a noninferiority hypothesis. Perceived quality of care can be measured validly and reliably using the Consumer Quality Index (CQI) [21], which has been used to monitor the quality of outpatient clinics in all Dutch hospitals [22].

Methods

Study Design

The study was designed as a between-subjects, multicenter, randomized controlled trial among patients visiting the outpatient memory clinics at two teaching hospitals: Radboud university medical center and Canisius Wilhelmina Ziekenhuis. The study was conducted between July and December 2019. The care pathways of both clinics consisted of six steps: a welcome, a physical examination, an interview using a PROM and a frailty questionnaire, a discussion of the results, a discussion on any other relevant medical issues, and a farewell (see Multimedia Appendix 1). We selected a care pathway describing older patients' repeated outpatient visits to control for safe and effective use of medications, such as cholinesterase inhibitors in patients with early-stage dementia. In the intervention pathway, the PROM and frailty questionnaires were administered by the robot, with all other actions performed

by the health care professional. In the nonintervention pathway, all tasks were performed by the health care professionals. The Older Patients and Informal Caregiver Survey – Short Form (TOPICS-SF) was used as the PROM and frailty questionnaire. It consists of 36 questions on general health outcome measures: pain and discomfort, memory, activities of daily living, feelings, social activities, and current diseases [23]. The questionnaire results are used to generate a Frailty Index (FI), which is calculated as the summation of the values associated with each answer, divided by the total of answered questions. The feasibility, validity, and reliability of the instrument as a frailty questionnaire has been established in previous studies [24,25]. It has also been validated as a PROM [23]. The TOPICS-SF is currently accepted by the Dutch Geriatrics Society as a PROM for older patients throughout the Netherlands, and it is being implemented within several hospitals throughout the country [25]. The TOPICS-SF is included in Multimedia Appendix 2.

Patient Population

Patients were recruited from the group of patients scheduled to visit the outpatient clinics of the geriatrics departments of both hospitals. These outpatient clinics subsequently welcomed a total of 419 patients during the study. Inclusion criteria were the ability to speak and read Dutch and being assisted by one of the regular staff nurses or physicians taking part in the study. Exclusion criteria were serious hearing or vision problems, serious cognitive problems, and paranoia or similar psychiatric problems, all as judged by the health care professional, as well as situations in which the patient had previously been asked to complete the TOPICS-SF. The patient population for this noninferiority trial was similar to the population that would be included in a trial for establishing the efficacy of social robots. Patients were selected by their responsible health care professionals, based on the inclusion criteria, upon reviewing the patient visits scheduled in the electronic health record (EHR) system. Patients were screened for exclusion criteria and consent was requested, all according to a standardized script.

Public and Patient Involvement

Patients were involved in the study as subjects; the public was involved in the study through the patients' accompanying informal caregivers and patient organization representatives. The study hypothesis explicitly refers to the measurement of patients' opinions by using the CQI. The public has also been involved in the study design through the preceding studies among community-dwelling older adults [26,27] and through advice given from patient organization representatives during pilot tests. The minimization of the burden on, and time required of, the patients was an important criterion in the study design.

Randomization

Patients were randomized using a computer-generated list and assigned to either the intervention or the control group, in sequence of admission. The nature of the intervention prevented the blinding of group allocation, and data acquisition could not be blinded from the patient perspective, given that the data were self-reported.

Study Procedure

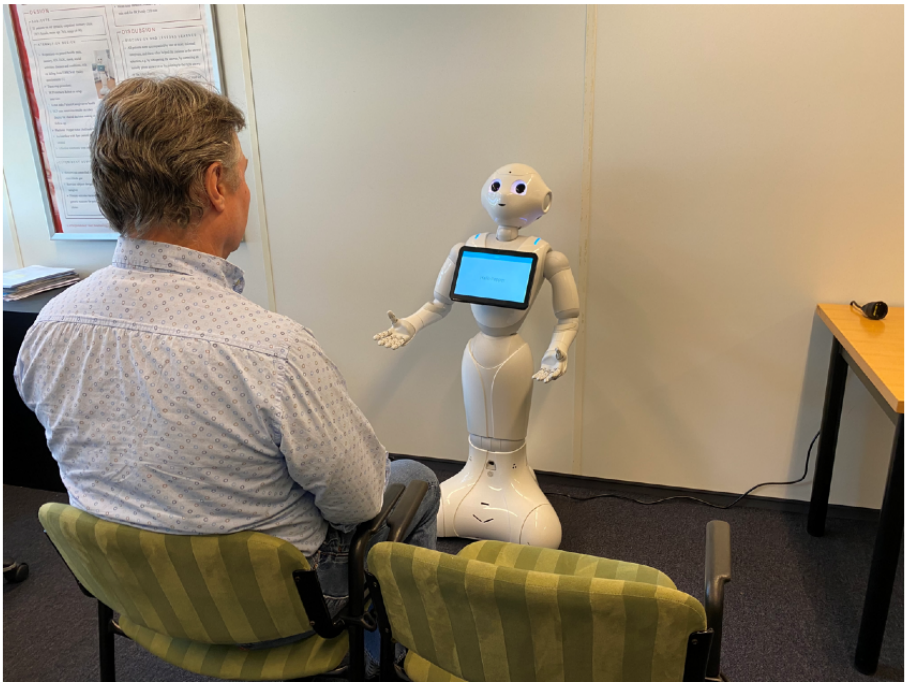
The health care professional guided the patient from the waiting room to an examination room, where the robot was or was not present, depending on randomization.

In the intervention pathway, the health care professional started the interview with several open-ended questions on the patient's general health status. This was followed by the introduction of the robot, which subsequently conducted the TOPICS-SF interview. Upon completing the interview, the robot generated a report of the PROM and FI results, including the activities of daily living and the instrumental activities of daily living scores. This report was the input for subsequent interactions between the patient and the health care professional within the context of shared decision making [28]. The robot-patient interaction is detailed in Multimedia Appendix 3. The interview setup is shown in Figure 1. A video of the interaction is shown in Multimedia Appendix 4.

In the control group, following the initial general talk, the health care professional started the structured TOPICS-SF. The results were discussed with the patient, and the other parts of the medical examination and management plan were carried out.

If needed, these steps were followed by other medical procedures that had been scheduled for the patient's care pathway (eg, blood samples, electrocardiogram, and MoCA) [8]. In both scenarios, if there were no more medical issues to handle, a research assistant asked both the patient and the caregiver to complete the CQI questionnaire. After the CQI questionnaire was checked for completeness, the health care professional completed the visit and said farewell.

Figure 1. The robot-patient interview setup. The patient sits on the right (not shown) and the informal caregiver sits on the left (person shown was not part of the study population). The distance between the patient and the robot is 90 cm.



Human-Robot Interaction Design

The social robot used in this study was a Pepper robot, version 1.8a, using the NAOqi operating system, version 3.9 (SoftBank Robotics) [10]. The robot software necessary for the intervention was designed and programmed using Android Studio, version 3.1 (Google Inc), and Java, version 8 (Oracle Corp). The software managing the dialogue included rules for introducing question groups, for providing variability in how similar questions were asked, and for generating affective and connecting statements. Answers were stored directly in the hospital's EHR system. Ethical design considerations were taken into account by incorporating the fundamentals of care [29] into the communication design. For example, for each question, the default answer set was divided into two groups: (1) answers indicating serious conditions, which could possibly invoke empathy on the part of a health care professional, and (2) answers indicating minor conditions, which would not require separate discussion. The robot looked mostly at the patient and sometimes at the caregiver, in order to create engagement with both. The robot's tablet display was used to show each question and the associated answer options. The layout of the interaction design was based on guidelines for older adults [26]. After hearing the patient's answer, the robot repeated it and showed it on the display, then proceeded to the next question. More details are provided in Multimedia Appendix 3.

Training Health Care Professionals

For this experiment, secretarial staff members were trained in using a telephone script and a list of answers to frequently asked questions about the robot. These answers were used in the event that patients or caregivers called with questions. Health care professionals were trained in how to start the robot, interact with the EHR system through the robot, initiate the questioning, and use the questionnaire report on the tablet.

Primary Outcome

The most relevant part of the validated, general medical CQI questionnaire for outpatient clinics was selected as the primary outcome measure [27,30]. Most of the list items were not applicable to our study and in the attempt to minimize the burden to the patients, the 10 most relevant questions were selected in advance (see Multimedia Appendix 5, Table MA5-1). This selection was done in line with recommendations for shortening the CQI questionnaires [22,31]. Furthermore, the subscales regarding the clinic and the treatment by the health care professional showed Cronbach α values of .845 and .880, respectively, thus indicating a high degree of correlation in the subscales [32]. Therefore, we considered our selection of relevant questions as allowed. Answers were evaluated for the scale as a whole, for the two subscales (ie, regarding the clinic and regarding the robot-supported health care professional), and individually.

Answers to the CQI questions are generally scored categorically, including *no*, *not at all*; *a little*; *largely*; and *yes*, *completely*. The granularity of this scale is small, however, and pilot evaluations revealed ceiling effects and skewed distributions. The patients were, therefore, asked to assign scores on a scale from 1 to 10, with references to these categories (see Figure 2).

Figure 2. An example of one of the 10 Consumer Quality Index (CQI) questions; this is presented in its 10-point scale version.

Did you feel welcome at the outpatient clinic? (Please encircle the number)

No, not at all		A little			Largely			Yes, completely	
1	2	3	4	5	6	7	8	9	10

The opinion of the informal caregiver accompanying the patient was also recorded using the same questions, albeit reformulated for the informal caregiver's perspective. The answers to each CQI question were averaged across all patients and caregivers in each group. The primary outcome was then calculated as the mean sum of the individual question outcomes. The same method was used for the two aforementioned subscales.

Secondary Outcomes

The time duration of the TOPICS-SF interview was registered as a secondary outcome by observers who witnessed each interview. These observers further used an observation form to record, for each question, the extent to which the patient and caregiver exchanged information on the TOPICS-SF answers (see Multimedia Appendix 6, Figure MA6-1). Other potentially relevant events were also recorded (eg, patient remarks on the interaction). The observers were instructed not to intervene at all. Given that such self-recording of secondary outcomes could not be blinded, observation bias was limited by using alternating trained observers. The general medical situation of the patient group was categorized according to the mean FI as follows: *robust* (FI ≤ 0.095), *prefrail* (0.095 < FI < 0.20), and *frail* (FI ≥ 0.20) [33]. The total number of reported comorbidities per patient was calculated, resulting in a value between 0 and 18.

In the intervention group, four questions based on the Almere model [34] were asked to evaluate the usability of the robot (see Multimedia Appendix 7, Table MA7-1). This made it possible to compare these results to our previous work [17,18]. To limit patient burden, survey questions were restricted to three variables: *perceived ease of use* (two items), *perceived enjoyment*, and *trust* [34].

Sample Size Calculation

In our two previous robot studies, which were conducted with 30 and 40 community-dwelling older volunteers, respectively, we found hardly any difference between the answers given to the robot and those given to the health care professional [17,18]. In this study, therefore, we focused on the quality of care perceived by patients and caregivers, hypothesizing that the robot interview would also not be valued less by the intervention group. For this reason, a noninferiority, sample size calculation was applied, specifying that the mean CQI of the intervention group should not be lower than the mean CQI of the control group minus 1.0, with a standard deviation of 1.5, α=.05, and power=1–β=.90 [35]. The difference value of 1.0 is based on the guideline proposed by Ringash et al, which defines 10% of the PROM scale range as a meaningful difference [36]. This calculation resulted in a sample size of 39 patients per group (78 in total).

Statistical Analysis

Data were stored in Castor, a cloud-based medical data management system (Castor EDC). Intention-to-treat analysis was performed using SPSS Statistics for Windows, version 25.0 (IBM Corp), and Microsoft Excel (Office 365, Microsoft). Because not all data were reported by patients or caregivers, the number of patients to which variables relate are reported separately. Missing values were not considered random and, thus, not imputed. Normally distributed values are presented as means, with standard deviations in parentheses. Because the target sample size was larger than 25, we applied the central limit theorem and assumed normality on the part of the summed score for the CQI questionnaire. Groups were compared using independent-samples *t* tests and, in case of nonnormality, the Mann-Whitney U test. For significant effects or effect trends, effect sizes were calculated as Cohen *d*.

Ethical Considerations

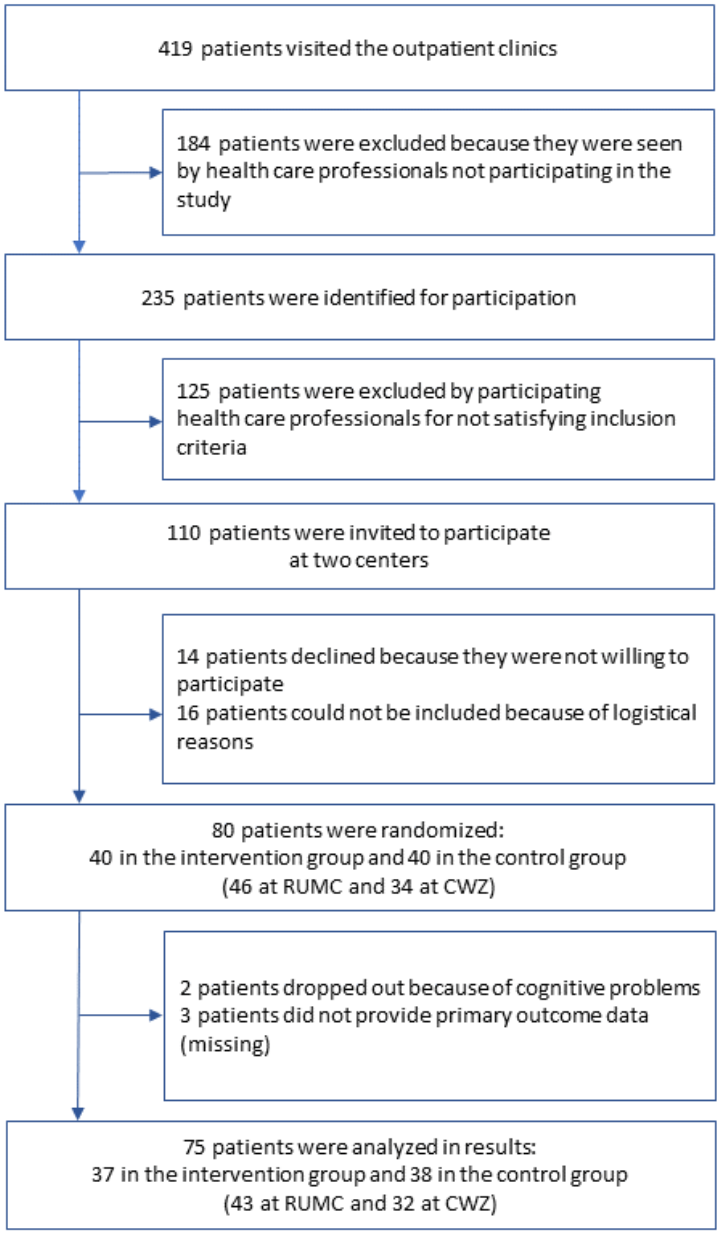
The study was conducted according to the principles of the Declaration of Helsinki (2013), in accordance with the Medical Research Involving Human Subjects Act (*Wet medisch-wetenschappelijk onderzoek met mensen* [WMO] in Dutch) and the CONSORT (Consolidated Standards of Reporting Trials) guidelines for randomized controlled trials, including the extension for noninferiority trials [37]. The study protocol was approved by the Institutional Review Board from each hospital. All patients granted written informed consent. Caregivers had the option to grant consent on behalf of their relatives, but this situation did not occur. This trial was registered at ClinicalTrials.gov (NCT03857789).

Results

Patient Population

The patient flowchart is provided in Figure 3. Recruitment was stopped upon reaching 80 included patients. However, 2 patients dropped out during the experiment after randomization: 1 patient turned out to have cognitive problems that made it impossible to complete the robot interaction, and 1 patient chose to discontinue the interview with the robot after nine questions because "she did not like the robot." Another 3 patients were lost to follow-up because of the unavailability of their CQI ratings. Therefore, the dataset used consisted of 75 patients: 36 female (48%) and 39 male (52%); mean age 77.4 years (SD 7.3), range 60-91 years. Of the 75 patients, 37 were in the intervention group (49%) and 38 were in the control group (51%).

Figure 3. Patient flowchart. CWZ: Canisius Wilhelmina Ziekenhuis; RUMC: Radboud university medical center.



All 75 patients were accompanied by an informal caregiver: 34 were partners of a patient (45%), 23 were children of a patient (31%), and 1 was a friend of a patient (1%); 8 informal caregivers had other affiliations (11%) and 9 did not disclose their relationship to the patient (12%).

None of the 14 patients (see Figure 3) who declined the invitation due to unwillingness to participate mentioned the robot as the reason (14/75, 19%).

The consultations were conducted by 13 different health care professionals. The patient-robot interactions were observed by 11 different trained observers. No important incidents of harm or unintended effects were observed or reported.

The FI for the group as a whole ranged from 0.07 to 0.68 (mean 0.26, SD 0.15). The mean FI for the control group (mean 0.26, SD 0.15) and the intervention group (mean 0.25, SD 0.15) were similar (P=.99). Out of 75 patients, 4 (5%) patients could be categorized as robust, 30 (40%) as prefrail, and 36 (48%) as frail; in addition, 21 patients (28%) had been diagnosed with dementia. The average number of comorbidities per patient was 3.9 (SD 2.6). The main patient baseline clinical data for each group are included in Table 1; extended data are provided in Multimedia Appendix 8.

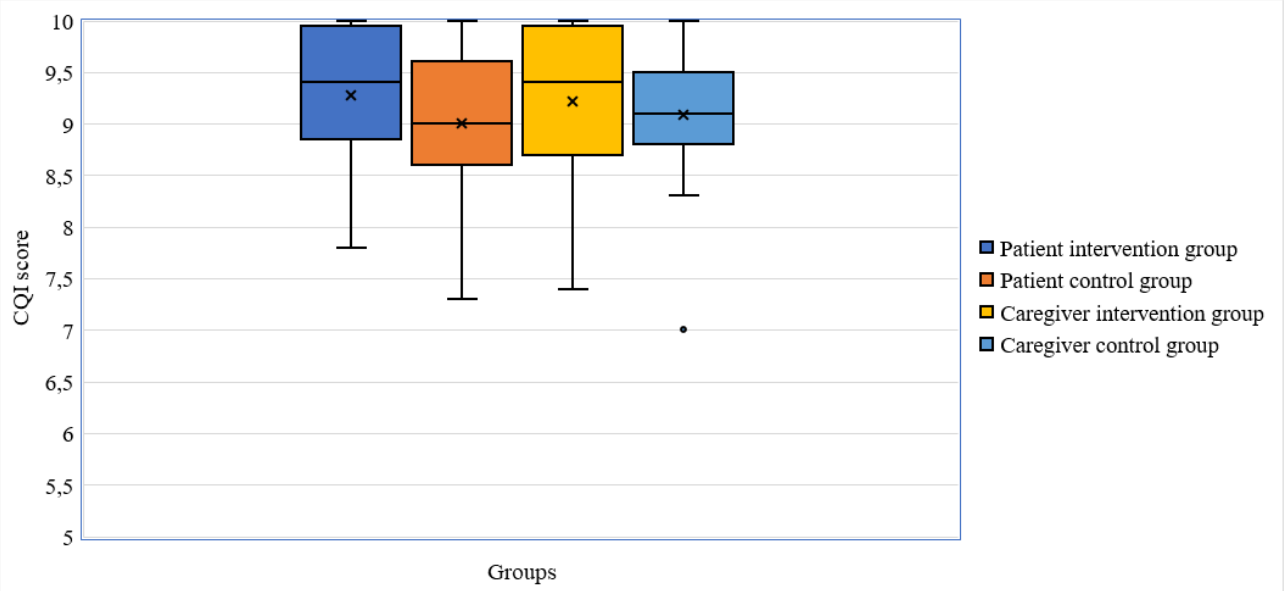
Table 1. Baseline characteristics of the study population (N=75).

Characteristic	Intervention group (n=37)	Control group (n=38)
Sex (female), n (%)	16 (43)	20 (53)
Age (years), mean (SD)	78.1 (7.0)	76.7 (7.7)
Self-indicated quality-of-life score (0-10), mean (SD)	7.5 (1.9)	7.1 (1.6)
Frailty Index (0-1), mean (SD)	0.25 (0.15)	0.26 (0.15)
Frailty value, n (%)		
Robust	3 (8)	1 (3)
Prefrail	13 (35)	17 (45)
Frail	21 (57)	15 (39)
Comorbidities, n (%)		
Memory complaints	19 (51)	26 (68)
Pain		
None	11 (30)	14 (37)
A little	12 (32)	9 (24)
Moderate	8 (22)	10 (26)
Severe	6 (16)	3 (8)
Extreme	0 (0)	1 (3)
Dementia	11 (30)	10 (26)
Hearing problems	9 (24)	8 (21)
Vision problems	10 (27)	3 (8)

Primary Outcome

The total CQI scores recorded for patients and caregivers are presented graphically in Figure 4. There was no significant difference in the total patient CQI scores for the intervention group (mean 9.27, SD 0.65) and the control group (mean 9.00, SD 0.70) (t_{73}=1.76, *P*=.08, 95% CI –0.04 to 0.58). There was also no significant difference in the total informal caregiver CQI scores for the intervention group (mean 9.21, SD 0.76) and the control group (mean 9.09, SD 0.60) (t_{63}=0.73, *P*=.47, 95% CI –0.21 to 0.46).

Figure 4. Box plots for total Consumer Quality Index (CQI) scores for patients (two box plots on the left) and caregivers (two box plots on the right).



A *t* test on each of the 10 individual CQI questions (see Multimedia Appendix 9) revealed that patients found that health care professionals, when supported by a robot, listened better (mean 9.46, SD 0.69) than health care professionals not supported by the robot (mean 9.11, SD 0.76) (t_{73}=2.104, *P*=.04, 95% CI 0.019-0.690; Cohen *d*=0.48). Patients also found that

health care professionals, when supported by the robot, had more time for the patient (mean 9.54, SD 0.56) compared to those not being supported by the robot (mean 9.13, SD 0.70) (t_{73}=2.784, *P*=.007, 95% CI 0.116-0.702; Cohen *d*=0.64). The other eight questions, individually, did not reveal any significant differences. A *t* test on the group of questions about the care provided by the health care professional (see Multimedia Appendix 9, questions 4-8) showed that patients found that health care professionals supported by the robot provided better care (mean 9.42, SD 0.62) than health care professionals not supported by the robot (mean 9.11, SD 0.69) (t_{73}=2.086, *P*=.04, 95% CI 0.014-0.619; Cohen *d*=0.48). The patients' answers to the group of questions about the clinic (see Multimedia Appendix 9, questions 1-3, 9, and 10) did not show significant differences. Regarding informal caregivers accompanying the patients, there were no significant differences found between health care professionals supported by a robot or not, nor between a clinic using a robot or not. The CQI scores for all questions are included in Multimedia Appendix 9, and the total CQI distributions are presented in Multimedia Appendix 10.

Secondary Outcomes

Within the care pathways, the mean duration for completing the TOPICS-SF with the robot was 17.9 minutes (SD 5.2), as compared to 14.8 minutes (SD 10.8) for the control group. The difference was not significant: t_{70}=1.60, *P*=.11, 95% CI –0.79 to 7.18. It should be noted that observations showed that health care professionals regularly skipped questions.

It was observed that patients and caregivers did not discuss the TOPICS-SF answer options any more during the interviews with the robot (mean 3.5, SD 3.8) than was the case in the control group (mean 2.9, SD 2.5): t_{53}=0.58, *P*=.56, 95% CI –1.32 to 2.42. It was further observed that, at the start of the interview, patients sometimes answered before the robot was finished speaking. This well-known barge-in effect occurred despite the fact that the robot had instructed patients to wait for the blue bar to appear at the top of the tablet before speaking [38]. Most patients learned after three or four questions that it was better to wait a short while before answering, as they would otherwise have to repeat their answers. Informal caregivers occasionally helped the patients when necessary (10% of the questions); for example, because one patient spoke a local Dutch dialect that was not understood by the robot, the patient's caregiver answered instead.

For the intervention group only, the mean scores for *perceived enjoyment*, *perceived ease of use* (2 items), and *trust* with regard to the robot interaction were recorded (see Multimedia Appendix 11). There was no significant difference in *perceived enjoyment* between patients (mean 7.81, SD 2.01) and caregivers (mean 7.56, SD 2.11): t_{56}=0.47, *P*=.64, 95% CI –0.85 to 1.37. In addition, there was no significant difference in *perceived ease of use* in terms of having sufficient response time between patients (mean 8.51, SD 1.63) and caregivers (mean 8.45, SD 1.10): t_{55}=0.15, *P*=.88, 95% CI –0.73 to 0.85. There was also no significant difference in *perceived ease of use* in terms of easy answering between patients (mean 8.11, SD 1.89) and caregivers (mean 7.86, SD 1.67): t_{56}=0.50, *P*=.62, 95% CI –0.74 to 1.23. Trust scores were higher for patients (mean 8.42, SD 1.38) than for caregivers (mean 7.59, SD 1.76): t_{55}=2.00, *P*=.05, 95% CI <–0.001 to 1.68; Cohen *d*=0.55. Of the 36 caregivers in the intervention group who answered the CQI questions, only 24 (67%) also answered the questions on robot appreciation. The caregivers who did not answer argued that it was better for the patients to answer themselves, as they had been the ones to talk to the robot.

Discussion

Principal Findings

To our knowledge, this study is the first to provide an assessment of patients' perceived quality of care in integrated care pathways with and without the support of social robots. We found that the perceptions of older patients and caregivers concerning quality of care were no different from the perceptions of quality of care in a pathway in which all interactions were carried out by health care professionals. This confirmed our hypothesis of noninferiority. The opinions of the patients and caregivers concerning the robot were in line with previous findings regarding the positive appreciation results on robot interaction among community-dwelling older adults [26,27], as well as with the results reported in our exploratory study among hospitalized patients [39].

Older adult patients participating in this study who had been diagnosed with dementia (11/37, 30%) were still able to answer the questions asked by the robot. The preselection of participants by the health care professionals probably resulted in a group with mild-to-moderate cognitive problems, who were still able to communicate verbally with either the health care professional or the robot. It was also observed that patients with auditive (9/37, 24%) or visual (10/37, 27%) problems were capable of completing the interview. This indicates that the design measures taken to improve robot communication (ie, quiet environment, adjusted voice volume and speed, font size of text on the tablet, and minimalistic layout) were adequate. When they deemed it necessary, informal caregivers assisted patients; this occurred for 10% of the questions.

The observers noted that, in the control group, health care professionals regularly skipped questions from the TOPICS-SF. When asked about this, the professionals responded that they had skipped questions to which they already knew the answers or that they considered inappropriate to ask explicitly. The robot always asked all of the questions. This could be a potential advantage, as it ensures that no items will be missed inadvertently.

Strengths and Weaknesses of the Study

The major strength of this study is that this is the first multicenter, randomized controlled trial on the acquisition of routine, collected PROM data with a social robot among older adult patients within an integrated care pathway. The noninferiority results of this trial suggest that an adequately designed social robot could be acceptable for use with older adult patients and their informal caregivers as part of an integrated care pathway, under the indirect supervision of a health care professional.

Despite this strength, this study is also subject to several limitations. First, after analysis, it turned out that the planned sample size was not met because of 2 dropouts and 3 participants with missing data, which was more than our margin of 2 patients. However, by imputing the dataset with 2 intervention group patients with scores of mean–2σ and 1 control group patient with a score of mean+2σ, which was considered as the worst case scenario, it was found that this did not affect the conclusion of a nonsignificant difference in perceived quality of the care pathway. Secondly, it was not possible to blind the assignment of patients to groups. Thirdly, the between-subjects design did not allow any comparative-accuracy analyses of the answers. In our previous study, however, the results indicated moderate-to-good agreement between scores with and without the robot [18].

Comparison With Prior Work

The results confirm and extend those of previous studies on the use of robots outside the hospital context [10,12,15,16]. For example, Olde Keizer et al concluded that social robots could potentially monitor and train the health of frail older adults, but they also identified some critical usability challenges [40]. Furthermore, the functionality of the Lio robot (F&P Robotics), given its reported voice communication capabilities, could be extended with verbal interviewing functions as described herein.

Riek has provided a comprehensive overview of robot applications in health care with many examples of physical support [8]. This study adds to the knowledge base a multicenter, randomized controlled trial examining the verbal support option of a robot interviewing older adult patients in an outpatient clinic regarding their health and, as such, resolves part of the paucity in effective clinical trials that Riek noted [8].

Meaning of the Study

In terms of generalizability, the patient group in this study was more frail and had more substantial multimorbidity than is the case for the general hospital population. Communication with the robot could possibly be even easier for the general hospital population. For this reason, and because the TOPICS-SF is similar to many available PROMs, it is plausible that the results can be generalized to most adults admitted to hospitals, as well as to most care pathways. The results thus suggest that robot assistance could be implemented more broadly without affecting perceived quality of care.

The observations and experiences gained in this experiment could also be translated into a number of recommendations. First, the introduction of a social robot should lead to a carefully prepared rearrangement of tasks among the health care professionals within a pathway of care. Second, for reasons of patient privacy and the intelligibility of the patient's utterances to the robot, the robot should be a fixed element in an outpatient room. Third, participating health care professionals appreciated the direct availability of all collected data in the EHR system. Therefore, we recommend implementing real-time data export from the robot to the hospital's EHR system for successful implementation. Fourth, technologies like these may support clinical care during pandemics, since they limit person-to-person contact and allow for social distancing.

Our findings suggest that this social-robot technology could be implemented more broadly for obtaining PROM data, as well as for other standardized parts of functional assessments and medical history taking. The assistance of social robots could, thus, potentially contribute to reducing problems related to the scarcity of health care personnel, while maintaining the quality of care, as perceived by patients and caregivers.

Unanswered Questions and Future Research

In the course of our study, we learned that one important further step in improving robot technology involves developing the ability to speak and listen at the same time, thus allowing for *barging-in* by patients. Although such technology does exist, it was not implemented in the robot used in this study. Moreover, the quality of the robot's speech recognition depended on its focus on the interlocutor, which was controlled by the built-in *human engagement* function. Improving the controllability of this function, in terms of both speech and body motions, would help to build rapport with users.

Acknowledgments

The authors would like to thank the following people for their valuable contributions to this research (in alphabetical order): Anoek Aben, Lisa van Beelen, Monique Boon, Jurgen Claassen, Nicole van der Cruijsen, Guido van der Hart, Rianne de Heus, Christa Hummelen, Trudy Jacobs, Ruud de Jong, Madeleine Koolen, Renée Koster, Liesbeth Langenhuysen, Lidwine Luijkx, Nicky Mols, Pascal van Nispen, Nadja Ommering-van den Elzen, Dorien Oostra, Romy van Os, Renke van Rijckevorsel, Yvonne Schoon, Lies Smits, Sevilay Tokgöz, Maarten van der Ven, and Jeroen Wildenbeest. This work has been funded as a research partnership project between the Delft University of Technology, Delft, Netherlands, and the Radboud university medical center, Nijmegen, Netherlands.

Authors' Contributions

RB was responsible for conceptualization of the study; design, development, and coding of the software; designing the methodology; conducting the formal analysis; conducting the investigation; and writing the original draft. FvM was responsible for writing, reviewing, and editing the manuscript. WvA was responsible for conducting the investigation and for writing, reviewing, and editing the manuscript. JA was responsible for conducting the investigation, conducting the formal analysis, and writing the original draft. MJ was responsible for conducting the investigation, conducting the formal analysis, and writing the original draft. MPK was responsible for conducting the investigation, conducting the formal analysis, and writing the original draft. GHdW was responsible for conceptualization of the study and for writing, reviewing, and editing the manuscript. AvdP was responsible

for funding acquisition, conceptualization of the study, securing resources, and providing supervision at Canisius Wilhelmina Ziekenhuis. KH was responsible for conceptualization of the study and for writing, reviewing, and editing the manuscript. MN was responsible for funding acquisition, conceptualization of the study, and writing, reviewing, and editing the manuscript. MOR was responsible for funding acquisition; conceptualization of the study; securing resources; writing, reviewing, and editing the manuscript; and providing overall supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Care pathway treatment.
[DOCX File , 58 KB - jmir_v22i9e18787_app1.docx]

Multimedia Appendix 2
The Older Patients and Informal Caregiver Survey – Short Form (TOPICS-SF).
[DOCX File , 17 KB - jmir_v22i9e18787_app2.docx]

Multimedia Appendix 3
Detailed description of the robot-patient interaction.
[DOCX File , 45 KB - jmir_v22i9e18787_app3.docx]

Multimedia Appendix 4
Video of a short part of the human-robot interaction.
[MP4 File (MP4 Video), 11189 KB - jmir_v22i9e18787_app4.mp4]

Multimedia Appendix 5
Consumer Quality Index (CQI) questions.
[DOCX File , 13 KB - jmir_v22i9e18787_app5.docx]

Multimedia Appendix 6
Observation form.
[DOCX File , 13 KB - jmir_v22i9e18787_app6.docx]

Multimedia Appendix 7
Robot-usability questions.
[DOCX File , 13 KB - jmir_v22i9e18787_app7.docx]

Multimedia Appendix 8
Baseline data tables.
[DOCX File , 16 KB - jmir_v22i9e18787_app8.docx]

Multimedia Appendix 9
Consumer Quality Index (CQI) scores by question.
[DOCX File , 15 KB - jmir_v22i9e18787_app9.docx]

Multimedia Appendix 10
Consumer Quality Index (CQI) distributions.
[DOCX File , 76 KB - jmir_v22i9e18787_app10.docx]

Multimedia Appendix 11
Patient and caregiver opinions regarding robot usability.
[DOCX File , 14 KB - jmir_v22i9e18787_app11.docx]

Multimedia Appendix 12
CONSORT-EHEALTH checklsit (V 1.6.1).
[PDF File (Adobe PDF File), 1638 KB - jmir_v22i9e18787_app12.pdf]

References

1. Buchan J, Campbell J, Dhillon I, Charlesworth A. Labour market change and the international mobility of health workers. The Health Foundation. 2019 Mar. URL: https://www.health.org.uk/sites/default/files/upload/publications/2019/Labour%20market%20change%20and%20the%20international%20mobility%20of%20health%20workers%20Working%20Paper_0.pdf [accessed 2019-11-25]
2. United Nations, Department of Economic and Social Affairs. World Population Prospects 2019: Highlights. New York, NY: United Nations; 2019. URL: https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf [accessed 2019-11-25]
3. Archibald MM, Barnard A. Futurism in nursing: Technology, robotics and the fundamentals of care. J Clin Nurs 2018 Jun;27(11-12):2473-2480. [doi: 10.1111/jocn.14081] [Medline: 28940657]
4. Institute of Medicine. Future of Nursing: Leading Change, Advancing Health. Washington, DC: The National Academies Press; 2011.
5. Breazeal C. Toward sociable robots. Rob Auton Syst 2003 Mar;42(3-4):167-175. [doi: 10.1016/S0921-8890(02)00373-1]
6. Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. BMJ 1998 Jan 10;316(7125):133-137 [FREE Full text] [doi: 10.1136/bmj.316.7125.133] [Medline: 9462322]
7. Care pathways. European Pathway Association. 2018. URL: http://www.e-p-a.org/care-pathways [accessed 2018-05-09]
8. Riek LD. Healthcare robotics. Commun ACM 2017 Oct 24;60(11):68-78. [doi: 10.1145/3127874]
9. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
10. Di Nuovo A, Varrasi S, Lucas A, Conti D, McNamara J, Soranzo A. Assessment of cognitive skills via human-robot interaction and cloud computing. J Bionic Eng 2019 May 22;16(3):526-539. [doi: 10.1007/s42235-019-0043-2]
11. Pandey AK, Gelin R. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. IEEE Robot Autom Mag 2018 Sep;25(3):40-48 [FREE Full text] [doi: 10.1109/mra.2018.2833157]
12. Bandera A, Bandera J, Bustos P, Calderita L, Dueñas A, Fernández F, et al. CLARC: A robotic architecture for comprehensive geriatric assessment. In: Proceedings of the 17th Workshop of Physical Agents (WAF). 2016 Jun Presented at: 17th Workshop of Physical Agents (WAF); June 16-17, 2016; Malaga, Spain.
13. García-Olaya A, Fuentetaja R, García-Polo J, González J, Fernández F. Challenges on the application of automated planning for comprehensive geriatric assessment using an autonomous social robot. In: Proceedings of the 19th International Workshop of Physical Agents (WAF). 2018 Presented at: 19th International Workshop of Physical Agents (WAF); November 22-23, 2018; Madrid, Spain p. 179-194. [doi: 10.1007/978-3-319-99885-5_13]
14. Ellis G, Langhorne P. Comprehensive geriatric assessment for older hospital patients. Br Med Bull 2004;71:45-59 [FREE Full text] [doi: 10.1093/bmb/ldh033] [Medline: 15684245]
15. Broadbent E, Garrett J, Jepsen N, Li Ogilvie V, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: Pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
16. D'Onofrio G, Sancarlo D, Raciti M, Burke M, Teare A, Kovacic T, et al. MARIO Project: Validation and evidence of service robots for older people with dementia. J Alzheimers Dis 2019;68(4):1587-1601. [doi: 10.3233/JAD-181165] [Medline: 30958360]
17. Boumans R, van Meulen F, Hindriks K, Neerincx M, Olde Rikkert M. A feasibility study of a social robot collecting patient reported outcome measurements from older adults. Int J of Soc Robotics 2019 May 22;12(1):259-266 [FREE Full text] [doi: 10.1007/s12369-019-00561-8]
18. Boumans R, van Meulen F, Hindriks K, Neerincx M, Olde Rikkert MGM. Robot for health data acquisition among older adults: A pilot randomised controlled cross-over trial. BMJ Qual Saf 2019 Oct;28(10):793-799 [FREE Full text] [doi: 10.1136/bmjqs-2018-008977] [Medline: 30894423]
19. Bendel O, Gasser A, Siebenmann J. Co-robots as care robots. In: Proceedings of the Association for the Advancement of Artificial Intelligence 2020 Spring Symposium. 2020 Presented at: Association for the Advancement of Artificial Intelligence 2020 Spring Symposium; March 23-25, 2020; Palo Alto, CA.
20. Kittmann R, Fröhlich T, Schäfer J, Reiser U, Weißhardt F, Haug A. Let me introduce myself: I am Care-O-bot 4, a gentleman robot. In: Pielot M, Diefenbach S, Henze N, editors. Mensch Und Computer 2015 - Tagungsband. Berlin, Germany: De Gruyter; 2015.
21. Zuidgeest M. Measuring and Improving the Quality of Care From the Healthcare User Perspective: The Consumer Quality Index [doctoral thesis]. Tilburg, the Netherlands: Tilburg University; 2011. URL: https://nivel.nl/sites/default/files/bestanden/Proefschrift-Marloes-Zuidgeest.pdf [accessed 2018-11-09]
22. De Boer D, Bos N, Zuidgeest M. Ontwikkelingen in het Meten en Gebuiken van Patiëntervaringen en Patiëntgerapporteerde Uitkomsten (Developments in Measuring and Using Patient Experiences and Patient Reported Outcomes) [document in Dutch]. Utrecht, the Netherlands: Nivel; 2018. URL: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2018/07/25/ontwikkelingen-in-het-meten-en-gebruiken-van-patientervaringen-en-patientgerapporteerde-uitkomsten/

6937+Nivel_Ontwikkelingen+in+meten+en+gebruiken+pati%C3%ABntervaringen_20180724_+OPGEMAAKT.pdf [accessed 2020-08-26]
23. Olde Rikkert MGM, van der Wees PJ, Schoon Y, Westert GP. Using patient reported outcomes measures to promote integrated care. Int J Integr Care 2018 Apr 19;18(2):8 [FREE Full text] [doi: 10.5334/ijic.3961] [Medline: 30127692]
24. Lutomski JE, Baars MAE, Schalk BWM, Boter H, Buurman BM, den Elzen WPJ, TOPICS-MDS Consortium. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): A large-scale data sharing initiative. PLoS One 2013;8(12):e81673 [FREE Full text] [doi: 10.1371/journal.pone.0081673] [Medline: 24324716]
25. Santoso AMM, Lutomski JE, Hofman CS, Metzelthin SF, Blom JW, van der Wees PJ, TOPICS-MDS Consortium. Development of a patient-reported outcome measure for geriatric care: The Older Persons and Informal Caregivers Survey Short Form. Value Health 2018 Oct;21(10):1198-1204 [FREE Full text] [doi: 10.1016/j.jval.2018.02.011] [Medline: 30314621]
26. Czaja SJ, Boot WR, Charness N, Rogers WA. Designing for Older Adults: Principles and Creative Human Factors Approaches. 3rd edition. Boca Raton, FL: CRC Press; 2019.
27. Delnoij DM, Rademakers JJ, Groenewegen PP. The Dutch Consumer Quality Index: An example of stakeholder involvement in indicator development. BMC Health Serv Res 2010 Apr 06;10:88 [FREE Full text] [doi: 10.1186/1472-6963-10-88] [Medline: 20370925]
28. Jamieson T, Goldfarb A. Clinical considerations when applying machine learning to decision-support tasks versus automation. BMJ Qual Saf 2019 Oct;28(10):778-781 [FREE Full text] [doi: 10.1136/bmjqs-2019-009514] [Medline: 31147420]
29. Kitson A, Conroy T, Wengstrom Y, Profetto-McGrath J, Robertson-Malt S. Scholarly paper: Defining the fundamentals of care. Int J Nurs Pract 2010;16(4):434. [doi: 10.1111/j.1440-172x.2010.01861.x]
30. Van Linschoten CP, Barf HA, Moorer P, Spoorenberg S. CQ-index poliklinische ziekenhuiszorg: Meetinstrumentontwikkeling [article in Dutch]. University of Groningen. 2011. URL: https://www.rug.nl/research/portal/nl/publications/cqindex-poliklinische-ziekenhuiszorg(7fe138fc-3d54-44ee-a230-d000e6701368).html [accessed 2018-11-26]
31. Triemstra M, Hendriks M, Delnoij D, Rademakers J. Doelspecifieke versies van CQ-index meetinstrumenten: Korter, krachtiger, en specifieker meten? Tijdschr Gezond Swetenschappen 2008 Aug;86(8):486-493. [doi: 10.1007/bf03082147]
32. Van Plaggenhoef W. Analyserapport: CQI Poliklinische Ziekenhuiszorg 2011. Barneveld, the Netherlands: Significant BV; 2011. URL: http://docplayer.nl/113435337-Analyserapport-cqi-poliklinische-ziekenhuiszorg-miletus-barneveld-2-december-2011-versie-2-0-auteur-s-wijnand-van-plaggenhoef.html [accessed 2020-08-26]
33. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr 2008 Sep 30;8:24 [FREE Full text] [doi: 10.1186/1471-2318-8-24] [Medline: 18826625]
34. Heerink M, Kröse B, Evers V, Wielinga B. Assessing acceptance of assistive social agent technology by older adults: The Almere Model. Int J Soc Robot 2010 Sep 4;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]
35. Power calculator for continuous outcome non-inferiority trial. Sealed Envelope Ltd. 2012. URL: https://www.sealedenvelope.com/power/continuous-noninferior/ [accessed 2019-11-29]
36. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. Cancer 2007 Jul 01;110(1):196-202 [FREE Full text] [doi: 10.1002/cncr.22799] [Medline: 17546575]
37. CONSORT Group. Consolidated Standards of Reporting Trials (CONSORT). 2018. URL: http://www.consort-statement.org/ [accessed 2018-11-30]
38. Heins R, Franzke M, Durian M, Bayya A. Turn-taking as a design principle for barge-in in Spoken Language Systems. Int J Speech Technol 1997 Dec;2(2):155-164. [doi: 10.1007/bf02208827]
39. Van der Putte D, Boumans R, Neerincx M, Olde Rikkert M, de Mul M. A social robot for autonomous health data acquisition among hospitalized patients: An exploratory field study. In: Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). New York, NY: IEEE; 2019 Presented at: 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 11-14, 2019; Daegu, South Korea. [doi: 10.1109/hri.2019.8673280]
40. Olde Keizer RACM, van Velsen L, Moncharmont M, Riche B, Ammour N, Del Signore S, et al. Using socially assistive robots for monitoring and preventing frailty among older adults: A study on usability and user experience challenges. Health Technol 2019 Apr 9;9(4):595-605 [FREE Full text] [doi: 10.1007/s12553-019-00320-9]

Abbreviations

CLARC: CLinical Assistant Robot for Comprehensive geriatric assessment
CONSORT: Consolidated Standards of Reporting Trials
CQI: Consumer Quality Index
EHR: electronic health record
FI: Frailty Index
MARIO: Managing active and healthy Aging with use of caRing servIce rObots
MoCA: Montreal Cognitive Assessment
PROM: patient-reported outcome measure

TOPICS-SF: The Older Patients and Informal Caregiver Survey – Short Form
WMO: Medical Research Involving Human Subjects Act (Wet medisch-wetenschappelijk onderzoek met mensen in Dutch)

Edited by G Eysenbach; submitted 19.03.20; peer-reviewed by T Jamieson, O Korn, O Bendel; comments to author 12.06.20; revised version received 03.07.20; accepted 22.07.20; published 09.09.20.

Please cite as:
Boumans R, van Meulen F, van Aalst W, Albers J, Janssen M, Peters-Kop M, Huisman - de Waal G, van de Poll A, Hindriks K, Neerincx M, Olde Rikkert M
Quality of Care Perceived by Older Patients and Caregivers in Integrated Care Pathways With Interviewing Assistance From a Social Robot: Noninferiority Randomized Controlled Trial
J Med Internet Res 2020;22(9):e18787
URL: http://www.jmir.org/2020/9/e18787/
doi:10.2196/18787
PMID:32902387

Original Paper

Combating Health Care Fraud and Abuse: Conceptualization and Prototyping Study of a Blockchain Antifraud Framework

Tim Ken Mackey[1,2,3,4*], MAS, PhD; Ken Miyachi[2,5,6*], BS; Danny Fung[2], BS; Samson Qian[2], BS; James Short[2], PhD

[1]UC San Diego - School of Medicine, Department of Anesthesiology and Division of Infectious Diseases and Global Public Health, La Jolla, CA, United States

[2]San Diego Supercomputer Center, BlockLAB, La Jolla, CA, United States

[3]Global Health Policy and Data Institute, San Diego, CA, United States

[4]UC San Diego - Extension, Department of Healthcare Research and Policy, La Jolla, CA, United States

[5]LedgerSafe Corporation, San Diego, CA, United States

[6]Institute of Electrical and Electronics Engineers, San Diego, CA, United States

[*]these authors contributed equally

Corresponding Author:
Tim Ken Mackey, MAS, PhD
UC San Diego - School of Medicine
Department of Anesthesiology and Division of Infectious Diseases and Global Public Health
8950 Villa La Jolla Drive
A124
La Jolla, CA, 92037
United States
Phone: 1 9514914161
Email: tmackey@ucsd.edu

Abstract

Background: An estimated US $2.6 billion loss is attributed to health care fraud and abuse. With traditional health care claims verification and reimbursement, the health care provider submits a claim after rendering services to a patient, which is then verified and reimbursed by the payer. However, this process leaves out a critical stakeholder: the patient for whom the services are actually rendered. This lack of patient participation introduces a risk of fraud and abuse. Blockchain technology enables secure data management with transparency, which could mitigate this risk of health care fraud and abuse.

Objective: The aim of this study is to develop a framework using blockchain to record claims data and transactions in an immutable format and to enable the patient to act as a validating node to help detect and prevent health care fraud and abuse.

Methods: We developed a health care fraud and abuse blockchain technical framework and prototype using key blockchain tools and application layers including consensus algorithms, smart contracts, tokens, and governance based on digital identity on the Ethereum platform (Ethereum Foundation).

Results: Our technical framework maps to the claims adjudication process and focuses on Medicare claims, with the US Centers for Medicare and Medicaid Services (CMS) as the central authority. A prototype of the framework system was developed using the blockchain platform Ethereum (Ethereum Foundation), with its design features, workflow, smart contract functions, system architecture, and software implementation outlined. The software stack used to build the system consisted of a front-end user interface framework, a back-end processing server, and a blockchain network. React was used for the user interface framework, and NodeJS and an Express server were used for the back-end processing server; Solidity was the smart contract language used to interact with a local Ethereum blockchain network.

Conclusions: The proposed framework and the initial prototype have the potential to improve the health care claims process by using blockchain technology for secure data storage and consensus mechanisms, which make the claims adjudication process more patient-centric for the purposes of identifying and preventing health care fraud and abuse. Future work will focus on the use of synthetic or historic CMS claims data to assess the real-world viability of the framework.

(J Med Internet Res 2020;22(9):e18623) doi:10.2196/18623

KEYWORDS

fraud; blockchain; medical informatics; delivery of healthcare; Medicare; information science

Introduction

Background

Fraud and abuse is a major financial, legal, and policy challenge in the US $3.5 trillion United States health care system, with the Department of Justice (DOJ) and the Department of Health and Human Services (HHS) reporting recoveries estimated at US $2.6 billion in the fiscal year 2019 alone [1]. In fact, recoveries for health care fraud and abuse have steadily risen in the past 5 years, with settlements consistently exceeding US $2 billion over the past decade [2]. In 2018, the DOJ announced the largest national health care fraud takedown in history, which included over 601 defendants charged across 58 federal districts and involved 165 health care professionals, equating to a total of US $2 billion in false billings, including illegal distribution of opioids and narcotics [3].

Health care fraud and abuse involve all sectors of the health care industry, including drug and device manufacturers, hospitals, pharmacies, physicians, wholesalers, distributors, laboratories, and payers. Arguably, the most significantly impacted group is payers, including public agencies such as Medicare, Medicaid, and Tricare as well as private payers, who are defrauded of billions in health care claims yearly [4,5]. Fraudulent health care occurs in different forms, including kickbacks, false claims (eg, billing for services not rendered, upcoding, and provisioning of medically unnecessary services), and illegal self-referrals [5,6]. Fraud and abuse have a direct negative impact on health care utilization as it leads to a waste of limited resources and potentially endangers patients by providing them unnecessary care or precluding their access to medically needed services, which can lead to a higher risk of all-cause mortality and emergency hospitalization [4,6].

Enforcement against health care fraud and abuse comes in the form of well-established legal mechanisms focused on penalizing such actions, including (1) the False Claims Act, United States Code (USC) section 3729 to 3733; (2) the Antikickback statute, 42 USC section 1320a to 7b(b); (3) the Physician Self-Referral Act (Stark law); (4) the Exclusion Statute; and (5) the Civil Monetary Penalty Law [5,7]. Prosecution for fraud and abuse can lead to civil (monetary) penalties (including triple damages) for each claim or service, and in some cases, criminal penalties, including possible exclusion from federal and state reimbursement programs. These legal frameworks serve as a strong deterrent to fraud and abuse schemes, but detection and prevention remains an ongoing challenge.

Although efforts have been made to automate the detection of fraud and abuse through computational methods involving data mining of Medicare and Medicaid reimbursement claims data sets, most of the fraud and abuse prosecutions continue to originate from whistleblowers [8-10]. Whistleblowers are incentivized to report fraud and abuse activities through *qui tam* provisions that allow private individuals acting as *realtors* to bring a suit on behalf of the government [5,11]. Once a lawsuit is filed, the DOJ then has the option to intervene and join one or all of the counts of a pending qui tam action. If the claim concludes in a prosecution and settlement, then the whistleblower may be entitled to 15% to 30% of what is recovered, a clear incentive for reporting, although *blowing the whistle* may come at a high personal and professional cost [5,11].

This current system of relying on whistleblowers to detect and report fraud and abuse is subject to certain challenges, including court cases, some of which limit the protection for prospective whistleblowers [5,12]. Furthermore, prosecutions based on whistleblowers' reports are not always successful, are often skewed toward prosecution of higher-amount cases, and by nature are reactive and punitive rather than proactive in preventing fraud and abuse. Hence, new technology approaches are needed to enable better resilience, provenance, and verifiability of health care claims that may be susceptible to fraud and abuse, an activity that is aligned with antifraud and program integrity priorities currently being pursued by the US Centers for Medicare and Medicaid Services (CMS).

Objective

One technology with the potential to address these challenges is blockchain, a distributed ledger technology with use cases across several industries, including the energy sector, transportation, finance, and health care [13]. Blockchain use cases in health care are beginning to mature, primarily to improve the governance of health care data and processes [14-16]. One of the primary uses involves improving management; enabling sharing; and improving exchange of patient health data, consumer health data, and genomic data [15,17-20]. This also extends to the use of blockchain for clinical research to improve trial data management and electronic consent [21]. Use cases in other health care sectors are also taking shape, including blockchain for pharmaceutical supply chain challenges (eg, detection of falsified medicines) and integration with medical devices and the internet of things [22,23]. Many of these uses focus on patient-centric approaches to manage and preserve the privacy of health care data with blockchain [24-27].

Importantly, incorporating blockchain into a systems software architecture can enable immutability, consensus, create incentives, and manage external data into a self-executing system with transparent rules across multiple stakeholders [14]. On the basis of these benefits, we proposed a technical framework for a blockchain-based system that includes 3 key stakeholder groups in the health care claims workflow process to enable a more proactive antifraud and abuse system. Although several companies are exploring blockchain to enhance health care reimbursement and revenue cycle management, few have explicitly assessed whether the technology can improve claims verification and better enable the detection and prevention of health care fraud and abuse [28,29]. Hence, this paper will explore the utility of blockchain by developing a fraud and

abuse technical design framework and prototype built on the blockchain environment Ethereum.

Methods

Overview

Blockchain's core utility is as a distributed database of transactions, securely connected in chronological order, enabling efficient and cost-reducing improvements to current systems and business processes [14]. Driving the efficiency of these blockchain processes is the enforcement and execution of rules by software, a shared governance environment, and use of smart contracts to create a more transparent and rule-based system that can tackle issues of trust, such as addressing fraud and abuse. Furthermore, health care information and communication technology systems are increasingly moving toward more *patient-centric* designs, not only for receiving the input from the patient but also for involving the patient in the design and solution implementation process [14,27,30].

Our blockchain technical framework leverages key principles of establishing trust through shared accountability and governance and enables the patient to be a stakeholder in addressing health care fraud and abuse. The primary aims of the blockchain solution are to (1) improve detection of potential fraudulent and illegal health care transactions and reimbursements, (2) create a more inclusive process for validating claims deploying a patient-centric approach, and (3) enhance efficiency in the claims adjudication process through smart contract automation.

To conceptualize this approach, we adopted the *fit-for-purpose* theoretical framework as published by Mackey et al [14] for designing health blockchain use cases that outline design and technical principles. These features are outlined below and are also described in the context of our early prototype version of the technical framework. On the basis of the central need for an environment that enables shared governance, we conceptualized our technical framework and prototype on the Ethereum decentralized platform, which enables 3 specific technical features that map to our use case, including (1) democratic autonomous organizations (DAOs), (2) smart contract execution environment, and (3) tokens via the ERC-20 token standard.

Hence, our design framework and prototype are based on the combination of *fit-for-purpose* design principles and feature layers on Ethereum.

Ethics Approval and Consent to Participate

Ethics approval and consent to participate was not required for this study. All information collected from this study was from the public domain, and the study did not involve any interaction with the users. User indefinable information was removed from the study results.

Results

General Design

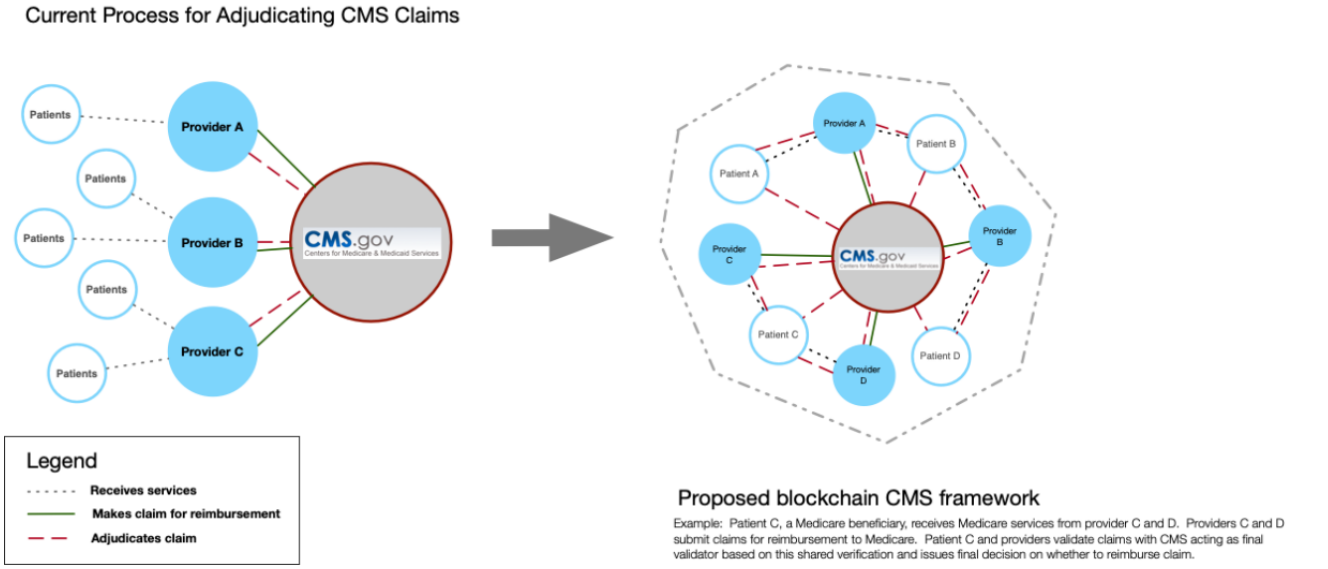
The proposed solution to improve health care fraud and abuse detection will utilize a *hybrid* or consortium permissioned blockchain model that incorporates relevant stakeholders in the claims and reimbursement workflow, a transactional process that currently does not involve the patient (Figure 1). A consortium blockchain consists of both *public* and *private* blockchain characteristics by restricting participation to certain trusted users who act as nodes on the blockchain and are required to meet the criteria set forth by the consortium.

Importantly, a consortium-based blockchain model enables a high throughput of information to be validated and stored on the blockchain. This is due to the restricted number of validation nodes that both process and distribute information throughout the system. In contrast, the Bitcoin blockchain can process approximately 8 transactions per second, and the public Ethereum blockchain can only process approximately 15 transactions per second, partly owing to the massive number of nodes connected to their respective networks. This is not sufficient for a health care fraud detection system, which requires significantly more than 100 transactions per second. The consortium blockchain model enables the system to scale to process large number of transactions while restricting participation and enhancing the security of entities who can access and interact with the system.

Our technical framework includes all subgroups (eg, providers, payers, and patients) that are eligible to provide or receive services through Medicare. We focus on Medicare as it is the largest public payer system in the United States (with projections that Medicare expenditures will increase from US $705 billion in 2017 to US $1.436 billion in 2027) and as many of the legal frameworks associated with health care fraud and abuse only apply to public sector reimbursement, although states may have their own laws and regulations when it comes to the private pay or the employer insurance–based market [31]. We primarily focus on Medicare claims under Parts A and B, but not C (Medicare Advantage), as these are capitated payments.

Next, we discuss the framework's core design features of shared data governance, interoperability, the smart contract claims validation process, proposed system architecture, and privacy considerations.

Figure 1. Visualization of the current process for Centers for Medicare and Medicaid Services claims reimbursement and proposed blockchain framework. CMS: Centers for Medicare and Medicaid Services.



Shared Data Governance

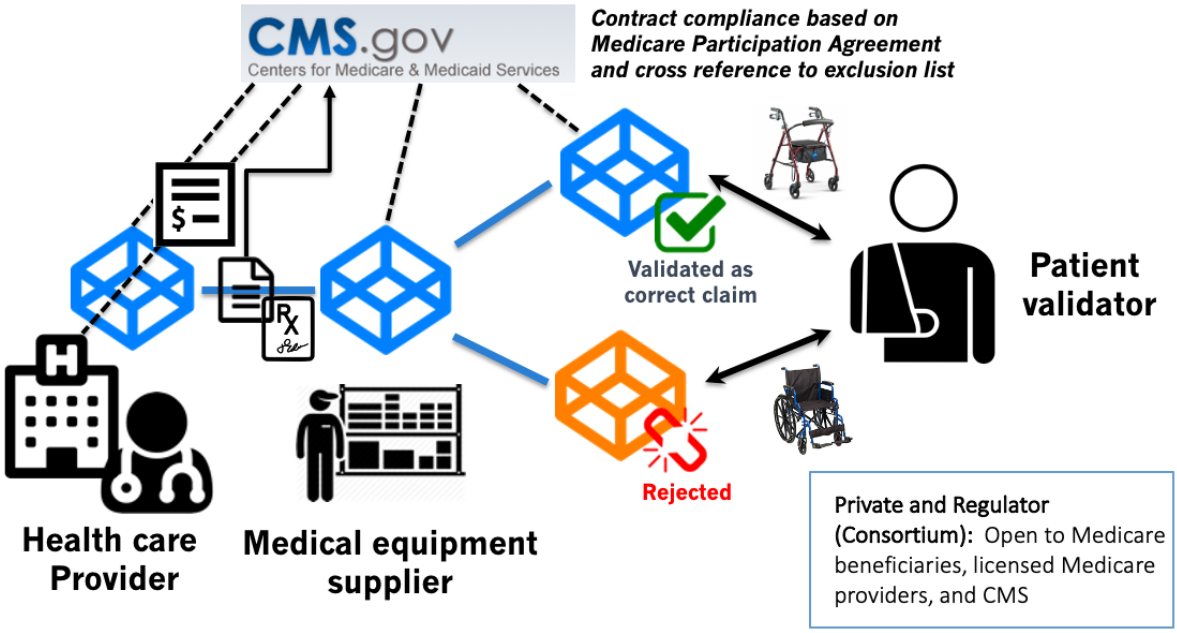
The first step in setting up a shared governance system is to define the membership of the distributed community that will participate in the consortium technical framework. A DAO will govern the framework with its subgroups of providers, payers, and patients based on verification that they are eligible to provide and receive Medicare services. Validating membership in these DAO subgroups will be accomplished by cross-referencing information directly from CMS Medicare identifier data (see Multimedia Appendix 1 [14] for more details on DAO subgroup matching). This will include matching CMS information on eligible providers (eg, clinicians accepting Medicare-approved payments), eligible provider organizations (eg, health care facilities accepting Medicare patients and clearinghouse organizations authorized to bill on their behalf), and eligible Medicare beneficiaries (eg, patients).

Furthermore, although blockchain is generally described as a decentralized network, this use case relies on a central payer to make a final adjudication decision on claims. Hence, the central authority of this consortium blockchain will be CMS, although the process of claim validation will be shared by the DAO. In this role, CMS will have privileges to add providers, gain access to universal claim information from providers and their associated patients, and adjudicate claims based on CMS' own set of rules and regulations that meet specific statutory criteria to validate that a service was actually billed and received by the patient appropriately. The framework also allows CMS to delegate and determine participation permission and access to different subgroups and determine validation rules in the network. For example, CMS may determine that there must be complete agreement among the associated patient, provider, and CMS for a claim to be confirmed and validated.

The DAO participants will act as *authority* nodes, which will validate the claims data that are proposed to be written to the blockchain. Importantly, the permissions structure for the validation procedure will be claims and patient-specific, with only the relevant provider, organization and beneficiary gaining access to identifiable reimbursement claims data, subject to identity verification. Data join fields (when contents of one database or table are joined based on a common attribute field), including type of payment (eg, fee-for-service and prospective payment systems), medical billing codes (eg, International Statistical Classification of Diseases and Related Health Problems-ICD-10, Current Procedural Terminology-CPT, Diagnosis Related Group-DRG, and Healthcare Common Procedure Coding System-HCPCS level 1 and 2 codes), and National Provider Identifier and patient's Medicare ID number of Medicare Beneficiary Identifier, can be used to validate identity and permissions for claims verification processes.

This validation of participating nodes will help prevent fraudulent actors from submitting claims (including *ghost* patients, ie, patients who do not exist or never received services, and providers and organizations debarred from Medicare participation). An example of how this would apply to a medical equipment claim is provided in Figure 2.

Figure 2. An example of a claim for a durable medical equipment, where a health care provider submits a claim for a walker that is supplied by a durable medical equipment provider, with the identity of the provider, durable medical equipment provider, and patient beneficiary validated by CMS. The claim is then either validated or rejected based on whether the patient actually receives the correct medical equipment. In this example, CMS was billed for a wheelchair, but the patient received a walker and validated that the claim was incorrect, leading to a potential denial of claim and detection of a false claim. CMS: Centers for Medicare and Medicaid Services.

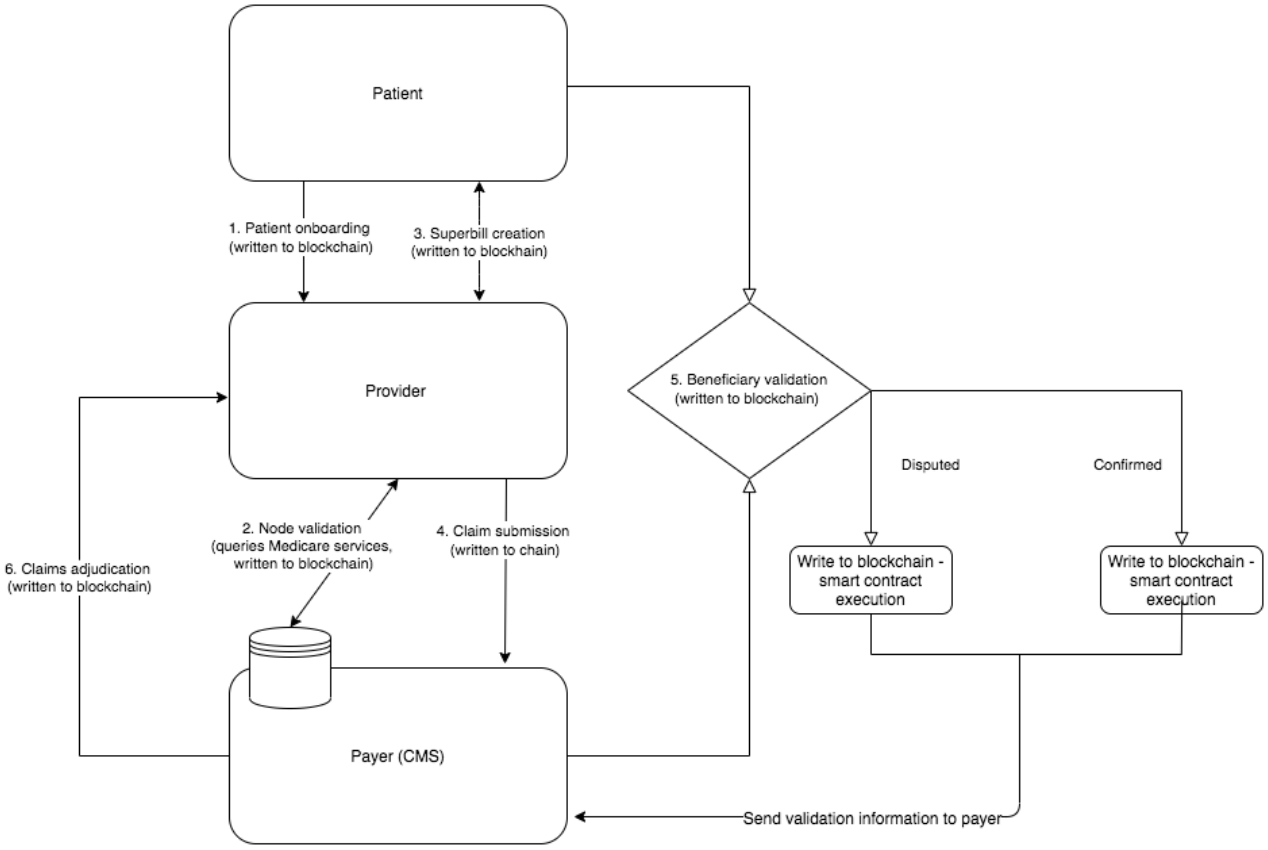


Interoperability and System Integration

Key to the functioning of processes to validate participating nodes and health care claims is establishing the interoperability of the system to interact with off-chain databases and systems, including CMS databases of verified Medicare providers and beneficiaries, medical coding and billing systems (including front-end and back-end billing and health care electronic data interchange), revenue cycle management systems, and current Medicare billing and antifraud solutions. Integration with CMS databases of validated providers and beneficiaries, either through a data clearinghouse model (ie, not directly integrated with CMS systems but using a third-party data clearinghouse) or through existing application programming interface (APIs; with query function such as the CMS Data portal API in JavaScript Object Notation-JSON) to directly query CMS or provider databases, will be a priority. In addition, integration with new CMS patient-centered initiatives, such as the Medicare's Blue Button 2.0 (explained later in the *Discussion* section), will also be explored.

To better ensure broader health informatics interoperability, our technical framework is also intended to ingest data solely related to the electronic claims submission process for Medicare Part A and B using the Accredited Standards Committee X12 standard transmission format (also known as Health Insurance Portability and Accountability Act [HIPAA] 5010), although it could also integrate into electronic health records (EHRs) systems using the Health Level Seven International Fast Healthcare Interoperability Resources (HL7 FHIR) standard. Finally, to enable interoperability while maintaining privacy, a secure off-chain key-value store, which can only be accessed and modified through a trusted execution environment within the system, will map the address or identifiers of users in our framework to the CMS identifiers. Validation of correctness and integrity between an off-chain database and the blockchain network may be performed using 2 patterns known as the challenge response pattern and the off-chain signature pattern (Figure 3 and Multimedia Appendix 1) [32].

Figure 3. Process for validating data off-chain in proposed framework. CMS: Centers for Medicare and Medicaid Services.



Smart Contract Claims Validation Process

For the desired utility of our technical framework, it is necessary to map the Medicare claims submission, adjudication, and reimbursement processes to our framework's data governance and smart contract feature layer. Below, we describe the basic claims workflow for the proposed framework that maps to the key procedures of verifying Medicare eligibility, creating a claim, submitting a claim, patient verification, and final adjudication. The proposed claims workflow, which will be automated by smart contracts in the framework, is described in Table 1.

Importantly, at each step of the Medicare claims adjudication process, smart contracts will govern how information is shared, what data are required, and how consensus is established for what is written to the blockchain. For consensus of validating and writing to the blockchain, our framework utilizes a proof-of-authority (POA) consensus mechanism. POA, a modified version of proof of stake, uses the validator's identity as a form of stake to validate blocks to be written to the blockchain. In this sense, consensus of decisions on validating a claim for the network (eg, submission of claim, viewing claim, and final validation of claim) will be visible to all stakeholders involved in the claim (ie, the provider who rendered services, the patient who received the services, and CMS which acts as the payer for services) and will also be tied to the user's identity. POA validation nodes should have their identity validated with CMS (as explained above) in the DAO, and their validation privileges can be revoked in cases of fraudulent behavior or if they are disbarred from Medicare. The rules of validation will be determined by the DAO to ensure that all stakeholders are accounted for. For example, the DAO may determine that over 90% of all authority nodes must validate information before being written to the blockchain.

As the central authority on our framework, CMS will set the rules for smart contracts, the parameters of consensus at each step, and the permission and data governance structure for the network, subject to federal laws such as 42 USC section 1395 and the following, 42 Code of Federal Regulations section 400 and the following, and HIPAA. The provider will have access to the framework and smart contract layer to add patients, provide services, and file claims based on the services they have rendered and where they have submitted a claim on behalf of a Medicare beneficiary. Patients will have access to view the services and associated claims that have been submitted with their digital beneficiary identity (eg, Medicare number) and have a role in validating those claims.

Table 1. Description of mapping the framework's smart contract processes to the Centers for Medicare and Medicaid Services general claims adjudication process.

Step	Claims process	Description
1	Patient onboarding	Patient registers and onboards at the provider location, confirms Medicare eligibility, and schedules an appointment—written to chain
2	Node validation	Patient, provider, and/or organization is validated for eligibility for Medicare services and benefits—query network and written to chain
3	Superbill creation	Medicare eligible services are provided to the patient by a health care provider and organization and a "superbill" (comprising claim codes and patient information) is created—written to chain
4	Claims submission	Provider submits claim directly to CMS[a] or uses a third party (ie, clearinghouse)—written to chain
5	Beneficiary validation	Patient beneficiary to the claim is queried to validate the services received upon a filed claim—written to chain
6	Claims adjudication	Payer (CMS) adjudicates the claims with validation information from both the provider and patient records and executes proof-of-authority consensus across other validating nodes (ie, patient and provider)—consensus results—written to chain
7	Electronic remit advice form	Payer (CMS) assesses whether to accept, deny, or reject a claim and provides payment information via an electronic remit advice form—written to chain

[a]CMS: Centers for Medicare and Medicaid Services.

Prototype System Architecture

Our technical framework consists of a web application (front end), blockchain network, and off-chain database storage (back end). The web application will display the system information on a graphical user interface (GUI). The blockchain network will validate and record all transactions that occur in the system. The off-chain database storage will store information regarding user credentials as validated by CMS as well as protected health information (PHI) or personally identifiable information (PII) not stored on the blockchain network but already present in existing clinical and EHR systems.

The system will implement a hybrid consortium blockchain model, wherein authorization is required to join the network. POA consensus will enable validation power among different stakeholders in the system in a distributed manner to mitigate collusion and false records among patient and provider workflows. Different read and write permissions will be given to different stakeholders in the network depending on the smart contract claims adjudication processes previously outlined. All transactions will be indexed in a system administration database to audit and enable efficient queries of aggregated data from the entirety of the network.

A prototype of this blockchain Medicare fraud and abuse framework can be explored and run by downloading and following instructions from our GitHub Repository (San Diego Supercomputer Center-BlockLAB Medicare-Claim-Verification). The software stack used to build the system consists of a front-end user interface, a back-end processing server, and a blockchain network (Figure 4). The blockchain application can be executed by running a local Ethereum blockchain via Ganache; deploying smart contracts onto the Ethereum blockchain via Truffle; installing all NodeJS packages; and running the NodeJS application, which is connected to the Ethereum network through Web3.

The front end of the web application will comprise 3 separate GUIs corresponding to the different roles in the system (eg, payer, provider, and patient) and will map directly to the smart contract claims adjudication function inputs (Figure 5). The roles of the system will be determined by a registration process in which users' credentials will be validated against CMS registries, confirming proper identification and roles in the Medicare reimbursement process. For example, a patient of the system must provide their Medicare beneficiary number, which will be cross-referenced off-chain against the Medicare database to validate access to the system. Upon proper registration, the user credentials will be stored as well as their associated role, which will be validated with input information from a log-in GUI.

The payer (ie, CMS) GUI presents information about the providers and claims. The current implementation breaks down the claims into 2 lists, verified claims and unverified claims, for the payer to distinguish whether a patient has verified a Medicare claim submitted. The provider GUI presents information regarding patients and allows providers to write information to the blockchain regarding both providing a service and filing a corresponding Medicare claim. The patient GUI presents information regarding the current filed claims made on their behalf and allows a patient to confirm or dispute whether they were provided the health care service or benefit associated with the claim.

Users designated with the payer role also have access to onboard providers who have been verified and registered in the system, pay and/or adjudicate a claim, and read all the claim information regarding providers and patients that have been registered and validated with their organization or agency (in our case, CMS). Users designated with the provider role have access to onboard verified patients and have access to read patient information that has been registered with their entity only. Users designated with the patient role only have access to confirm or dispute claims associated with their verified Medicare digital identity in the system.

The data storage of the current implementation can be categorized into on-chain storage, which is data written to the blockchain, and off-chain storage, which is data stored in a

traditional database (structured query language-SQL or not only SQL-NoSQL) external to the blockchain network. On-chain data comprise entity relationship information, such as which payer-provider relationships and provider-patient relationships need to be validated as part of the claim adjudication process. On-chain data also comprise information regarding the adjudication of services provided and claims submitted. The on-chain information is stored on Solidity smart contracts, which are deployed to a private Ethereum blockchain and are used to populate the GUIs of the web application via event listeners and Solidity contract calls (Multimedia Appendix 1). Off-chain data comprise user credential information, PHI, Medicare data, and other information needed to authenticate and integrate into the Medicare claims adjudication workflow, as previously discussed.

Finally, patient verification is a crucial step to adjudicate Medicare claims under our framework. To further incentivize patients to participate in the blockchain-based validation process, we have implemented an ERC-20 token to encourage active patient-generated claims validation (see Multimedia Appendix 1 for details on ERC-20 tokens). The ERC-20 token is meant to have dual utility within our framework to encourage patient validation of claims and to incentivize other population health benefits. Further research will explore different utilities for the ERC-20 token to maximize patient participation, such as allowing use of tokens to lower patient cost sharing (ie, co-pays) and incentivizing other health behaviors (eg, issued tokens for claims validation used to lower the cost of prescription drugs and for fitness club memberships).

Figure 4. Overall system architecture of the framework, with React used for the user interface and NodeJS and an Express server used for the back-end processing server. Solidity was the smart contract language used to interact with a local Ethereum blockchain network. The Application Programming Interface is a set of functions and procedures allowing communication between the front-end user interface, back end server, blockchain network as well as access to functions and data of the system. The Ethereum Virtual Machine is the runtime environment for smart contracts in Ethereum. JavaScript Object Notation remote procedure protocol is a specification that defines several data structures and the rules around their processing. Interaction with the Ethereum blockchain starts with sending a request via JSON RPC. API: Application Programming Interface; EVM: Ethereum Virtual Machine; JSON: JavaScript Object Notation; RPC: remote procedure protocol.

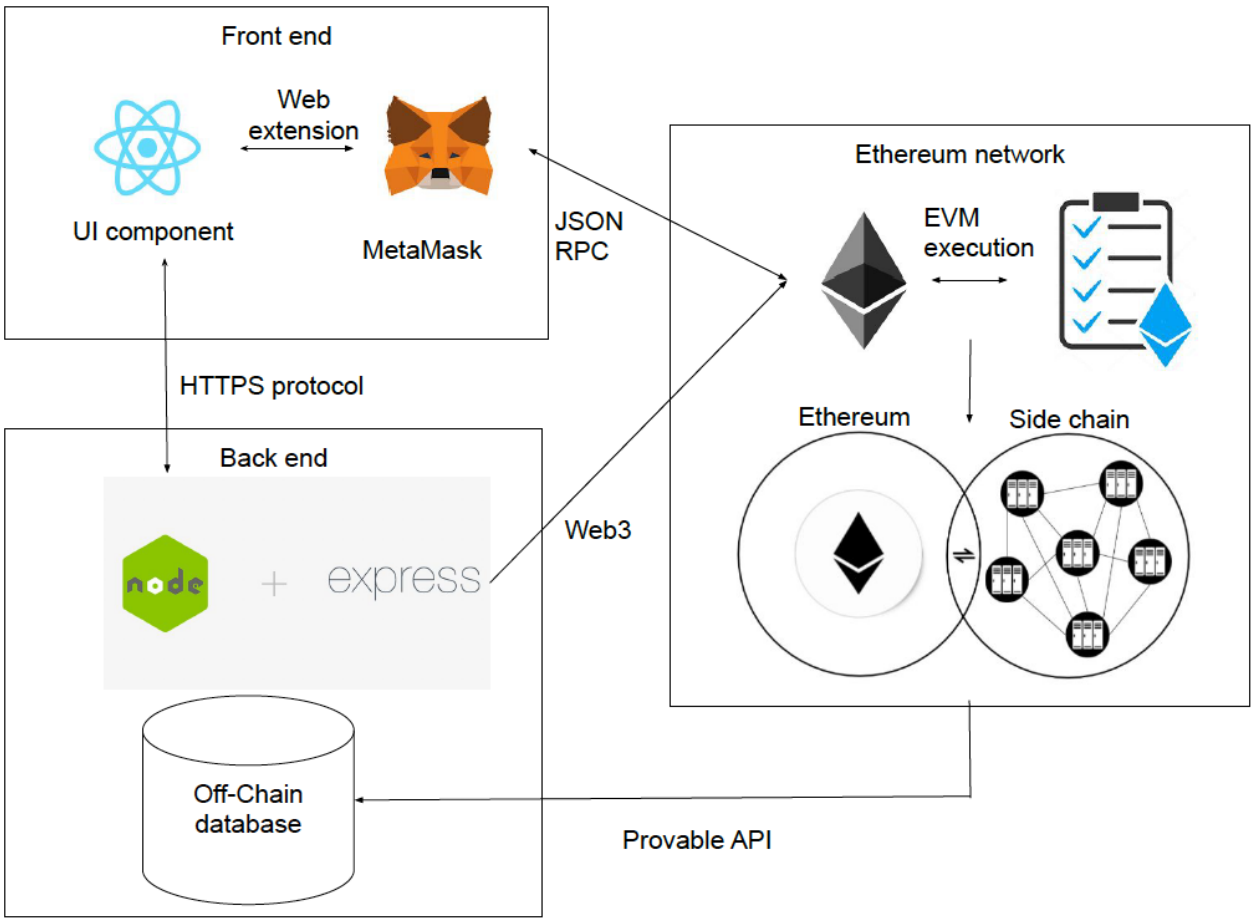


Figure 5. Description of the framework's smart contract function inputs.

Insurer

Method Name	Description
getInfo	Emits the contract address and the name of the insurer
preloadInfo	Sets up the dashboard by adding the insurer's providers and patients
addProvider	Adds a new provider for the insurer
getProviders	Gets all the current providers for the insurer
payProvider	Confirm to pay specified provider for the total of verified services
getAllVerifiedClaims	Gets all current verified claims for each provider
getAllUnverifiedClaims	Gets all current unverified claims for each provider
getAllServices	Lists all services that have been provided by all providers

Patient

Method Name	Description
verifyClaim	Verifies a claim made by a provider and emits an event to the insurer dashboard, to appear in verified claims
recordClaim	Records the claim in the patient contract
recordService	Pushes a service provided to all unclaimed services of the patient and emits an event for added claim
getUS	Gets all of patient's unclaimed services
getUC	Gets all of patient's unverified claims
getVC	Gets all of patient's verified claims
getLastSC	Gets patient's most recent unverified claim

Provider

Method Name	Description
addPatient	Adds a new patient to the system to provide services for, with specified name and id
getPatients	Gets a list of all patients for this provider
getPatient	Emits the details for a specified patient based on address
provideService	Provides a service to a certain patient and records the total cost
fileClaim	Files a claim for the patient after providing a service
getPatientName	Returns specified patient's name based on their address

Service Claim

Method Name	Description
file	Emits an event after filing a claim to record the amount and time of filing
verify	Verifies that a claim has been filed and returns true
getAmount	Gets the amount of the claim
isVerfied	Gets the verification status of the claim
getPatientAddress	Gets the patient address associated with the service claim

Privacy Considerations

As patient claims data are subject to privacy and confidentiality under HIPAA, no PHI or PII will be written to the public Ethereum blockchain. Instead, access to all patient-level claims data will be restricted to permissions validated to the identity of providers and organizations of the DAO with specified roles involved in the claim (ie, HIPAA covered entities); any organization subject to a HIPAA business associate agreement, which needs access to such data; and the patient who is the sole beneficiary of the claim.

All PHI will be stored off-chain isolated in existing systems with identified join and match data attribute fields the only event queried (such as Medicare claim number or Medicare Beneficiary Identifier). The information that will be written to the chain is simply the adjudication of the claim information itself, with encrypted data regarding any patient identifiers, wherein management of public and private keys to access

third-party databases will be needed to correlate any claim information with PHI or PII. Properly deidentified claims and associated metadata can be written to a separate public chain of the framework for inspection by all stakeholders, including regulators and law enforcement, which can use these data to detect larger patterns of fraud and abuse and suspicious claims or reimbursement activities through existing data mining and predictive analytics approaches.

Discussion

Principal Findings

The purpose of our blockchain fraud and abuse technical framework and prototype is to enable a shared governance approach to addressing health care fraud and abuse, while also empowering patients with the option and authority to become active participants in the claims verification process. We base this central design principle on 2 key facts: (1) for certain health care claims, patients are best suited to verify whether appropriate health care services have actually been rendered, and (2) patient verification or lack of verification of claims provides an important indicator of potential fraud and abuse risk, which can later be confirmed by investigation and can also enable targeted fraud and abuse prevention that is more proactive than current approaches. Hence, our framework focuses on a patient-centered design to address health care fraud and abuse by engaging the patient as a key stakeholder in the claims verification process.

We chose blockchain over existing technologies, such as cloud computing or traditional database storage methods, as our focus is on *shared* validation of claims in a distributed and immutable ledger that is supported by cryptography. Central to this model is establishing *trust* in a shared governance approach across multiple parties in the same transaction that also enables the patient to be an additional validating node in this process. These core features, along with the technology application layers enabled by blockchain (eg, smart contracts, digital identity, tokens, and consensus mechanisms), are why blockchain architecture is ideal for this use case.

Specifically, blockchain improves both the security and data integrity of the information stored in a system. However, blockchain technology is not suited for all systems as a data storage solution. To determine the viability of blockchain technology for a health care fraud detection system, we used a structured methodology and flowchart published by Wust and Gervais [33] to determine whether a blockchain is the appropriate technical solution. Properties such as public verifiability, transparency, privacy, integrity, redundancy, and the anchor of trust are all considered when determining whether a blockchain is a viable technology for a given problem. The decision flowchart applied to our fraud and abuse framework is available in Multimedia Appendix 1. By applying this methodology, we concluded that a consortium permissioned blockchain is an appropriate technical solution to solve the problems posed to current health care fraud detection systems.

Furthermore, the use of smart contracts on the blockchain enables an agreed-upon rule set to be automatically executed based on events. This automated event-driven architecture can also be accomplished by traditional systems, but these are generally controlled by a centralized entity. Health care insurance systems involve multiple stakeholders with different incentives. Stakeholders are often forced to trust an entity based on perceived compliance with applicable regulations. Trust in the health care claims system could be significantly improved by automating the execution of rule-based logic through smart contracts. For example, a specific action can be automatically invoked if a patient disputes a health care claim in our system, whereas current systems require detection and auditing of fraud mostly retrospectively.

Our framework also focuses on a patient-centric design, which is compatible with government-wide initiatives led by CMS, including the MyHealthEData initiative (which aims to provide patients with more access and control of their health care records from a device or mobile app of their choice) and the Medicare's Blue Button 2.0 (which enables patients to access and share claims data) [34]. Importantly, Medicare's Blue Button initiative provides Medicare beneficiaries with claims data in a universal and secure format, which can be integrated into our proposed framework. Although Blue Button 2.0 provides access to claims data to providers, it does not provide methods for beneficiaries to validate claims or report possible fraud and abuse directly to CMS mediated by technology. Hence, our patient-centric blockchain framework complements this and other CMS data access initiatives, while also ensuring that patients are not just consumers of their claims data but can also take action in the event of a discrepancy.

Central to our approach is also the fact that health care fraud detection has traditionally been implemented via reactive systems that analyze fraudulent claims activity after a claim has already been submitted and has likely been paid. These traditional systems can be costly (as defrauded amount may never be recovered or requires lengthy litigation) and rarely have incorporated any patient feedback when validating the integrity of a Medicare claim. Here, blockchain technology offers a potential solution by creating a tamper-evident and near-immutable audit log of health care claims and transaction data that can be viewed and agreed upon in a distributed ledger by providers, payers (eg, CMS), and patients to collectively verify claims and work collaboratively to identify fraud and abuse.

Our framework also aligns with specific programmatic priorities of HHS to combat health care fraud and abuse, including initiatives to reduce fraud, waste, and improper payments across its different agencies. The Affordable Care Act (ACA) has provided resources to CMS to improve prevention of fraud, waste, and improper payments through its CMS Fraud Prevention Initiative and Fraud Prevention toolkit that enables enhanced collaboration with state and law enforcement partners using predictive modeling technology. Furthermore, the ACA has empowered CMS to jointly develop many Medicare, Medicaid, and Children's Health Insurance Program antifraud policies, leading to enhanced screening requirements for new providers and suppliers, a concept that aligns well with our stakeholder blockchain validation approach [35,36]. The overall objective of these approaches is to enable health care programs

to do less *paying-and-chasing* of fraudulent claims and do more proactive and transparent fraud prevention [35].

CMS also has plans to develop a preventative model that will help identify potential fraud before it occurs by utilizing analytical techniques to improve payment accuracy by identifying, in real time, atypical trends that could be indicators of waste or fraud to appropriately intervene, again representing a good use case for patient validation data that can improve the precision of the proposed analytical models [35]. The rules provide new CMS enforcement tools to fight fraud, such as the ability to suspend payments in cases of credible allegations of fraud that could arise from a patient, and requires a more rigorous screening process for providers and suppliers enrolled in Medicare, including possible cross-termination for federal and state health programs [35]. Using these tools, Medicare and state agencies will be watching for trends that may indicate significant potential for health care fraud and can temporarily stop enrollment of a category or geographic area of providers or suppliers that has been identified as high risk.

Finally, HHS has been given new authority to prevent problematic providers from participating in Medicare. Specifically, the ACA increased the federal sentencing guidelines related to health care frauds involving US $1 million or more in losses to federal health care programs to create more disincentives for this activity. With this new authority also comes the responsibility of both determining and proving that a health care fraud has occurred. Hence, the collection of current CMS' antifraud goals, initiatives, and authorities provides an opportunity to develop a blockchain-based Medicare fraud detection system that aligns with these objectives for the purposes of integrating and developing a fraud and abuse prevention model modernized for today's health care and technology offerings.

Possible Benefits, Limitations, and Challenges

The potential benefits of our framework focus on creating verified claims transaction logs and more efficient and validated workflows. First, instead of waiting for fraud and abuse to occur and then reacting to it retrospectively, the system will be designed to actively detect and prevent potential fraud and abuse and other noncriminal activities (eg, overbilling, unintentional upcoding, and billing errors) using a layer of patient validation that is not currently available in legacy claims adjudication systems.

It will also add a layer of aggregated data to detect more systemic forms of fraud and abuse that can be mined for geographic areas, vulnerable patient populations, and specific health care providers that may be prone to fraud and abuse activities. This claims workflow data generated by the framework, which can also be properly deidentified, could be written to a public chain for the purpose of data mining and research. In addition, cryptographically validated multistakeholder claims data (eg, the claim submitted, validating the identities of stakeholders, and consensus established about the claim) could also enable more efficient machine learning approaches to detect patterns and risk factors of fraud and abuse not available from current static claims data.

If our approach was implemented to augment the current health care fraud detection system and was able to prevent just 1% of the current lost value, it would have saved the US health care system over US $25 million. Furthermore, by automating a larger portion of the health care fraud detection system, cumbersome and tedious tasks such as the human review of health care claims could be reduced. Hence, incorporating a trust-based technology such as blockchain into health care fraud detection systems can have economic benefits and technical utility but needs further testing with real-world or synthetic data to assess feasibility.

However, there are also certain limitations and challenges associated with implementing our proposed Medicare blockchain fraud and abuse prevention system. First, full participation from all stakeholders in the Medicare claims lifecycle (eg, CMS, providers, and patients) will require a comprehensive process of integrating with existing information technology systems, identification of interoperability challenges, and ensuring the use of appropriate data standards (such as Blue Button 2.0, Accredited Standards Committee X12, and HL7 FHIR). However, integrating provider billing and revenue management cycle systems, existing Medicare databases and APIs, and a front end that can interact with the patient will likely prove challenging.

Furthermore, abuse of the framework system itself must be considered and mitigated. For example, there may be an attempt to manipulate the consensus between all stakeholders regarding a record of events. This includes situations where the beneficiary may be complicit with a fraud and abuse scheme, a situation where patient validation may actually lead to incorrect adjudication of a fraudulent claim. This is of specific concern when a patient may have a clear incentive to participate and benefit from fraud and abuse, such as in the context of opioid use disorder and drug diversion [37]. Our system will take multiple measures to identify and proactively prevent cases where both the provider and patient cooperate in fraud and abuse, with modifications to the smart contract claims adjudication process and consensus mechanism specific to high-risk claims and patient profiles. However, adding the patient to validation may also enhance anomaly detection when unusual validation behavior occurs and could also act as a cryptographically hashed affirmation and evidence of wrongdoing by a provider or patient for use in prosecution.

Owing to the different incentives available to providers, payers, and patients, a proper and mutually agreed upon consensus algorithm will need to be implemented to address many of these challenges. In addition, we have discussed the use of ERC-20 tokens in our framework to incentivize patients to verify claims. Along with ensuring that claims data are represented in a way that patients can understand and verify (including the translation of claims codes in lay terms and education on health literacy), proper incentives need to be in place to encourage patients to validate claims correctly. Tokens may also disincentivize bad behavior. For example, if a patient is complicit in a fraud and abuse scheme, token payments can be withheld, and even possible additional penalties could be applied to a patient's validated Medicare identity. Hence, the *tokenomics* of reporting

fraud and abuse and the benefits to the patient and Medicare itself, need further design and testing.

The framework we developed builds upon an emerging body of innovation seeking to transform the health care data and claims workflows using blockchain technology. However, all these proposals, including our own framework, face barriers to adoption and implementation that require further experimentation, assessment, and active collaboration with the health care community. In fact, although our framework focuses on CMS and Medicare, a similar consortium blockchain design tailored to a private payer's own closed network of providers and beneficiaries might represent a more pragmatic approach to detecting fraud and abuse and enable better integration with more centralized systems. Future work on our framework will focus on reference models for different payer and provider network consortium types.

Conclusions

Our blockchain framework proposes a tamper-evident and near-immutable audit log of health care claims and transaction data that can be viewed and agreed upon in a distributed ledger by providers, health care organizations, payers, regulators, and most importantly patients to verify claims for the purposes of limiting the loss of more than US $2 billion to health care fraud and abuse every year. Future studies of our proposed framework and prototype will need to focus on using synthetic or historic CMS claims data to assess the real-world viability of the framework.

Acknowledgments

The authors thank the San Diego Supercomputer Center for support and the AEEC Innovation Lab for feedback and support on the research presented.

The authors received funding from the San Diego Supercomputer Center BlockLAB and greatly acknowledge this support.

Authors' Contributions

TM, KM, DF, and SQ collected data and built the prototype for this study. TM and KM designed the study. All authors conducted the data analyses, wrote the manuscript, and approved the final manuscript.

Conflicts of Interest

KM is the principal owner of the blockchain startup company LedgerSafe. TM is the CEO and cofounder of S-3 Research, LLC, a big data startup company funded by the National Institutes of Health–National Institute on Drug Abuse that also conducts research on blockchain technology related to public health challenges including the opioid crisis. LedgerSafe and S-3 Research, LLC had no financial role in this study. The authors report no other conflict of interest associated with this manuscript.

Multimedia Appendix 1
Details on technical framework.
[DOCX File , 17 KB - jmir_v22i9e18623_app1.docx]

References

1. Justice Department Recovers over $3 Billion from False Claims Act Cases in Fiscal Year 2019. US Department of Justice. 2019. URL: https://www.justice.gov/opa/pr/justice-department-recovers-over-3-billion-false-claims-act-cases-fiscal-year-2019 [accessed 2020-08-14]
2. Livingston S. Feds Amassed $2.6 Billion From 2019 Healthcare Fraud Cases. Modern Healthcare. 2019. URL: https://www.modernhealthcare.com/legal/feds-amassed-26-billion-2019-healthcare-fraud-cases [accessed 2020-08-14]
3. National Health Care Fraud Takedown Results in Charges Against 601 Individuals Responsible for Over $2 Billion in Fraud Losses. US Department of Justice. 2019. URL: https://www.justice.gov/opa/pr/national-health-care-fraud-takedown-results-charges-against-601-individuals-responsible-over [accessed 2020-08-14]
4. Shrank WH, Rogstad TL, Parekh N. Waste in the US health care system: estimated costs and potential for savings. J Am Med Assoc 2019 Oct 7:- epub ahead of print. [doi: 10.1001/jama.2019.13978] [Medline: 31589283]
5. Mackey T, Liang B. Off label promotion reform: a legislative proposal addressing vulnerable patient drug access and limiting inappropriate pharmaceutical marketing. J Law Reform 2011;45(1):- [FREE Full text]
6. Nicholas LH, Hanson C, Segal JB, Eisenberg MD. Association between treatment by fraud and abuse perpetrators and health outcomes among medicare beneficiaries. JAMA Intern Med 2019 Oct 28:- epub ahead of print. [doi: 10.1001/jamainternmed.2019.4771] [Medline: 31657838]
7. Office of Inspector General (OIG), HHS. Medicare and state health care programs: fraud and abuse; revisions to the safe harbors under the anti-kickback statute and civil monetary penalty rules regarding beneficiary inducements. Final rule. Fed Regist 2016 Dec 7;81(235):88368-88409 [FREE Full text] [Medline: 27992158]
8. Herland M, Khoshgoftaar TM, Bauder RA. Big Data fraud detection using multiple medicare data sources. J Big Data 2018 Sep 4;5(1):21. [doi: 10.1186/s40537-018-0138-3]

9. Joudaki H, Rashidian A, Minaei-Bidgoli B, Mahmoodi M, Geraili B, Nasiri M, et al. Using data mining to detect health care fraud and abuse: a review of literature. Glob J Health Sci 2014 Aug 31;7(1):194-202 [FREE Full text] [doi: 10.5539/gjhs.v7n1p194] [Medline: 25560347]
10. Joudaki H, Rashidian A, Minaei-Bidgoli B, Mahmoodi M, Geraili B, Nasiri M, et al. Improving fraud and abuse detection in general physician claims: a data mining study. Int J Health Policy Manag 2015 Nov 10;5(3):165-172 [FREE Full text] [doi: 10.15171/ijhpm.2015.196] [Medline: 26927587]
11. Kesselheim AS, Studdert DM. Whistleblower-initiated enforcement actions against health care fraud and abuse in the United States, 1996 to 2005. Ann Intern Med 2008 Sep 2;149(5):342-349. [doi: 10.7326/0003-4819-149-5-200809020-00009] [Medline: 18765704]
12. Chung A. Supreme Court Declines to Broaden Whistleblower Protections. Reuters. 2018. URL: https://www.reuters.com/article/us-usa-court-whistleblower-idUSKCN1G51Y3 [accessed 2020-08-14]
13. Casino F, Dasaklis T, Patsakis C. A systematic literature review of blockchain-based applications: current status, classification and open issues. Telemat Inform 2019 Mar;36:55-81. [doi: 10.1016/j.tele.2018.11.006]
14. Mackey TK, Kuo T, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?' - challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Med 2019 Mar 27;17(1):68 [FREE Full text] [doi: 10.1186/s12916-019-1296-7] [Medline: 30914045]
15. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 1;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
16. Agbo C, Mahmoud Q, Eklund J. Blockchain technology in healthcare: a systematic review. Healthcare (Basel) 2019 Apr 4;7(2):56 [FREE Full text] [doi: 10.3390/healthcare7020056] [Medline: 30987333]
17. Zhang A, Lin X. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain. J Med Syst 2018 Jun 28;42(8):140. [doi: 10.1007/s10916-018-0995-5] [Medline: 29956061]
18. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. J Med Syst 2018 Nov 22;43(1):5. [doi: 10.1007/s10916-018-1121-4] [Medline: 30467604]
19. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating health data sharing: a solution based on the internet of things and distributed ledger technologies. J Med Internet Res 2019 Jun 6;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]
20. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
21. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials 2017 Jul 19;18(1):335 [FREE Full text] [doi: 10.1186/s13063-017-2035-z] [Medline: 28724395]
22. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]
23. Dwivedi A, Srivastava G, Dhar S, Singh R. A decentralized privacy-preserving healthcare blockchain for IoT. Sensors (Basel) 2019 Jan 15;19(2):326 [FREE Full text] [doi: 10.3390/s19020326] [Medline: 30650612]
24. Gordon WJ, Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
25. Hylock RH, Zeng X. A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study. J Med Internet Res 2019 Aug 31;21(8):e13592 [FREE Full text] [doi: 10.2196/13592] [Medline: 31471959]
26. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design choices and trade-offs in health care blockchain implementations: systematic review. J Med Internet Res 2019 May 10;21(5):e12426 [FREE Full text] [doi: 10.2196/12426] [Medline: 31094344]
27. Ghafur S, Schneider E. Engaging Patients Using Digital Technology — Learning from Other Industries. NEJM Catalyst. 2019. URL: https://catalyst.nejm.org/doi/abs/10.1056/CAT.19.0654 [accessed 2020-08-25]
28. Bresnick J. Five Blockchain Use Cases for Healthcare Payers, Providers. HealthITAnalytics. 2017. URL: https://healthitanalytics.com/news/five-blockchain-use-cases-for-healthcare-payers-providers [accessed 2020-08-14]
29. Khatoon A. A blockchain-based smart contract system for healthcare management. Electronics 2020 Jan 3;9(1):94. [doi: 10.3390/electronics9010094]
30. Chu L, Shah A, Rouholiman D, Riggare S, Gamble J. Patient-Centric Strategies in Digital Health. Cham, Switzerland: Springer; 2018:43-54.
31. Sisko AM, Keehan SP, Poisal JA, Cuckler GA, Smith SD, Madison AJ, et al. National health expenditure projections, 2018-27: economic and demographic trends drive spending and enrollment growth. Health Aff (Millwood) 2019 Mar;38(3):491-501. [doi: 10.1377/hlthaff.2018.05499] [Medline: 30785832]
32. Eberhardt J, Tai S. On or off the blockchain? Insights on off-chaining computation and data. In: Service-Oriented and Cloud Computing. Cham, Switzerland: Springer; 2017.

33. Wust K, Gervais A. Do you Need a Blockchain? In: Crypto Valley Conference on Blockchain Technology. 2018 Presented at: CVCBT'18; June 20-22, 2018; Zug, Switzerland URL: https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1109%2FCVCBT.2018.00011?_sg%5B0%5D=31HfyjHuwAwMCMclG2v-D_CZTjw4ZLZvkz-betGYQf65FW-vA6wlO9mXlqTuF-_u7daEft5QqSDOX5TGSabGMrTF-A.4yIiOsuOWeOMaEKQyhYDreetEOLNQTLE4L6nwnPlxgtE0tjUEcNzC710E8M12ibGmiqnLotBPp0hG5kTojks8g [doi: 10.1109/cvcbt.2018.00011]
34. Trump Administration Announces MyHealthEData Initiative to Put Patients at the Center of the US Healthcare System. CMS. 2018. URL: https://www.cms.gov/newsroom/press-releases/trump-administration-announces-myhealthedata-initiative-put-patients-center-us-healthcare-system [accessed 2020-08-14]
35. CMS Fraud Prevention Initiative: New Tools to Fight Fraud and Protect Taxpayer Dollars. CMS. 2018. URL: https://www.cms.gov/outreach-and-education/outreach/partnerships/downloads/backgrounderfraudpreventioninitiative.pdf [accessed 2020-08-14]
36. Medicare Fraud & Abuse: Prevent, Detect, Report. CMS. 2019. URL: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/Fraud-Abuse-MLN4649244.pdf [accessed 2020-08-14]
37. Raghavendra M. Can blockchain technologies help tackle the opioid epidemic: a narrative review. Pain Med 2019 Oct 1;20(10):1884-1889. [doi: 10.1093/pm/pny315] [Medline: 30848821]

Abbreviations

ACA: Affordable Care Act
API: application programming interface
CMS: Centers for Medicare and Medicaid Services
DAO: democratic autonomous organization
DOJ: Department of Justice
EHR: electronic health record
GUI: graphical user interface
HHS: Health and Human Services
HIPAA: Health Insurance Portability and Accountability Act
HL7 FHIR: Health Level Seven International Fast Healthcare Interoperability Resources
PHI: protected health information
PII: personally identifiable information
POA: proof-of-authority
SQL: structured query language
USC: United States Code

Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by FC Wang, T Ueno; comments to author 12.06.20; revised version received 02.07.20; accepted 26.07.20; published 10.09.20.

Please cite as:
Mackey TK, Miyachi K, Fung D, Qian S, Short J
Combating Health Care Fraud and Abuse: Conceptualization and Prototyping Study of a Blockchain Antifraud Framework
J Med Internet Res 2020;22(9):e18623
URL: http://www.jmir.org/2020/9/e18623/
doi:10.2196/18623
PMID:32909952

Original Paper

Current Challenges of Digital Health Interventions in Pakistan: Mixed Methods Analysis

Abdul Momin Kazi[1*], MBBS, MPH; Saad Ahmed Qazi[2,3*], PhD; Nazia Ahsan[1], BSc, MA, MSc; Sadori Khawaja[1], MBBS; Fareeha Sameen[4], BE; Muhammad Saqib[4], BE, MCS, MS, MEng; Muhammad Ayub Khan Mughal[1], Grad Dip; Zabin Wajidali[1], BSc, Msc; Sikander Ali[5], MSc; Rao Moueed Ahmed[4], BE, MS; Hussain Kalimuddin[1], MCS; Yasir Rauf[1], BCS; Fatima Mahmood[1], MD; Saad Zafar[6], BSc, MS, PhD; Tufail Ahmad Abbasi[6], MS; Khalil-Ur-Rahmen Khoumbati[7], PhD; Munir A Abbasi[8], PhD; Lampros K Stergioulas[8*], MSc, PhD

[1]Department of Pediatrics and Child Health, Aga Khan University, Karachi, Pakistan

[2]Department of Electrical Engineering, NED University of Engineering and Technology, Karachi, Pakistan

[3]Neurocomputation Lab, National Center of Artificial Intelligence, Karachi, Pakistan

[4]NED University of Engineering and Technology, Karachi, Pakistan

[5]Biomedical Department, NED University of Engineering and Technology, Karachi, Pakistan

[6]Riphah International University, Islamabad, Pakistan

[7]University of Sindh, Jamshoro, Pakistan

[8]Surrey Business School, University of Surrey, Guildford, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Abdul Momin Kazi, MBBS, MPH
Department of Pediatrics and Child Health
Aga Khan University
Stadium Road
P.O. Box 3500
Karachi, 74800
Pakistan
Phone: 92 2134864232
Email: momin.kazi@aku.edu

Abstract

Background: Digital health is well-positioned in low and middle-income countries (LMICs) to revolutionize health care due, in part, to increasing mobile phone access and internet connectivity. This paper evaluates the underlying factors that can potentially facilitate or hinder the progress of digital health in Pakistan.

Objective: The objective of this study is to identify the current digital health projects and studies being carried out in Pakistan, as well as the key stakeholders involved in these initiatives. We aim to follow a mixed-methods strategy and to evaluate these projects and studies through a strengths, weaknesses, opportunities, and threats (SWOT) analysis to identify the internal and external factors that can potentially facilitate or hinder the progress of digital health in Pakistan.

Methods: This study aims to evaluate digital health projects carried out in the last 5 years in Pakistan with mixed methods. The qualitative and quantitative data obtained from field surveys were categorized according to the World Health Organization's (WHO) recommended building blocks for health systems research, and the data were analyzed using a SWOT analysis strategy.

Results: Of the digital health projects carried out in the last 5 years in Pakistan, 51 are studied. Of these projects, 46% (23/51) used technology for conducting research, 30% (15/51) used technology for implementation, and 12% (6/51) used technology for app development. The health domains targeted were general health (23/51, 46%), immunization (13/51, 26%), and diagnostics (5/51, 10%). Smartphones and devices were used in 55% (28/51) of the interventions, and 59% (30/51) of projects included plans for scaling up. Artificial intelligence (AI) or machine learning (ML) was used in 31% (16/51) of projects, and 74% (38/51) of interventions were being evaluated. The barriers faced by developers during the implementation phase included the populations' inability to use the technology or mobile phones in 21% (11/51) of projects, costs in 16% (8/51) of projects, and privacy concerns in 12% (6/51) of projects.

Conclusions: We conclude that while digital health has a promising future in Pakistan, it is still in its infancy at the time of this study. However, due to the coronavirus disease 2019 (COVID-19) pandemic, there is an increase in demand for digital health and implementation of health outcomes following global social distancing protocols, especially in LMICs. Hence, there is a need for active involvement by public and private organizations to regulate, mobilize, and expand the digital health sector for the improvement of health care systems in countries.

(J Med Internet Res 2020;22(9):e21691) doi:10.2196/21691

KEYWORDS

digital health; eHealth; LMICs; mHealth; Pakistan; SWOT; telehealth

Introduction

The World Health Organization (WHO) addresses digital health as a global strategy in 2020-2024, and its digital health policy draft defines digital health as "the field of knowledge and practice associated with the development and use of digital technologies to improve health" [1]. It is a broad umbrella term encompassing mobile health (mHealth), eHealth, telemedicine, and advanced computing sciences like genomics, artificial intelligence (AI), and big data [2]. Digital health is a rapidly growing industry that, according to some estimates, is expected to be valued at US $504.4 billion by the end of 2025 [3]. It is being viewed as an accessible and affordable solution for people who do not have access to the traditional health system, and an important tool in achieving sustainable development goals [4].

Digital technologies like artificial intelligence (AI) and machine learning (ML) are an integral part of many businesses and companies in the developed world. They are a driving force in multimillion-dollar industries, such as automotive manufacturers, who rely on AI to, for example, predict when cars might need repair to ensure the safety of passengers [5]; the banking sector, where AI is applied to detect fraudulent transactions [6]; and Silicon Valley giants like Google, Facebook, and others, who have formed a consortium for conducting research intended to improve the understanding of AI technologies for the improved welfare of society [7]. The health sector is also embracing the digital revolution; digital health technology in developed countries is being employed in major aspects of the health care process, including consultation, diagnosis, treatment, monitoring, patient education, behavioral modification, and medication adherence [8-11]. In the fast-growing field of health AI, a recent study in Japan used an AI-based diagnostic system that demonstrated a higher diagnostic accuracy for esophageal carcinoma than those from conventional methods [12]. A retrospective study conducted to evaluate the accuracy of AI in predicting the mammograms of biopsy-proven breast cancer patients showed that AI outperformed all of the human readers [13]. Similar studies conducted on dermatological lesions using deep convolutional neural networks have shown that the ability of AI to classify malignant and nonmalignant conditions correctly is comparable to that of board-certified dermatologists [14].

A meta-analysis evaluating the impact of digital health interventions on noncommunicable diseases (NCDs) showed that the interventions significantly reduced the occurrence of cardiovascular outcomes such as stroke and myocardial infarction through positive behavior change theory [15]. mHealth-based interventions such as MyAirCoach (Asthma UK) are helping patients achieve better control of asthma-related symptoms by educating them on proper inhaler use and providing personalized treatment plans in coordination with health care providers in cases of exacerbations [16]. As seen in relevant randomized controlled trials, this intervention is part of an initiative to digitalize asthma care in the National Health Service (NHS) in the United Kingdom [16-18]. In the United States, over 61% of health care institutions provide telemedicine services, which are covered by insurance policies and Medicare programs [19,20]. According to estimates, telemedicine services helped the US Department of Veterans Affairs reduce its hospitalization for mental health diseases by over 40% in 2012 [21,22].

The expansion of mobile and wireless technologies offers an unprecedented opportunity for global health delivery in low and middle-income countries (LMICs). Digital health innovations are addressing issues such as maternal, newborn, and child health; low immunization coverage; lack of access to life-saving medications; infectious disease outbreaks; and the increasing burden of NCDs [23,24]. Sub-Saharan African countries are using an SMS text messaging technology project, SMS for Life, to ensure accurate reporting of real-time facility stock data to reduce antimalarial drug stock-outs [25,26]. India is using digital health tools to combat tuberculosis pandemics through eCompliance, which helps to monitor patients in real-time, ensure better medication adherence, and decrease treatment default rates [27-29]. Bangladesh is storing its health information data in a common data repository called the District Health Information Software 2 (DHIS2, Health Information Systems Programme), which allows for real-time monitoring, accurate localization of under-resourced areas, and better resource allocation [30,31].

In Pakistan, the doctor-to-patient ratio is close to 0.83 physicians per 1000 individuals in the population. Digital health interventions are being designed to address various health care needs. Several SMS-based interventions are being used to improve medication compliance in patients with NCDs [32], and telemedicine tools are being used to educate patients and to keep health care professionals abreast of medical advancements [33]. Moreover, as many female doctors leave clinical practice due to household and childcare responsibilities, telemedicine initiatives such as eDoctor (SE Software Technologies) and Sehat Kahani (Grocode.io) enable them to conduct their medical practices remotely via online patient consultation through a telehealth platform [34-36]. Another intervention, a mobile app called Teeku (Aga Khan University)

aimed at helping vaccinators record immunization data, generated reliable data for better monitoring and improved the coverage of expanded program on immunization (EPI) vaccines such as the pentavalent and pneumococcal conjugate vaccine [37]. Outbreak investigation and surveillance is another relevant public health domain that can be improved with digital health interventions. In 2011, after a particularly severe Dengue outbreak, a GPS-enabled mobile application called the Dengue Activity Tracking System (Punjab Information Technology Board) was developed to track suspected and confirmed cases of dengue [38]. In 2 recent epidemics of the extensively drug-resistant typhoid and human immunodeficiency virus (HIV) in the province of Sindh, geospatial mapping helped considerably with identifying the root causes of the outbreaks and with isolating cases and their spreads [39,40].

These aforementioned programs are just a few examples of successful digital health interventions carried out in Pakistan, despite limited resources available due to the financial constraints of the country. Mobile app solutions and social media have been shown to be quite effective in various programs worldwide, but there is limited data from LMICs on the use of emerging technologies in improving health care services. In this study, we propose to identify the current digital health projects being carried out in Pakistan and the key stakeholders involved in these initiatives. Further, we evaluate these projects and studies through strengths, weaknesses, opportunities, and threats (SWOT) analysis to identify the internal and external factors which can potentially facilitate or hinder the progress of digital health in Pakistan.

This will enable us to highlight specific challenges and areas that require digital health, to assess gaps in the current system, and to identify the roles of health care providers, technology partners, public and private partners, and policymakers in creating an environment for digital health fraternity to sustain prosperity.

Methods

Identification of the Targeted Population

The main objective of this study is to investigate the major digital health interventions in Pakistan during the last 5 years. The target population includes individuals or teams who have used technology-based interventions in the field of digital health.

Research Design

This study is conducted through mixed methods [41,42]. The data were gathered following the conceptual framework of Arksey and Mallor [43], categorized according to the WHO's recommended building blocks for health systems research and analyzed using the SWOT analysis strategy [44].

Figure 1 describes the flow of the study and the major components of the study methodology. The study was divided into 3 phases: Phase 1 was related to a literature search to identify relevant authors, papers, and interventions; Phase 2 was related to questionnaire development, piloting, and data collection using a mixed-methods approach; Phase 3 involved the SWOT analysis of the collected data.

Figure 1. Study flow and major components of the study methodology.

Inclusion and Exclusion Criteria

All studies in digital health conducted in Pakistan over the last 5 years were included.

Projects, innovations, and interventions that were completed more than 5 years ago in Pakistan were excluded from our study.

Research Population

An electronic, systematic, in-depth literature search was conducted in March 2019 using PubMed, Google Scholar, Web of Science, J.Store, Academia, and Pak Medinet. The key medical subject headings (MeSH) terms used for these databases included “mHealth” [MeSH Terms] AND “Pakistan” [MeSH Terms], “digital health” [MeSH Terms], OR “telehealth” [MeSH Terms], OR “Mobile phones” [MeSH Terms], OR “Cell Phone”[MeSH Terms] OR “Mobile Applications”[MeSH Terms] OR “Text Messaging”[MeSH Terms] OR “Telemedicine”[MeSH Terms] AND “Pakistan”[MeSH Terms].

A search strategy similar to that of a systematic review was used to identify the maximum number of interventions, projects, studies, or smartphone applications that had been carried out in the field of digital health in Pakistan. The search for published studies was not limited to any particular study design; all types of study designs were included.

In addition to published work, applications uploaded on Google Play store were also evaluated and shortlisted for further analysis. Google Play was thoroughly searched, as 95% of the smartphones used in Pakistan are Android-based, and usually, the apps available in other app stores are already on this platform [45,46]. We also contacted the eHealth Association of Pakistan (eHAP) for information on current digital health studies. Studies were excluded from the review if the project was not based in Pakistan, did not explicitly involve digital health interventions (eg, telemedicine, other types of eHealth, or used other telecommunication technologies such computers, internet, or email), or examined only feasibility or acceptance of eHealth in society.

The data collected was then independently assessed for inclusion and eligibility. Any duplicated titles were removed, and the final list of published studies, projects, innovations, and smartphone applications was then imported on to a Microsoft Excel file (Multimedia Appendix 1). The projects were divided into different categories, including telemedicine projects, mobile apps, studies published in the last five years, as well as their authors, provinces, and institutions that the authors were

affiliated with. This system of organization eased the process of approaching respondents.

Identification of Stakeholders and Mapping

The project team then made efforts to contact the stakeholders of the projects identified in the literature search and brief them regarding the study objectives. This data was used for stakeholder engagement activities, including questionnaire surveys and co-design workshops. The co-design workshops were carried out in various cities in Pakistan. For this study, 51 stakeholders were approached. Each participant or stakeholder represented an individual study or intervention. The organizing team members met the stakeholders, both in-person and online, to ensure streamlined communication with everyone involved. This also provided information regarding other unreported or unpublished technical innovations.

Questionnaire Development and Administration of the Stakeholder Survey

A questionnaire with qualitative and quantitative sections was developed for data collection, which included both open-ended and close-ended questions. The questionnaire included sections on (1) the respondents' details, (2) project details, (3) team details, (4) project technology, (5) project evaluation, and (5) ethics in digital health, which were captured in line with the WHO's building blocks for health systems [44,47]. Pilot testing of the questionnaire was conducted initially to help us refine the questionnaire. The finalized study questionnaire was administered to the core project team members who had health and technology expertise.

Data Analysis

A mixed-methods strategy was used for data analysis. The collected data was divided into 3 sets of analysis: quantitative, qualitative, and SWOT analysis. The quantitative data were analyzed in Matrix Laboratory (MATLAB R2015b 9.7, MathWorks) and SPSS (version 19.0, IBM Corp) to run basic frequencies, while the qualitative data were reviewed for familiarization, coding, and the generation of themes. The findings were presented as a descriptive summary and thematically analyzed according to the WHO's defined health system building blocks [44,47]. The main frequencies of the quantitative data and the themes derived from the qualitative data were summarized and extracted into a table of 4 main groups: strengths, weaknesses, opportunities, and threats (SWOT).

SWOT analysis was performed to identify the internal factors (including the demographics of the researcher and the institute, scalability, sources of funding, and feasibility) and external factors (including the diseases targeted by the projects, technologies being used to carry out the intervention, monitoring and evaluation, and ethical challenges) that influenced the implementation of the project or study [48]. SWOT analysis also provided insight into the often complex relationships between the various factors and aspects necessary to create sustainable improvements. Lastly, SWOT analysis was used to serve as a main strategic project planning tool to better inform decision makers in evaluating the impact of different digital health initiatives in Pakistan.

Data Management

The finalized questionnaire was developed as a web-based application connected to a MySQL database server, hosted on a secured local cloud. The development of the web application was tested for correctness, timely data inputs, and secure data saving. The team administering the questionnaire was trained in 2 separate group sessions, in which the study methodology and design were explained, and each question and the associated probes were discussed in detail. This team was then divided to approach the stakeholders individually and administer the questionnaire through face-to-face, telephonic, and Skype interviews lasting 20-30 minutes, after obtaining the interviewees' written consent. The responses were uploaded in real-time and stored anonymously using a coded respondent ID, and both quantitative and qualitative inputs were stored digitally.

Ethical Considerations

The study protocol and associated study instruments, including the consent form, were approved by NED University of Engineering and Technology and Aga Khan University's Ethics Review Board before the commencement of any study activities.

Results

Quantitative Results

Baseline information of the quantitative survey is shared in Table 1. Most of the 51 respondents were public sector (17/51, 34%) and private sector (17/51, 34%) employees working in health care organizations, followed by academia (9/51, 17%) and nongovernmental organizations (1/51, 2%). Of the 51 projects, 57% (29/51) commenced after 2016, 33% (17/51) commenced between 2011 and 2015, and 6% (3/51) commenced between 2006 and 2010. Regarding funding, 37% (19/51) of interventions were funded internally, followed by 29% (15/51) funded internationally and 16% (8/51) funded nationally. Of the applications designed, 56% (28/51) were close source and 14% (7/51) were open source.

Table 1. Baseline information of the digital health studies and projects in Pakistan (N=51).

Variables	n (%)
Age (years)	
20-30	5 (10)
31-40	31 (61)
41-50	10 (20)
>50	5 (9)
Gender	
Male	41 (80)
Female	10 (20)
Field of expertise (n=47)	
Information technology	22 (47)
Engineering	6 (13)
Public health	6 (13)
Digital health	9 (19)
Others	4 (9)
Department (n=44)	
Computer science and information technology	11 (25)
Pediatrics	9 (21)
Primary and secondary health care department	8 (18)
Engineering	8 (18)
Other	8 (18)
Province/territory where the project is being conducted	
Sindh	32 (65)
Punjab	9 (18)
Federal capital	5 (10)
Gilgit Baltistan	2 (4)
Khyber Pakhtunkhwa	1 (2)
Not answered	2(4)
Developer of the application (n=43)	
University	16 (37)
Private institute	12 (28)
Semi-government institute	8 (19)
Individual developer	7 (16)
How do you define work on technology in health? (n=43)	
Digital health	19 (44)
Technology in health	16 (37)
General health	5 (12)
Geospatial	2 (5)
Device development	1 (2)
Did you report the study/project dissemination? Select all that apply (n=31).	
Publications	12 (39)
Websites/blogs	11 (35)
Government and funding agency	5 (16)

Variables	n (%)
Not yet	3 (10)

Of the 51 interventions, 45% (23/51) were centered around research, 29% (15/51) on implementation, 12% (6/51) on application or software development, 4% (2/51) each on prototype or device development, and 2% (2/51) each for commercial projects and system development. Core teams consisted of professionals from different backgrounds; information technology (IT) professionals comprised 40% (21/51) of the core teams, followed by investigators (5/51, 10%), epidemiologists (5/51, 10%), students (4/51, 8%), administrators (3/51, 7%), health care professionals (3/51, 7%), field staff (3/51, 7%), electrical engineers (2/51, 4%), biomedical engineers (2/51, 4%), psychologists (2/51, 4%), and cell phone providers (1/51, 2%).

The health domains targeted by the interventions included general health (23/51, 46%), immunization (13/51, 26%), diagnostics (5/51, 10%), and mental health and behavioral change (3/51, 6%). The disease outcomes targeted were NCDs in 56% (29/51) of the interventions, infectious diseases in 33% (17/51) of the interventions, and mental health in 7% (4/51) of the interventions. In regard to target populations, 33% (17/51) of the interventions catered to the general population, 30% (15/51) were specifically for children and adolescents, and 20% (10/51) targeted adults only. The targeted population belonged to all the various socioeconomic classes, with 39% (20/51) belonging to the middle class, 38% (19/51) to lower socioeconomic class, and 23% (12/51) to upper socioeconomic class, as per the participants' self-rating.

These projects are also using innovative ideas for their implementation: 30% (15/51) of the respondents reported using a new technique or method, 17% (9/51) implemented the technology in a novel way, and 15% (8/51) used a centralized system; 9% (5/51) of the interventions were low-cost and mobile app–based and another 9% (5/51) were based in remote areas with constraint settings.

Smartphones and devices were the most commonly used component of digital health, used in 55% (28/51) of the interventions; websites and portals were employed in 25% (13/51), telehealth was used in 9% (5/51), special devices like e-stethoscopes and e-ultrasounds were used in 9% (5/51) of the projects, and 3% (2/51) used feature mobile phones and functional phones. When asked about barriers encountered in study implementation, 21% (11/51) reported that the population could not use the technology or mobile phones. Cost was the second most common barrier, reported by 16% (8/51) of the responders, followed by privacy concerns, which were raised by 12% (6/51) of the responders.

At the time of the study, 37% (19/51) of projects were fully operational, 26% (13/51) were completed, and 15% (8/51) were in the process of being scaled up. In regard to the development stage of the technologies, 44% (22/51) of the technologies used in the interventions were fully launched, 34% (17/51) were in the pilot phase, 16% (8/51) were in the conceptual stage, and only 3% (2/51) of the technologies employed were regularized.

Major barriers faced in implementing the project-specific technology were the cost of technology for 28% (14/51) of the projects, lack of skilled personnel for 18% (9/51), and internet connectivity issues for 16% (8/51). The major costs involved in most projects were equipment and infrastructure costs in 42% (21/51) of the projects, followed by human resource personnel for 23% (12/51) and field implementation for 18% (9/51). For storage, 73% (37/51) of projects used a local server for data storage while the rest (14/51, 27%) engaged cloud technology. MySQL was used in 66% (34/51) of projects as a server for the database, while 11% (6/51) used SQLite, 9% (5/51) used Firebase, and 9% (5/51) used Postgres Structure Query Language (PostgreSQL). Programming languages used include Java (15/51, 30%), C/C++ (12/51, 24%), PHP (10/51, 20%), Python and Java (4/51, 7% each), DOT Net and ROR (3/51, 6% each).

Users were required to undergo some training specific to the technology being used in 76% (39/51) of the interventions, and this training varied in terms of duration. The individuals and groups most likely to benefit from the interventions are health care providers (22/51, 44%), followed by global health organizations (17/51, 33%) and digital health administration (17/51, 33%). At the time of the survey, 74% (38/51) of the projects were being evaluated.

AI or ML is being used in 31% (16/51) of the projects, with most projects using TensorFlow (19/51, 38%), MATLAB (16/51, 31%), and General Architecture for Text Engineering (GATE; 8/51, 15%). Only 2% (1/51) of the projects were connected to a national database. Plans for scaling up were present in 59% (30/51) of the projects; 13% (7/51) were fully commercialized and 7% (4/51) were awaiting patent approval. Data were accessible to only the administrators in 76% (39/51) of the projects. Funding agencies and universities were given access to 5% (3/51) of the projects. All respondents agreed on having a data retention policy beyond the project duration. Ethics board approval before implementation was required for 51% (26/51) of the projects, 33% (17/51) required licensed software, and 6% (3/51) required approval from regulatory bodies before implementation. Written consent from participants was taken in 42% (21/51) of projects, 33% (17/51) utilized verbal consent, 7% (4/51) used electronic consent, 9% (5/51) each did not take consent or did not require it for the intervention.

Qualitative Results

The 4 SWOT themes that were identified are the following: (1) project novelty, (2) technology constraint, (3) catalyzing culture, and (4) workforce training duration.

Project Novelty

The results of the study show that the research population, including professionals in IT, engineering, public health, and digital health across different provinces, prefer project innovation to be focused on the domains of telehealth, AI diagnostics, and the digitization of EPI data. At present, the

projects are in different phases of development, such as exploratory, research, ideation, delivery, and others.

Technology Constraints

The professionals with diverse backgrounds (which include individuals from IT, engineering, public health, and digital health) face many barriers in implementing the technology due to the following: (1) patentability; (2) little deployment of mHealth apps and new digital tools in current clinical practices; (3) a communication gap between multiple stakeholders (which include technology entrepreneurs, investors, developers, researchers, and practicing physicians) because of the complexity and involvement of experts from diverse domains in digital health projects (the data shows that clinical experts were only involved in the implementation phase and were not represented or asked for input at the planning and initiation phases of the interventions); (4) lack of evidence on the validations of digital health devices and smartphone apps; (5) unavailability of regulatory frameworks.

On the other hand, the existing technologies currently used do not support large data sets, the replacement of gadgets is costly, and the tools for natural language processing for local languages are unavailable.

Catalyzing Culture

There is a communication gap when it comes to exchanging ideas and initiatives, primarily between physicians, technology experts, and researchers. Hence, there is a need for better communication to scale up and integrate digital health into the health care system in Pakistan. In addition, there should be advanced collaborative pathways to improve user wellbeing where the roles of investors, their responsibility, rights, and transparency should be clearly described. The result of this study shows that participants felt that there is a lack of guidance on how to utilize technologies to better suit the objectives of digital health interventions.

Workforce Training Duration

On-the-job training becomes pivotal in enabling individuals to acquire essential skills (related to technology advancement) for a meaningful contribution. For the teams involved in these projects to acquire a basic understanding before initiation, 2 to 3 training days were considered. Another important point raised was the need for digital health certificate programs for an average of 6 months, as these can advance one's career further.

SWOT Analysis

Profiles of the Projects

The set of projects selected for this study is representative of all the provinces in Pakistan. The academic sector is contributing significantly to digital health in Pakistan. There were a good number of projects which were funded internally through their parent organization; this is an important factor in increasing the chances of the scalability and sustainability of the intervention as self-reported by the participants. The interventions in these projects targeted the general health of the population, focusing on both communicable and noncommunicable diseases, and providing services to individuals from all socioeconomic groups. There was, however, less focus on subspecialty medicine in the interventions as most projects were found to revolve around family medicine and general health domains. The interventions under study did not use geospatial technology as part of their interventions for local health problems such as regional outbreaks, epidemics, and surveillance activities. The use of AI and ML was observed to be used at the rudimentary level, either due to a lack of big data sets or access to the main datasets.

A significant proportion of the interventions were based on smartphones, which are owned by less than one-third of the population [49]. Digital health is steadily growing as an industry in Pakistan, and there are a lot of opportunities to work with different health care providers in both programmatic and research domains as part of public health and clinical settings. Globally, digital health has been used to address domains like mental health and maternal and child health; however, in our SWOT analysis, there were limited studies in this domain and hence this could be a very important opportunity—perhaps even more relevant now due to the COVID-19 pandemic [50]. AI and ML are currently revolutionizing the digital health landscape across the world, and the Pakistani health care community needs to take advantage of such technologies to derive significant benefits for the population's health.

In our study findings, one of the major constraints in conducting and scaling up technology-based interventions and programs in Pakistan was the low availability of human resources with technology-specific skills in addition to restricted local training and capacity building opportunities. There is also a low trend of sharing data publicly due to multiple reasons, as reflected in the low percentage of studies published in the public domain, an issue that raises serious concerns. The technologies being used in most projects are old generation, and this can make stability and sustainability challenging to achieve. Currently, there is a trend in Pakistan to develop custom-made software in-house, thus limiting value for money and creativity. Further details are displayed in Multimedia Appendix 2.

Project Team Characteristics

The teams were diverse with professionals belonging to different disciplines, reflecting collaboration across different sectors. A good representation of IT professionals is central to making sure the projects are technologically sound. The results of the study show that professionals belonging to different industries prefer project novelty to be a focus in the domains of telehealth, AI diagnostics, and digitization of EPI data. At present, the projects are in various phases, including exploratory, research, ideation, proof of concept, and implementation.

There is a noticeably low representation of individuals with research and evaluation backgrounds in the project teams, and IT experts and clinical teams might not have the bandwidth to evaluate the impact of these projects. The basics of digital health and its importance can be introduced to students at the undergraduate and postgraduate levels in health care and IT disciplines. There is room for better collaboration across various sectors. Health care and IT professionals especially have an important role to play in shaping the future of digital health and therefore need to get involved. There was a significant proportion of projects designed entirely by individual developers. While this may have certain advantages, often these

developers lack the insight that health care professionals have regarding the needs of the patients and the health system in general.

The Technology Used in Current Projects

Most projects are employing advanced technologies to store data. MySQL is a popular, easy-to-use framework [51,52]. Most initiatives are currently saving data on local servers, which ensure data privacy. The local servers used for storage come with the drawback of high maintenance cost and high bandwidth. Also, these servers may not be practically usable when the interventions are scaled up. Projects studied are also employing interactive modes of communication to engage the users better.

Some projects are still using outdated programming frameworks that have been replaced worldwide, inhibiting the use of state-of-the-art analytical and decision support features such as Python and geospatial technology such as imaging, analytics, mapping tools, and AI and ML. These limitations can be overcome with the use of technology frameworks that consume lower bandwidth, capacity building for Python and other modern technical software languages, AI and MI, and cloud-based services. It was also noted that the majority of the end-users of the projects required some training before use. There is little to no cybersecurity and regulation of the technology used to gather the data, and little expertise to develop more secure systems. There is also a lack of an established mechanism to determine the effectiveness of training and capacity in the area of cybersecurity.

Project Evaluation

At the time of the survey, most of the projects were undergoing an evaluation process using a variety of tools, and a high proportion of them were being scaled up, which reflects the acceptability and success of digital health interventions in the society. Most of the participants consider the strengths of the project in terms of technology, focusing on the use of hospital information systems, AI models, patient monitoring and information tracking system, cloud base, time consultation, and square database in their existing projects. On the other hand, the participants consider a lack of resources, unavailability of technology solutions, and a lack of integration with allied systems as the major factors limiting project advancement and scalability. There is an extremely low ratio of manuscript publications in the digital health domain due to the lack of technical skills and ethical implications; hence, most of the relevant work is not published on scientific forums. Therefore, the majority of projects are only being reported on project websites and blogs, which may often exaggerate the impact of the interventions and may not be based on valid evidence or reality.

There are a lot of opportunities to employ AI and ML techniques in the evaluation of big data sets, as these tools are readily available in Pakistan. More projects need to be connected with national databases and registries. This will help in scaling up the projects from the implementation and commercialization perspective. Stringent monitoring and evaluation of the overall project, in addition to technology and clinical aspects, need to be improved. Furthermore, improvement in the reporting standards is also required to enhance the quality and efficacy of digital health-based interventions and programs. There is an increased reliance on conventional paper-based checklists and rudimentary tools lacking coverage of technology and health aspects.

Ethics in Digital Health

All our respondents agreed on having their digital health projects regulated by independent ethics committees, with access to data limited to relevant personnel only. The stakeholders were also mindful of the importance of data retention and archiving. Of the 51 projects examined, 10% (5/51) did not obtain consent from the end-users related to the project, which might include clinical or personal identifier data. Around half of the studies had ethical approval to implement the projects or the study. The majority of the participants felt the need for an independent ethics committee led by the local institution as opposed to a regional or public based ethical committee. However, one common critique of local ethical boards is the poor quality and the lack of adherence to national and international guidelines. Unfortunately, not following the standard guidelines might have ethical and legal implications not only for the study but to the parent institute as well. Hence, strict compliance with the ethical standards set by national and international standards is highly recommended.

There is also a pressing need at the national level to establish ethical frameworks, standardize consent norms, and identify the regulatory approvals needed before a project can be implemented and launched, especially when something is scaled at a larger level. The Ministry of Information Technology and Telecom in Pakistan appears to be the body with the relevant mandate. There is currently a limited national effort dedicated to addressing the ethical concerns that arise in digital health projects, such as data privacy, retention, and determining the situations when consent is not required.

Discussion

Principal Findings

This is the first study conducted in Pakistan that looks at digital health interventions at a national level using SWOT analysis, and that also examines the technology and ethical dimensions of digital health. In the baseline survey, both qualitative and quantitative data were collected using a mixed-methods strategy; therefore, SWOT analysis was the most suitable strategy to analyze the data and highlight the impact of different digital health initiatives in Pakistan, as it allowed for the identification of internal factors as well as external factors that influence the implementation of these initiatives.

Overall, digital health or the use of technology in all domains of health is an emerging factor globally, and especially in resource-constrained settings; it can be a powerful solution for improving health outcomes locally and at the grass-root levels. This study and SWOT analysis were conducted just before the global COVID-19 pandemic. However, the findings obtained resonate with the demand and improvement in digital health interventions and innovations in line with the health system. Since the COVID-19 pandemic, the role of and demand for

digital health has increased significantly, and in some scenarios, it is the only viable solution for moving forward while following a social distancing strategy. In Pakistan, the use of digital health technologies in handling the COVID-19 pandemic has surged, especially in the public sector; 3 notable interventions being implemented include: (1) real-time registries and dashboards to visualize and download positive cases and relevant data, (2) a COVID-19–specific telehealth portal where patients can consult a doctor online, (3) an SMS text messaging–based EHSAAS emergency cash transfer program providing financial support to citizens identified by the government's poverty criteria during the enforced COVID-19 lockdown [53–56].

Due to resource constraints, the current infrastructure of Pakistan still struggles with providing access to high-speed internet and smartphones to two-thirds of its population [46]. Furthermore, high budget requirements in the establishment and implementation of technology-based interventions, including human resources and infrastructure costs like software and servers, is a major barrier among countries under-budgeted for technology and health sectors such as Pakistan [53]. This limits the potential impact of technology-based interventions that could be carried out for a more extensive and socioeconomically diverse audience [46].

Our study revealed that 36% (18/51) of the current projects involve smartphones as a medium for service delivery, which is in line with Pakistan Telecommunication Authority (PTA) data which states a similar number of smartphone ownership. However, on closer inspection, most of the smartphone owners are concentrated in urban populations, sometimes with one person owning more than one smartphone [46]. Similarly, the nationwide Pakistan Demographic and Health Survey 2017-2018 revealed that 22.9% of urban households had an internet connection compared to 4.9% in rural areas [54]. While smartphones offer an interactive interface and allow for more complex web-based applications to be developed, the generalizability of these interventions or programs is limited due to the non-accessibility of smartphones to the majority of the population in present-day Pakistan. Hence, looking at the ground reality, any successful interventions with an excellent outreach would have to be generalizable to a majority of the population, even if it is not the most advanced technology.

Our SWOT analysis also highlighted that most hospitals, basic health care units, and tertiary centers do not have the capacity to incorporate digital health interventions, and therefore all the efforts being made in this area are fragmented and vertical. Electronic health records (EHRs) are an indispensable tool in aligning digital health projects with the existing health system [55,56]. A study done in the province of Khyber Pakhtunkhwa in 2017 revealed that of the 35 hospitals studied, only 1 (0.03%) fulfilled all the requirements of basic EHRs [57]. An eHealth readiness assessment in Pakistan reported low e-literacy in health care professionals and showed that the cost of technology is a major barrier faced by hospitals in introducing EHRs, reflecting the need for health care staff training on the proper usage of digital health tools and technology before technology implementation on a larger scale [61–63].

The projects in this study originated from various sectors, with equal representation from the public (17/51, 34%) and private (17/51, 34%) sectors. This is a novel approach, as previous research was less balanced (eg, in a study based in Bangladesh, the private sector was the primary contributor [44]). The active involvement of the public sector is very promising, as no sustainable impact can be made without the active participation of the public sector.

One major gap identified in the analysis was the lack of ethical and legal regulations at the national level. The Ministry of Health and the Ministry of Information Technology and Telecommunications in Pakistan have to work together to implement a policy for regulating the use of technology in the health sector and standardize the procedures for developing and executing digital health–based projects. However, health is a provincial matter in Pakistan while IT is a federal subject; this will present a challenge in this regard. Our results also showed that only 1 project was connected to a national database; this demonstrates the need for better collaboration and communication between the different stakeholders and the government toward sharing the data and unifying the currently fragmented system. Another barrier related to academia and digital health was the absence of undergraduate and postgraduate level degrees and courses in Pakistan. Several of the weaknesses and threats associated with human resource capacity are linked to the lack of any formal program in digital health in Pakistan. Lastly, as per our SWOT analysis findings, the main risks associated are (1) data storage and retention; (2) hidden costs due to storage, security, processing, archiving, analyzing, and cloud; (3) unorganized dataset leads to incompetent analytics and faulty trends and predictions; and (4) data privacy and security.

Strengths and Contribution of the Study

To the best of the authors' knowledge, this is the first study that looks into digital health interventions in Pakistan as a whole. Rather than just considering the PubMed published literature, the stakeholders with both published and unpublished projects were interviewed for the SWOT analysis to provide a more holistic picture of current digital health-related projects.

Limitations

This study was conducted by researchers from academic universities; thus, there is a possibility of low representation of projects from industry and public domains. In addition, only the interventions conducted in the last 5 years were taken into consideration, and the work done in this field before that might have been missed due to possible recall bias. The authors approached all potential stakeholders, but consent was denied in a few cases. The respondents self-reported whether their projects are or have been evaluated; however, this could not be verified by an independent source. Lastly, since digital health is a relatively new field in Pakistan, there are still some grey areas as to what interventions fall in the domain of digital health; therefore, some relevant work may have been missed. However, the investigators have tried their best to include all the interventions carried out in Pakistan in the domain of digital health in the last 5 years.

Conclusions

Digital health–based interventions are, slowly but steadily, being ushered into the existing health system of Pakistan. There are still significant hurdles, barriers, and roadblocks in the form of limited internet facilities, phone ownership, network coverage, unavailability of regulatory frameworks, data protection and security regulation, accessibility, affordability, and paper-based health records, limiting the types of technologies that can be utilized for effective interventions. However, despite all the challenges, digital health is steadily expanding through the efforts of multiple stakeholders in both the public and private sectors. It is difficult to say how effective these interventions have been, as not all interventions are being evaluated or published. The future for digital health does look bright, especially after the government's new initiative to digitalize the public sector in Pakistan [58].

Acknowledgments

We gratefully acknowledge the financial support of the Grand Challenges Research Fund (GCRF), UK. This research was carried out as part of the Capacity Building in Technology-Driven Innovation in Healthcare (CoNTiNuE) project, funded by the UK-GCRF program. We are also grateful to all the participants and facilitators of the conducted workshops. Thank you to the eHealth Association of Pakistan (eHAP) and Ms. Salima Gulzhar for their support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Resource list: details of all the projects, studies, and other grey literature considered for the strengths, weaknesses, opportunities, and threats (SWOT) analysis.
[XLSX File (Microsoft Excel File), 40 KB - jmir_v22i9e21691_app1.xlsx]

Multimedia Appendix 2
Strengths, weaknesses, opportunities, and threats (SWOT) analysis table.
[PDF File (Adobe PDF File), 178 KB - jmir_v22i9e21691_app2.pdf]

References

1. Dhingra D, Dabas A. Global Strategy on Digital Health. Indian Pediatr 2020 Apr 25;57(4):356-358. [doi: 10.1007/s13312-020-1789-7]
2. Organization W. mHealth. Use of appropriate digital technologies for public health: report by Director-General. 71st World Health Assembly provisional agenda item 2017;12:A71.
3. Global Market Insights, Inc. Digital Health Market Share Trends 2019-2025 Growth Forecast Report Internet. URL: https://www.gminsights.com/industry-analysis/digital-health-market [accessed 2020-01-23]
4. Organization W. Seventy-first world health assembly: digital health. Geneva: WHA 2018;7:71.
5. URL: https://www.forbes.com/sites/bernardmarr/2016/07/18/how-the-connected-car-is-forcing-volvo-to-rethink-its-data-strategy/#20cfeab13e8d [accessed 2020-01-25]
6. Mejia N. AI-Based Fraud Detection in Banking - Current Applications and Trends Internet. Emerj. URL: https://emerj.com/ai-sector-overviews/artificial-intelligence-fraud-banking/
7. URL: https://www.partnershiponai.org/ [accessed 2020-01-25]
8. Sophie P, Javier GP, Andrew C, Zsuzsanna V. From treatment to prevention: The evolution of digital healthcare. https://media.nature.com/original/magazine-assets/d42473-019-00274-6/d42473-019-00274-6.pdf. URL: https://www.nature.com/articles/d42473-019-00274-6 [accessed 2020-01-23] [WebCite Cache ID https://media.nature.com/original/magazine-assets/d42473-019-00274-6/d42473-019-00274-6.pdf]
9. Merchant RK, Inamdar R, Quade RC. Effectiveness of Population Health Management Using the Propeller Health Asthma Platform: A Randomized Clinical Trial. J Allergy Clin Immunol Pract 2016;4(3):455-463. [doi: 10.1016/j.jaip.2015.11.022] [Medline: 26778246]
10. Merchant R, Szefler SJ, Bender BG, Tuffli M, Barrett MA, Gondalia R, et al. Impact of a digital health intervention on asthma resource utilization. World Allergy Organ J 2018;11(1):28 [FREE Full text] [doi: 10.1186/s40413-018-0209-0] [Medline: 30524644]
11. Shull JG. Digital Health and the State of Interoperable Electronic Health Records. JMIR Med Inform 2019 Dec 01;7(4):e12712 [FREE Full text] [doi: 10.2196/12712] [Medline: 31682583]
12. Tokai Y, Yoshio T, Aoyama K, Horie Y, Yoshimizu S, Horiuchi Y, et al. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. Esophagus 2020 Jan 24:1-7. [doi: 10.1007/s10388-020-00716-x] [Medline: 31980977]

13. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020 Jan 1;577(7788):89-94. [doi: 10.1038/s41586-019-1799-6]
14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
15. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
16. Honkoop PJ, Simpson A, Bonini M, Snoeck-Stroband JB, Meah S, Fan Chung K, et al. MyAirCoach: the use of home-monitoring and mHealth systems to predict deterioration in asthma control and the occurrence of asthma exacerbations; study protocol of an observational study. BMJ Open 2017 Jan 24;7(1):e013935 [FREE Full text] [doi: 10.1136/bmjopen-2016-013935] [Medline: 28119390]
17. myAirCoach: mHealth assisted self-management in patients with uncontrolled asthma, a randomized control trial. URL: https://erj.ersjournals.com/content/54/suppl_63/PA745.abstract [accessed 2020-03-22] [WebCite Cache ID =]
18. Asthma UK. myAirCoach Internet. URL: https://www.asthma.org.uk/Testing/research2/future/myaircoach/ [accessed 2020-02-04]
19. American Hospital Association. Fact Sheet: Telehealth | AHA Internet. URL: https://www.aha.org/factsheet/telehealth [accessed 2020-01-23]
20. ASPE. Report to Congresshealth and Telemedicine Internet. 2016. URL: https://aspe.hhs.gov/pdf-report/report-congress-e-health-and-telemedicine [accessed 2020-03-20]
21. URL: https://www.mdedge.com/ccjm/article/189759/practice-management/telemedicine-past-present-and-future [accessed 2020-01-23]
22. Association A. Telehealth: Helping hospitals deliver cost-effective care. Issue Brief 2016;30:7.
23. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 3;14(1):1-8. [doi: 10.1186/s12992-018-0424-z]
24. Long L, Pariyo G, Kallander K. Digital Technologies for Health Workforce Development in Low- and Middle-Income Countries: A Scoping Review. Glob Health Sci Pract 2018 Dec 10;6(Suppl 1):S41-S48 [FREE Full text] [doi: 10.9745/GHSP-D-18-00167] [Medline: 30305338]
25. Githinji S, Kigen S, Memusi D, Nyandigisi A, Mbithi AM, Wamari A, et al. Reducing stock-outs of life saving malaria commodities using mobile phone text-messaging: SMS for life study in Kenya. PLoS One 2013 Jan;8(1):e54066 [FREE Full text] [doi: 10.1371/journal.pone.0054066] [Medline: 23349786]
26. Organization W. Preventing stock-outs of antimalarial drugs in sub-Saharan Africa: Novartis's SMS for life. Preventing stock-outs of antimalarial drugs in sub-Saharan Africa: World Health Organization; 2013.
27. Pai M, Yadav P, Anupindi R. Tuberculosis control needs a complete and patient-centric solution. Lancet Glob Health 2014 Apr;2(4):e189-e190 [FREE Full text] [doi: 10.1016/S2214-109X(14)70198-6] [Medline: 25103049]
28. Batra S, Ahuja S, Sinha A, Gordon N. eCompliance: Enhancing tuberculosis treatment with biometric and mobile technology. 2012 Feb 28 Presented at: Proceedings of M4D New Delhi, India 2012;28(29); 2012 February 2012; Operation ASHA, New Delhi, India p. 36-40.
29. Snidal SJ, Barnard G, Atuhairwe E, Ben AY. Use of eCompliance, an innovative biometric system for monitoring of tuberculosis treatment in rural Uganda. Am J Trop Med Hyg 2015 Jun;92(6):1271-1279 [FREE Full text] [doi: 10.4269/ajtmh.14-0413] [Medline: 25846297]
30. Khan MAH, Cruz VDO, Azad AK. Bangladesh's digital health journey: reflections on a decade of quiet revolution. WHO South East Asia J Public Health 2019 Sep;8(2):71-76 [FREE Full text] [doi: 10.4103/2224-3151.264849] [Medline: 31441440]
31. Biswas A. Shifting paradigm of maternal and perinatal death review system in Bangladesh: A real time approach to address sustainable developmental goal 3 by 2030. F1000Res 2017 Jul 12;6:1120. [doi: 10.12688/f1000research.11758.1]
32. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15(1):1-11. [doi: 10.1186/s12883-015-0471-5]
33. Hashmi NR, Khan SA. Interventional study to improve diabetic guidelines adherence using mobile health (m-Health) technology in Lahore, Pakistan. BMJ Open 2018 May 31;8(5):e020094. [doi: 10.1136/bmjopen-2017-020094]
34. Moazam F, Shekhani S. Why women go to medical college but fail to practise medicine: perspectives from the Islamic Republic of Pakistan. Med Educ 2018 Mar 06;52(7):705-715. [doi: 10.1111/medu.13545]
35. Ltd G. Media Internet. Sehat Kahani. URL: https://sehatkahani.com/media/ [accessed 2020-01-23]
36. URL: https://edoctor.net.pk/ [accessed 2020-01-25]
37. Zaidi S, Shaikh SA, Sayani S, Kazi AM, Khoja A, Hussain SS, et al. Operability, Acceptability, and Usefulness of a Mobile App to Track Routine Immunization Performance in Rural Pakistan: Interview Study Among Vaccinators and Key Informants. JMIR Mhealth Uhealth 2020 Feb 13;8(2):e16081 [FREE Full text] [doi: 10.2196/16081] [Medline: 32053115]
38. Zareen S, Mursalin SM. Managing Dengue Fever by Using the One Health Approach and Electronic Surveillance. OJPHI 2015 May 08;7(1):1-1. [doi: 10.5210/ojphi.v7i1.5965]

39. Yousafzai MT, Qamar FN, Shakoor S, Saleem K, Lohana H, Karim S, et al. Ceftriaxone-resistant Salmonella Typhi Outbreak in Hyderabad City of Sindh, Pakistan: High Time for the Introduction of Typhoid Conjugate Vaccine. Clin Infect Dis 2019 Feb 15;68(Suppl 1):S16-S21 [FREE Full text] [doi: 10.1093/cid/ciy877] [Medline: 30767003]
40. Mir F, Mahmood F, Siddiqui AR, Baqi S, Abidi SH, Kazi AM, et al. HIV infection predominantly affecting children in Sindh, Pakistan, 2019: a cross-sectional study of an outbreak. Lancet Infect Dis 2020 Mar;20(3):362-370. [doi: 10.1016/S1473-3099(19)30743-1] [Medline: 31866326]
41. Bentahar O, Cameron R. Design and Implementation of a Mixed Method Research Study in Project Management. Electronic Journal of Business Research Methods 2015;13(1):1-13 [FREE Full text]
42. edu Internet. (PDF) Yin Case Study Research Design and Methods | Sergio Gonzales Rodriguez - Academia. URL: https://www.academia.edu/32420108/Yin_Case_Study_Research_Design_and_Methods [accessed 2020-01-23]
43. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
44. Ahmed T, Lucas H, Khan AS, Islam R, Bhuiya A, Iqbal M. eHealth and mHealth initiatives in Bangladesh: a scoping study. BMC Health Serv Res 2014;14:260 [FREE Full text] [doi: 10.1186/1472-6963-14-260] [Medline: 24934164]
45. URL: https://gs.statcounter.com/os-market-share/mobile/pakistan [accessed 2020-01-27]
46. URL: https://www.pta.gov.pk/en/telecom-indicators [accessed 2020-01-27]
47. World HO. Everybody's business strengthening health systems to improve health outcomes: WHO's frmaework for action. Geneva: World Health Organization; 2007:978-992.
48. URL: https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-188 [accessed 2020-01-27]
49. AfterAccess: I. 0). Colombo: LIRNEasia. LIRNEasia; 2019.
50. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848. [doi: 10.2196/18848]
51. URL: https://www.mysql.com/customers/ [accessed 2020-02-23]
52. Walia E, Gill E. A framework for web based student record management system using PHP. International Journal of Computer Science and Mobile Computing 2014;3(8):33 [FREE Full text]
53. Pakistan Economic Survey 2017-18 Internet. Health and Nutrition. URL: http://www.finance.gov.pk/survey/chapters_18/11-Health.pdf [accessed 2020-01-23]
54. National IOPS(PICF. 2019. Pakistan Demographic Health Survey 2017;Islamabad, Pakistan, Rockville, Maryland, USA:1-573.
55. Au L, Oster A, Yeh GH, Magno J, Paek HM. Utilizing an electronic health record system to improve vaccination coverage in children. Appl Clin Inform 2010;1(3):221-231 [FREE Full text] [doi: 10.4338/ACI-2009-12-CR-0028] [Medline: 23616838]
56. Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform 2016 May 20;Suppl 1:S48-S61. [doi: 10.15265/IYS-2016-s006] [Medline: 27199197]
57. Qazi U, Haq M, Rashad N, Rashid K, Ullah S, Raza U. Availability and use of in-patient electronic health records in low resource setting. Comput Methods Programs Biomed 2018 Oct;164:23-29. [doi: 10.1016/j.cmpb.2018.06.005] [Medline: 30195429]
58. Tribune. com. pk. URL: https://tribune.com.pk/story/2112360/8-digital-pakistan-pm-imran-addresses-launch-ceremony/ [accessed 2020-04-16]

Abbreviations

AI: artificial intelligence
COVID-19: coronavirus disease 2019
DHIS 2: District Health Information Software 2
eHAP: eHealth Association of Pakistan
EPI: expanded program on immunization
HIV: human immunodeficiency virus
IT: information technology
LMICs: low and middle-income countries
MeSH: medical subject headings
mHealth: mobile health
ML: machine learning
NCDs: noncommunicable diseases
NHS: National Health Service
PostgreSQL: Postgres Structure Query Language
PTA: Pakistan Telecommunication Authority
SQLite: Structured Query Language
SWOT: strengths, weaknesses, opportunities, and threats
WHO: World Health Organization

Edited by G Eysenbach; submitted 24.06.20; peer-reviewed by J Li, M Adly, A Adly; comments to author 08.07.20; revised version received 13.07.20; accepted 14.07.20; published 03.09.20.

Please cite as:
Kazi AM, Qazi SA, Ahsan N, Khawaja S, Sameen F, Saqib M, Khan Mughal MA, Wajidali Z, Ali S, Ahmed RM, Kalimuddin H, Rauf Y, Mahmood F, Zafar S, Abbasi TA, Khoumbati KUR, Abbasi MA, Stergioulas LK
Current Challenges of Digital Health Interventions in Pakistan: Mixed Methods Analysis
J Med Internet Res 2020;22(9):e21691
URL: https://www.jmir.org/2020/9/e21691
doi:10.2196/21691
PMID:32880584

Original Paper

Examining Patterns of Information Exchange and Social Support in a Web-Based Health Community: Exponential Random Graph Models

Xuan Liu[1*], PhD; Shan Jiang[2*], PhD; Min Sun[1], MSc; Xiaotong Chi[1], MSc

[1]East China University of Science and Technology, Shanghai, China

[2]University of Massachusetts Boston, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Xuan Liu, PhD
East China University of Science and Technology
130 Meilong Rd East China University of Science and Technology
Shanghai
China
Phone: 86 86 21 64253177
Email: xuanliu@ecust.edu.cn

Abstract

Background: Although an increasing number of studies have attempted to understand how people interact with others in web-based health communities, studies focusing on understanding individuals' patterns of information exchange and social support in web-based health communities are still limited. In this paper, we discuss how patients' social interactions develop into social networks based on a network exchange framework and empirically validate the framework in web-based health care community contexts.

Objective: This study aims to explore various patterns of information exchange and social support in web-based health care communities and identify factors that affect such patterns.

Methods: Using social network analysis and text mining techniques, we empirically validated a network exchange framework on a 10-year data set collected from a popular web-based health community. A reply network was extracted from the data set, and exponential random graph models were used to discover patterns of information exchange and social support from the network.

Results: Results showed that reciprocated information exchange was common in web-based health communities. The homophily effect existed in general conversations but was weakened when exchanging knowledge. New members in web-based health communities tended to receive more support. Furthermore, polarized sentiment increases the chances of receiving replies, and optimistic users play an important role in providing social support to the entire community.

Conclusions: This study complements the literature on network exchange theories and contributes to a better understanding of social exchange patterns in the web-based health care context. Practically, this study can help web-based patients obtain information and social support more effectively.

(J Med Internet Res 2020;22(9):e18062) doi:10.2196/18062

KEYWORDS

web-based health communities; information exchange; social support; ERGM

Introduction

Background

The rapid evolution of the internet and related technologies has created advanced virtual platforms that allow popular and pressing health topics to be discussed on the web. Unlike in the past, patients now may seek help from experts, share stories with similar patients from across the world to obtain emotional support, and keep themselves informed with latest updates about related issues [1]. The resulting environment is often referred to as web-based health communities, where users can share relevant experiences about diseases, physical conditions, and therapeutic schedules. Users can also consult specialists and

seek opinions from experts [2]. Various web-based health communities such as PatientsLikeMe, CureTogether, DailyStrength, and Disaboom are emerging. In these platforms, patients can read stories shared by other peers, find information regarding diseases, post a new thread to initiate a conversation, and reply to others' messages to provide feedback. During these interactions, information exchange and social support occur [3]. Users can provide information by leaving messages, which generates a dynamic information exchange procedure [4]. Meanwhile, obtaining social support, especially emotional relief, from communicating with others has become one of the major purposes for patients with diseases to join web-based communities [5]. An increasing number of studies have attempted to understand how individuals interact with others in web-based health communities. For example, research has shown that patients with similar disease stages or health status are more likely to develop friendship on the web [6]. Urban users tend to provide social support to rural participants in web-based health communities [7]. A recent study has found that web-based interaction between doctors and patients does not decrease the effectiveness of information exchange compared with face-to-face communication [8]. However, studies focusing on understanding individuals' patterns of information exchange and social support in web-based health communities are still limited.

The aim of our research is to explore various patterns of information exchange and social support in web-based health care communities and identify factors that affect such patterns. On the basis of the network exchange framework [9], we discuss how patients' social interactions develop into social networks. We empirically validated the network exchange framework based on data sets collected from a leading web-based health community in China. The results from our analyses indicate that reciprocated information exchanges are likely to develop between patients, especially between web-based members who have different roles and members who are web-based friends. Some patients are more likely to receive social support, especially when they are new to the community and when they express polarized sentiment in messages. In addition, these patterns could vary depending on the topics being discussed by patients. To the best of our knowledge, this study is the first to validate the network exchange framework in the web-based health community context. This study is also the first to perform a stratified analysis of user sentiment to understand the complex information exchange and social support patterns between patients with various sentiments.

Related Work

Information Exchange and Social Support

In web-based health communities, patients can exchange information by sharing their experiences in overcoming illness, transferring medical knowledge to one another, and providing information regarding health care resources. Different communities usually specialize in different aspects. Some communities focus on specific types of disease, whereas others can provide unique services to patients. For example, in DailyStrength, patients with anxiety disorder can find a support group suggested by the community to discuss how to overcome stress or other disorders. In Tianmijiayuan, patients with diabetes and their families can post messages to share how they deal with different stages of diabetes in the long term. PatientsLikeMe provides a matching service for patients based on their profiles to quickly get in touch with other peers who have experienced or are experiencing similar diseases. Members of web-based health communities can benefit from their collective knowledge and skills by exchanging information among each other [10]. Information exchange is also a critical component in the development of many web-based communities [11]. Information exchange in web-based health communities can enhance effective communication between medical experts and patients by improving teamwork [12].

Previous studies have shown that social support plays a crucial role in helping individuals improve their health status or treat psychological problems [13,14]. The benefits of web-based social support come in 2 forms: informational support and emotional support [15].

By asking questions regarding health concerns on the web, users can obtain professional knowledge from experts. Moreover, patients may learn experiences from others who share similar diseases. Although such informational benefits can also come from information exchanges, obtaining social support differs from information exchange mainly in the patterns of interaction. Unlike information exchanges where the interaction is mutual, social support can be unilateral [16]. Many users provide informational support without expecting any return due to empathy [17].

Emotional support mainly comes from web-based users who share similar disease experiences or from friends of such patients. Patients can talk about what difficulties they have overcome, what they did to recover, and can encourage peers to be optimistic and fight against the disease. Some patients have reported that they received more understanding from web-based strangers than they did from offline families or friends [18]. This support could significantly enhance patients' emotional well-being [19]. As such, web-based health communities have become a platform for many patients to seek and provide emotional support [20].

Previous research has implied that social network is one of the key antecedents of social support and information exchange [15]. However, understanding information exchanges and social support in web-based health communities from a social network perspective has received limited research attention.

Network Perspective of Web-Based Health Communities

Members of web-based health communities develop a social network through communication. Such social networks provide users an opportunity to exchange information and seek social support in web-based communities [15,21]. By using social network analysis, attributes of nodes (ie, users) and relationships between users can be modeled and examined.

Recently, a network exchange framework has been proposed to theorize how social interactions between individuals aggregate into a social network [9]. The network exchange framework tries to explain social exchange from a network perspective by combining the social exchange theory and the network theory.

Social interactions are viewed as processes of exchanging resources such as information, knowledge, and emotional well-being. Individual characteristics play an important role in such exchange processes. When the pair of individuals is part of a larger network, the exchange processes can be further influenced by their positional configurations, such as their social connections with others [22]. The network exchange framework has been used to explain the formation of information exchange networks in web-based communities organized around various topics, such as software [9] and automobile [23]. However, we do not find studies that apply a network exchange framework to investigate information exchange patterns in web-based health communities. This endeavor is important because web-based health communities are distinct from other more traditional web-based communities in that information being exchanged is usually sensitive, private, and requires professional medical knowledge. Handling such information gives rise to special community norms that could lead to unique social exchange patterns. As evidenced by previous studies, patterns of social interactions could vary greatly across web-based communities in different contexts [9,24].

Network Exchange Framework

In this study, we develop our research hypotheses regarding information exchange in web-based health communities based on a network exchange framework for several reasons. First, we deem information and knowledge exchanged between patients as resources in web-based health communities, and thus, social exchange theory can help explain the patterns of information exchange. Second, information exchange between individuals aggregates into a network between patients because patients typically interact with multiple peers. Therefore, adopting a network perspective would better explain information exchange and social support between individual patients.

In our model, we include 3 major structural tendencies that comprise network formation in the network exchange framework: direct reciprocity, indirect reciprocity, and preferential attachment [9].

Direct Reciprocity

According to the reciprocity principle [25], individuals expect to receive information back after providing information to others. In web-based health communities, obtaining useful health care knowledge is regarded as one of the major objectives when users join a community [26]. As individuals' health care needs are usually complicated, users may ask further questions to obtain more information after receiving initial responses. Reciprocated information exchange develops in this way. Some users enjoy contributing their knowledge and receive thankful responses from others. Users are also likely to provide assistance to others who have provided them with support and then realize their intrinsic motivation [27,28]. Therefore, we propose the following hypothesis:

- Hypothesis 1: Reciprocated information exchange is likely to develop in web-based health communities.

The pattern of direct reciprocity could further manifest in subgroups of web-based community members. Social interaction helps people with similar characteristics to become acquainted and build trust with each other [29,30]. Homophily, the tendency for individuals to be attracted by others with similar characteristics, is an important dimension in social networks [31]. Homophily commonly occurs based on geographic and demographic characteristics such as race, religion, age, gender, residence, marital status, and interests [32,33]. In a web-based environment, similarity between users increases the frequency of their interactions [34]. A study reported that patients with similar health conditions and treatments are likely to develop friendships [6]. Another common user-related attribute in web-based health communities is user type, such as doctor, family, or patient, which users report when they join the community. As we expect homophily to exist in web-based health information exchange, we propose the following hypothesis:

- Hypothesis 2: Reciprocated information exchange is likely to develop between users who share similar concerns in web-based health communities.

Typically, web-based health community members can also become web-based friends so that they can send private messages or keep updated about each other's posts. The possibility of sending private or offline messages may decrease their visible web-based conversations. However, being capable of following friends' posts may increase their chances of reading friends' messages and initiating conversations. Furthermore, web-based friendship could develop into relational capital [35], which becomes the basis for information exchange to occur. We propose the following hypothesis for empirical tests to understand the combined effects:

- Hypothesis 3: Reciprocated information exchange is likely to develop between users who are web-based friends.

Indirect Reciprocity

Indirect reciprocity refers to returning an information exchange but not to the original provider [36]. Indirect reciprocity can be observed in web-based health communities because communicating health information usually requires specialized knowledge, but the expertise of patients is uneven. When a patient receives informational help from a knowledge provider, the patient may not be able to return the favor due to limitations in expertise. Instead, the patient may choose to provide help to others in the network as they could feel that the help is from the community as a whole, and they are willing to return the favor to the same community [37]. Previous studies have also found that new participants who received help tend to remain in the community to help others [38]. Therefore, we propose the following hypothesis:

- Hypothesis 4: Patients who receive social support tend to provide support to others who are not necessarily the support provider.

Preferential Attachment

Preferential attachment refers to a process in which a new node tends to establish connections with existing nodes that already possess many connections [39]. In the context of web-based health communities, preferential attachment translates to the tendency that a patient who is already involved in many

web-based social interactions is likely to receive further replies. This is intuitive because highly active members contribute more and influence more people in need [40]. Such contributions are visible to the entire community, and as a rewarding mechanism, the active members may receive more help in terms of incoming social support in the future. Therefore, we propose the following hypothesis:

- Hypothesis 5: Highly active users are more likely to receive replies as social support.

Another dimension of preferential attachment in the web-based health community context is connections with new members joining the community. Contrary to traditional preferential attachment notions, we do not expect new members to be able to select which nodes to form attachments with. This is because most new patients join a community to be helped, not to help, at least in the initial periods [26]. It is the subsequent replies to the new patient that initiate social support. In web-based health care communities, such subsequent replies are likely to occur. Although the new member has not yet contributed to the community, existing members can benefit from providing support to the new member insofar as the new member becomes part of the community and adds value through social capital to the community network [35]. As such, we expect that existing community members have greater motivation to provide social support to new members.

- Hypothesis 6: In web-based health communities, new patients are more likely to receive replies as social support.

Another factor that could affect preferential attachment in web-based health communities is sentiment. Previous studies have found that users with polarized sentiment tend to receive more attention on the web. For example, expressing positive emotions helps peers improve psychological and physical health conditions [41,42]. In web-based health communities, patients may feel more comfortable getting in touch with peers who are optimistic and show positive emotions. Negative emotions attract attention in another way. In web-based health communities, many individuals are inclined to help others avoid negative feelings, such as shame, guilt, or indebtedness, especially after they receive help from others [43]. Patients expressing negative emotions are often those who have a disease or are experiencing loss, and are in need of help from peers. Therefore, social support could also go toward patients with negative moods. Overall, we expect that users who express polarized sentiment (either positive or negative) are more likely to receive attention and hence receive more replies in web-based health communities.

- Hypothesis 7: In web-based health communities, patients with polarized sentiment are more likely to receive replies as social support.

A related question is who is providing social support to the users with polarized sentiment. On the one hand, homophily plays an important role in social networking [6,32,33], and we expect that patients with overall similar sentiment valence are likely to make friends and talk to each other very often. On the other hand, the web-based health community is a platform where patients not only make friends but also help other strangers voluntarily [5]. Without being friends or knowing someone, a patient who has gone through the most difficult time could be willing to help someone who is still suffering. Meanwhile, patients in a negative mood may seek emotional support from peers who seem to be optimistic. Therefore, in addition to the homophily effect, we also expect that users with opposite sentiments are likely to leave replies to each other. The following set of hypotheses is proposed:

- Hypothesis 8a: Patients are likely to receive replies from peers with similar sentiment valence.
- Hypothesis 8b: Patients are likely to receive replies from peers with opposite sentiment valence.

Methods

Data

To test our hypotheses, we collected data from Tianmijiayuan [44], a leading web-based diabetes community in China where patients, doctors, and relatives participate in various activities. It was established in 2005 and had 247,638 members in 2018. It is one of the largest and the most active web-based nonprofit Chinese diabetes communities, targeting individuals with diabetes and helping them share information about diabetes, exchange experiences of diabetes treatment, seek emotional support, and make friends with people who are facing similar diabetic conditions. From the entire forum, we extracted users' postreply networks as well as all the textual posting content and publicly available personal information of users, such as user type and web-based friendships. Data collection was performed using a Java web crawler, with a time range from 2005 to 2015.

Tianmijiayuan has separate subforums for different discussion topics. The most popular (in terms of the number of postings) ones include *Diabetes Knowledge*, *Communications Area for Diabetics*, and *Diabetic's Life*. The discussions in Diabetes Knowledge are usually related to symptoms and diagnoses of different types of diabetes, patients' diet and exercise, and diabetes news. Users can make friends and participate in community activities in the Communications Area for Diabetics. In addition, they can publish their own photography, life insights, and advice for the community in the Diabetic's Life subforum. To examine whether the information exchange and social support patterns vary depending on the topic of discussion, we also performed a separate analysis on each subforum.

Operationalization of Nodal Attributes

The following nodal attributes were modeled in our study.

Individual Type

Upon registration, users choose the type of their identity as one of the following: doctors, patients' family members, patients with type 1/2/X, web service staff, or other.

Activity Level

Tianmijiayuan [44] tracks a user's number of posts, replies, web-based time, peer reviews, and numerous other factors. These factors are integrated as a numerical score to represent users' level of activity. Users with higher scores are considered

active users. We collected this information, and users whose scores ranked among the top 25% were coded as highly active users. For robustness tests, we changed this threshold value to 20%, 23%, 27%, and 30% to examine how this operationalization affects the results (seethe *Robustness Tests* section).

Registration Time

We classified users as long-time users or new users based on their registration time. The number of months since registration was calculated for each user, and users in the bottom 25% of registration length were coded as new users. For robustness tests, we changed this threshold value to 20%, 23%, 27%, and 30% to examine how this operationalization affects the results (see the *Robustness Tests* section).

Emotion

Sentiment analysis was performed to determine each user's overall sentiment in the data set [45]. Specifically, a text analysis program, TextMind, was employed to assess users' sentiments on Tianmijiayuan [44]. It can identify the frequency of words associated with different emotions when users express opinions in community discussions. TextMind has been used in previous research to analyze emotional expressions in Chinese texts [46]. On the basis of the frequency of emotion-related words expressed by users in the entire forum, we found that approximately 5% of users used more negative words than positive words. These users were identified as pessimistic users with negatively polarized sentiments. An equal number of users were identified as optimistic users who used more positive words than negative words (the top 5% users with the highest frequency of positive words were selected). The remaining users did not have extremely high proportion of positive or negative words and were identified as sentiment neutral users.

Table 1 summarizes the operationalization of the nodal attributes of the users.

Table 1. Operationalization of nodal attributes.

Node attribute	Type	Measuring method
User type, %	Categorical variable	• 1-Users with type 1 diabetes, 23.7 • 2-Users with type 2 diabetes, 58.3 • 3-Users with type X diabetes, 3.6 • 4-Family members, 7.0 • 5-Doctors, 0.7 • 6-Web service staff, 1.1 • 7-Others, 5.6
Activity level	Binary categorical variable	• 1-Highly active users • 0-Other users
Registration time	Binary categorical variable	• 1-New users • 0-Other users
Emotion	Categorical variable	• 2-Optimistic users • 1-Pessimistic users • 0-Neutral users

Network Tie and Dichotomization

In this study, the extracted postreply network was used as the base network. If a user replied to another user's thread post or reply post, a network tie was developed. The number of ties was counted as the network tie intensity.

Network dichotomization was then performed based on the threshold values of the tie intensity. According to a previous study [47], the threshold values were determined as the mean tie intensity plus one standard deviation.

Exponential Random Graph Model

Exponential random graph model (ERGM) can simultaneously model structural relationships between nodes and the effects of nodes' individual attributes on network formation [7,48,49]. The research hypotheses in our study involve various nodal attributes of web-based patients (eg, sentiment and activity level) and structural relationships between them (eg, receiving replies and reciprocating replies). With ERGM, all the complex interactions of these nodes, nodal attributes, and network ties can be incorporated simultaneously in the same model.

In ERGM, the observed network is represented as $Y=\{Y_{ij}\}$, where Y_{ij} indicates whether there is a tie between nodes i and j ($Y_{ij}=1$) or not ($Y_{ij}=0$). The ERGM generates random networks based on hypothesized network patterns (ie, configurations) and compares the generated network with the actual observed network. The more similar they are, the more likely the hypothesized network patterns exist in the actual network. The general mathematical formulation of the ERGM is as follows:

where the summation is over all configurations A, y represents one kind of particular network graph y, and  is the parameter corresponding to the configuration A. is the network statistic corresponding to configuration A,  =1 if the configuration is observed in the network y and is 0 otherwise, and k is a normalizing quantity that ensures that (1) is a proper probability

distribution [50]. ERGM estimates parameters θ associated with each configuration, and positive and significant parameters indicate that corresponding network patterns are highly likely to occur in the network [51,52].

To test our hypotheses with ERGM, we transformed our hypotheses into network patterns. Multimedia Appendix 1 shows the hypotheses and the illustration of their network patterns.

Results

ERGM Results

Table 2 summarizes the estimated parameters and *P* values for all configurations. If a parameter is positive and significant, it indicates that the corresponding network pattern is more likely to develop than random chance [50]. During the initial tests, we found that the inclusion of a configuration for H4 (2-path) always resulted in model degeneracy [52]. This indicates that the pattern of indirect reciprocity hardly existed in the dichotomized postreply network. Therefore, H4 was not supported, and we excluded this network configuration from further tests.

In the subsequent section, for each hypothesis, we discuss our findings on the entire forum, and then, we compare the observations with the results in the subforums to examine how the patterns could vary depending on the topics of discussion. We deem a hypothesis to be supported only if it is supported in at least three tests.

Table 2. Results for exponential random graph model tests.

Configuration	Entire forum (sample size=1528)		Diabetes Knowledge (sample size=1188)		Communications Area for Diabetics (sample size=455)		Diabetic's Life (sample size=376)	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
H1: reciprocity	3.850	<.001	3.153	<.001	3.509	<.001	3.684	<.001
H2: type	0.240	<.001	−0.139	<.001	0.188	<.001	−0.287	<.001
H3: friend	3.473	<.001	3.712	<.001	0.019	.007	0.010	.14
H5: active_user	0.012	.76	−0.220	<.001	−0.176	.002	−0.411	<.001
H6: new_user	0.409	<.001	−0.110	0.004	−0.400	<.001	−0.338	.03
H7: optimistic	0.289	<.001	−1.383	<.001	−1.050	<.001	−0.819	<.001
H7: pessimistic	0.144	.045	0.692	<.001	−2.214	.04	−0.596	.21
H8a: opti-opti	−0.332	.23	1.294	0.03	0.829	.21	−0.100	.92
H8a: pessi-pessi	−0.140	.71	−0.180	0.45	N/A[a]	<.001	N/A	<.001
H8b: opti-pessi	0.168	.51	0.281	0.17	N/A	<.001	N/A	<.001
H8b: pessi-opti	−0.218	.48	1.345	0.004	N/A	<.001	N/A	<.001

[a]N/A: not applicable.

Hypotheses Testing Results

First, we found positive and significant coefficients for the *reciprocity* configuration in the entire forum as well as 3 popular subforums, indicating that directly reciprocated information exchange was common in web-based health communities. This observation conforms to the reciprocity principle that individuals are willing to return exchanges in favor [25,53]. In web-based health communities, patients appreciate the help received from others, and gratitude is expressed in many such reciprocated messages. In addition, we also observed that a number of patient pairs reciprocated replies in different threads, especially in the Diabetes Knowledge subforum. This indicates that patients are also willing to return favors to those from whom they have received support before. In summary, H1 was supported.

A positive and significant parameter was observed for the *type* configuration in the entire forum, indicating that users of the same type were more likely to reciprocate messages overall. However, the effect was negative and significant in the *Diabetes Knowledge* and *Diabetic's Life* subforums. This observation implies that conversations between users of different types were more common when the discussion topics were relevant to disease knowledge (diabetes) or personal life. For example, it is very likely that diabetes patients obtain information from doctors in the *Diabetes Knowledge* subforum. Moreover, when sharing personal life with web-based peers, users may be less concerned about whether others are in the same stage of diabetes as them. Note that our finding does not imply low chances of communication between any specific pair of user types in the subforums (eg, reciprocated ties specifically between two patients with type-2 diabetes in *Diabetes Knowledge* subforum was not tested). Instead, our finding simply implies that, overall, there was more reciprocated communication between users of different types in the 2 subforums. As a result, H2 was supported in the entire forum but not in the *Diabetes Knowledge* and *Diabetic's Life* subforums.

A positive and significant parameter was observed for the *friend* configuration in the entire forum as well as in the *Diabetes Knowledge* and *Communication* subforums, indicating that web-based friends were very likely to exchange information frequently with each other when discussing diabetes knowledge. Being web-based friends can increase one's attention and motivation to reply to health-related posts from other community members. In addition, patients were able to obtain some timely

health-related information and show empathy to others through this kind of virtual friendship [17]. This effect was not significant in *Diabetic's Life* subforum possibly due to the fact that when sharing personal life with web-based peers, users may be less concerned about whether others are their virtual friends. Overall, H3 was supported.

The *active_user* configuration was negative and significant in all subforums. Note that the activity level of a user was evaluated based on the user's log-in time and number of messages posted by the users in our data set. Therefore, our observation indicates that highly active users may stay on the web for a long time and leave many replies, but they may not necessarily receive an equally large number of replies back. This observation is different from previous findings that "popular friends get more friends" [6,40] but is consistent with prior research where preferential attachment was found to be in the opposite direction in knowledge sharing communities [9]. In a community where knowledge is frequently exchanged, new members do not preferably *attach* to existing active members, but instead, active members play an important role in helping new members stay in the community. In the context of our postreply network in health care communities, active members frequently help others by providing social support to them (outgoing links), but they receive relatively less support from new members (incoming links) because newcomers are usually not ready to provide help yet. Overall, H5 was not supported.

We found a positive and significant parameter estimate for the *new_user* configuration in the entire forum, indicating that new users are likely to receive replies. One of the important goals for web-based health communities is to increase community prosperity, and hence, web forums such as Tianmijiayuan [44] encourage users to help new members. Therefore, message postings from new members could be more easily noticed in the community, making the new members more likely to receive social support from other users in the web-based health community. Interestingly, this effect was negative and significant in all 3 subforums. Note that the reply networks in the subforums only counted user interactions within each subforum. Hence, our observation implies that users who recently registered tend to participate in discussions in multiple subforums rather than staying in one specific subforum. With the rapid development of internet technology, users have changed tremendously in recent years. Our results indicate that the newly joined web-based health community participants tend to utilize resources from multiple sources. Therefore, H6 was supported only in the entire forum.

Both *optimistic* and *pessimistic* configurations were positive and significant in the entire forum, indicating that patients with polarized sentiment were more likely to receive replies in the entire forum. This confirms prior findings that polarized emotion can entail more attention [41-43]. In the 3 subforums, users with polarized sentiment were less likely to receive replies in most cases, possibly because of the same reason discussed for new users. The only exception was observed in the *Diabetes Knowledge* subforum, where *the pessimistic* configuration remained positive and significant. This implies that most of the threads seeking informational support in the knowledge sharing subcommunity could be associated with negative mood. It is intuitive because patients are likely to be anxious and desperate during the information-seeking process. Moreover, giving informational support may be prioritized for patients in desperate needs due to negativity bias [54]. Overall, H7 was supported, and negative sentiment was found to have a unique impact when seeking informational support in web-based health communities.

For communication between users of similar sentiment, neither *opti-opti* nor *pessi-pessi* was significant, indicating that users with similar sentiment were exchanging messages just as normal. This observation differs slightly from findings in previous research where homophily effects manifested in more objective attributes such as gender and health status [6]. For personal attributes such as sentiment, we found that the influence effect was stronger than the homophily effect [55]. In the Diabetes Knowledge subforum, both *pessi-opti* and *opti-opti* configurations were positive and significant, indicating that optimistic users were more likely to provide support to other users who are polarized in sentiment when exchanging health care knowledge (ie, diabetes knowledge in our data set). The effect was stronger in the *pessi-opti* configuration, indicating that positive attitude can influence other users, especially those who are in a negative mood. By interacting with optimistic users, pessimistic users can obtain relief, receive encouragement, and improve emotional well-being overall. To summarize, H8a and H8b were partially supported: sentiment plays a key role in communication when information exchange is involved, and social support is more likely to come from optimistic users.

Robustness Tests

Robustness tests conducted to examine whether operationalization of active users and new users could have affected our results. Our base experiment used a 25.00 (%) threshold to identify new users and active users. In robustness tests, we used 20.00 (%), 23.00(%), 27.00 (%), and 30.00 (%) instead to operationalize these two nodal attributes and performed ERGM analysis on the entire forum. Tables 3 and 4 show the results of the robustness tests for new users and active users. Overall, the qualitative results did not change, with the exception that the configuration for *active_user* became significant when the top 27.00 (%) or 30.00 (%) users were operationalized as highly active users. This was due to the fact that several users newly included in the robustness tests 3 and 4 posted very popular threads that received a large number of replies. Considering that the effect did not change when the threshold was changed to 20.00 (%) or 23.00 (%), we argue that being highly active did not have significant correlations with receiving support, and our qualitative results remain the same as in the base test.

Table 3. Results of robustness tests, new users evaluated under different thresholds.

Configuration	Exponential random graph model parameters and *P* values									
	Base test, threshold=25.00 (%)		Robustness test 1, threshold=20.00 (%)		Robustness test 2, threshold=23.00 (%)		Robustness test 3, threshold=27.00 (%)		Robustness test 4, threshold=30.00 (%)	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
H1: reciprocity	3.850	<.001	4.070	<.001	4.050	<.001	3.785	<.001	3.799	<.001
H2: type	0.240	<.001	0.247	<.001	0.239	<.001	0.258	<.001	0.267	<.001
H3: friend	3.473	<.001	3.370	<.001	3.516	<.001	2.976	<.001	2.996	<.001
H5: active_user	0.012	.76	−0.005	.88	.03	.52	−0.019	.61	−0.013	.74
H6: new_user	0.409	<.001	0.487	<.001	0.432	<.001	0.362	<.001	0.316	<.001
H7: optimistic	0.289	<.001	0.305 (<.001)	<.001	0.331	<.001	0.320	<.001	0.340	<.001
H7: pessimistic	0.144	.05	0.050	.48	0.134	.08	0.118	.08	0.145	.05
H8a: opti-opti	−0.332	.23	−0.138	.64	−0.042	.89	−0.210	.43	0.117	.67
H8a: pessi-pessi	−0.140	.71	−0.244	.42	−0.432	.24	−0.319	.28	−0.757	.12
H8b: opti-pessi	0.168	.51	0.224	.38	0.200	.48	0.223	.44	0.071	.79
H8b: pessi-opti	−0.218	.48	−0.652	.05	−0.522	.11	−0.585	.12	−0.448	.13

Table 4. Results of robustness tests, active users evaluated under different thresholds.

Configuration	Exponential random graph model parameters and *P* values									
	Base test, threshold=25.00 (%)		Robustness test 1, threshold=20.00 (%)		Robustness test 2, threshold=23.00 (%)		Robustness test 3, threshold=27.00 (%)		Robustness test 4, threshold=30.00 (%)	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
H1: reciprocity	3.850	<.001	3.748	<.001	4.052	<.001	3.881	<.001	3.943	<.001
H2: type	0.240	<.001	0.243	<.001	0.259	<.001	0.260	<.001	0.271	<.001
H3: friend	3.473	<.001	3.071	<.001	3.635	<.001	3.293	<.001	3.130	<.001
H5: active_user	0.012	.76	−.067	.14	0.040	.31	0.181	<.001	0.232	<.001
H6: new_user	0.409	<.001	0.431	<.001	0.423	<.001	0.440	<.001	0.497	<.001
H7: optimistic	0.289	<.001	0.315	<.001	0.310	<.001	0.353	<.001	0.343	<.001
H7: pessimistic	0.144	.05	0.156	.04	0.177	.02	0.125	.04	0.168	.04
H8a: opti-opti	−0.332	.23	−0.069	.82	0.047	.82	−0.229	.36	−0.286	.24
H8a: pessi-pessi	−0.140	.71	−0.212	.54	−0.366	.31	−0.250	.37	−0.394	.20
H8b: opti-pessi	0.168	.51	0.204	.39	0.244	.37	0.166	.49	0.357	.30
H8b: pessi-opti	−0.218	.48	−0.354	.28	−0.591	.12	−0.529	.09	−0.400	.30

Discussion

Summary of Results

This study uses ERGM to explore patterns of information exchange and social support in web-based health communities. Table 5 summarizes the hypotheses testing results. For hypotheses that were not supported or only partially supported, additional implications were provided. Overall, we found that reciprocity could promote information exchanges effectively. When sharing health knowledge, the homophily effect was not strong in web-based health communities, and conversations were more likely to occur between users of different types (eg, patient and doctor, web service staff, and regular users). Web-based friends were very likely to exchange information frequently with each other. Newly registered users were overall associated with better chances of receiving replies from peers. Sentiment plays an important role in web-based health communities, and users with polarized sentiment tend to receive more replies. In particular, pessimistic users were associated with better chances of informational support when knowledge is exchanged. Most of such support came from optimistic users.

Table 5. Summary of research hypotheses and results.

Hypothesis	Result	Implications
Hypothesis 1: Reciprocated information exchange is likely to develop in web-based health communities.	Supported	In web-based communities, norm of reciprocity exists.
Hypothesis 2: Reciprocated information exchange is likely to develop between users who share similar concerns in web-based health communities.	Partially supported	In web-based communities, homophily effects are not strong when health information is exchanged.
Hypothesis 3: Reciprocated information exchange is likely to develop between users who are web-based friends.	Supported	In web-based communities, friends are likely to exchange messages often.
Hypothesis 4: Patients who receive social support tend to provide support to others who are not necessarily the support provider.	Not supported	Indirect reciprocity hardly exists in Web-based Health Community.
Hypothesis 5: Highly active users are more likely to receive replies as social support.	Not supported	Preferential attachment was found to be in the opposite direction in knowledge sharing communities.
Hypothesis 6: In web-based health communities, new users are more likely to receive replies as social support.	Partially supported	Users who recently registered tend to participate in discussions in multiple subforums rather than staying in one specific subforum.
Hypothesis 7: In web-based health communities, patients with polarized sentiment are more likely to receive replies on their posts.	Partially supported	Negative sentiment was found to have a unique promoting impact when seeking informational support in web-based health communities.
Hypothesis 8a: Patients are likely to receive replies from peers with similar sentiment valence. Hypothesis 8b: Patients are likely to receive replies from peers with opposite sentiment valence.	Partially supported	Communication between sentiment polarized patients has a complex pattern: only when information exchange is involved, optimistic users are more likely to give support to other sentiment polarized users.

Contributions

Our research makes several contributions to the literature. First, this study made the first attempt to test the network exchange framework on reply networks developed in web-based health communities. Web-based health discussions are distinct from other types of conversations in that they contain sensitive and private information and specialized knowledge. Handling such information gives rise to special community norms that could lead to unique social exchange patterns [9,24]. Our study applied ERGM under a network exchange framework and identified a number of such unique patterns. This study complements the literature on network exchange theories and contributes to a better understanding of social exchange patterns in the web-based health community context. Specifically, compared with conventional social networking sites where the formation of social ties is driven by homophily effects, we found that conversations between users of different types were more common when users discussed diabetes knowledge. It does not conflict with prior findings in the network exchange framework because information exchange is different from simply making friends. User heterogeneity could actually increase the effectiveness of knowledge sharing [7]. In terms of preferential attachment, we found that the sentiment of users interacts with discussion topics during the formation of reply networks. Generally, showing polarized sentiment resulted in better chances of receiving replies. However, when seeking knowledge regarding disease, expressing negative emotion could be a better strategy. We further found that most users who provided social support to such users were optimistic users.

Second, our research used sentiment analysis to identify optimistic users and pessimistic users from web-based health communities. To the best of our knowledge, our study is the first to examine how users with different sentiments participate differently in information exchange and social support activities.

Practically, findings from this study help patients in web-based health communities to obtain information and social support more effectively. For example, in addition to making friends, patients are encouraged to participate in discussions on health care knowledge as well as personal life to increase their visibility in the community. It is fine to express negative sentiment when seeking informational support, and showing a positive attitude could be more helpful when making friends with others.

Limitations

A limitation of this study is that our empirical analysis focused on a diabetes-related health community. Although we expect that similar patterns of information exchange and social support should be observed in other web-based health communities that provide web forums, it is interesting to see if the addition of other social features (eg, health platforms that provide feeds to users based on collaborative filtering) will affect how patients interact with each other. Moreover, the internet has changed dramatically over the 10-year time frame covered in this study. As the number of users increases over time, the resulting users in this study after network dichotomization might represent a more recent sample. Performing a temporal analysis to examine how social support patterns evolve dynamically could also be a future direction.

Acknowledgments

This research was supported by the National Natural Science Foundation of China with grants (71971082, 71471064, 71371005, and 91646205), the Key Soft Science Projects in Shanghai (grant number 19692106700), the Humanity and Social Science Youth Foundation of the Ministry of Education of China (grant number 18YJC630068), and Fundamental Research Funds for the Central Universities.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research hypotheses, levels of analysis, and graphical illustrations.
[PNG File , 143 KB - jmir_v22i9e18062_app1.png]

References

1. Wang X, Parameswaran S, Bagul DM, Kishore R. Can online social support be detrimental in stigmatized chronic diseases? A quadratic model of the effects of informational and emotional support on self-care behavior of HIV patients. J Am Med Inform Assoc 2018 Aug 1;25(8):931-944. [doi: 10.1093/jamia/ocy012] [Medline: 29684190]
2. Comerford M, Fogel R, Bailey JR, Chilukuri P, Chalasani N, Lammert CS. Leveraging social networking sites for an autoimmune hepatitis genetic repository: pilot study to evaluate feasibility. J Med Internet Res 2018 Jan 18;20(1):e14 [FREE Full text] [doi: 10.2196/jmir.7683] [Medline: 29348111]
3. Zhang S, O'Carroll BE, Owen J, Bakken S, Elhadad N. Online cancer communities as informatics intervention for social support: conceptualization, characterization, and impact. J Am Med Inform Assoc 2017 Mar 1;24(2):451-459 [FREE Full text] [doi: 10.1093/jamia/ocw093] [Medline: 27402140]
4. Sudau F, Friede T, Grabowski J, Koschack J, Makedonski P, Himmel W. Sources of information and behavioral patterns in online health forums: observational study. J Med Internet Res 2014 Jan 14;16(1):e10 [FREE Full text] [doi: 10.2196/jmir.2875] [Medline: 24425598]
5. Liu X, Sun M, Li J. Research on gender differences in online health communities. Int J Med Inform 2018 Mar;111:172-181. [doi: 10.1016/j.ijmedinf.2017.12.019] [Medline: 29425630]
6. Song X, Jiang S, Yan X. Collaborative Friendship Networks in Online Healthcare Communities: An Exponential Random Graph Model Analysis. Smart Healthpringer International Publishing 2014:-. [doi: 10.1007/978-3-319-08416-9_8]
7. Goh JM, Gao G, Agarwal R. The creation of social value: can an online health community reduce rural-urban health disparities? MIS Q 2016 Jan 1;40(1):247-263. [doi: 10.25300/MISQ/2016/40.1.11]
8. Tates K, Antheunis ML, Kanters S, Nieboer TE, Gerritse MB. The effect of screen-to-screen versus face-to-face consultation on doctor-patient communication: an experimental study with simulated patients. J Med Internet Res 2017 Dec 20;19(12):e421 [FREE Full text] [doi: 10.2196/jmir.8033] [Medline: 29263017]
9. Faraj S, Johnson SL. Network exchange patterns in online communities. Organ Sci 2011 Dec;22(6):1464-1480. [doi: 10.1287/orsc.1100.0600]
10. Poortvliet PM. Examining the effects of mastery and performance goals on information exchange processes. Netherland J Psychol 2009 Jun;65(2):53-61. [doi: 10.1007/bf03080127]
11. Jin J, Li Y, Zhong X, Zhai L. Why users contribute knowledge to online communities: an empirical study of an online social Q&A community. Inf Manag 2015 Nov;52(7):840-849. [doi: 10.1016/j.im.2015.07.005]
12. Aubé C, Brunelle E, Rousseau V. Flow experience and team performance: the role of team goal commitment and information exchange. Motiv Emot 2013 May 30;38(1):120-130. [doi: 10.1007/s11031-013-9365-2]
13. Johnsen MC. Social Support Measurement and Intervention: A Guide for Health and Social Scientists. PS 2001 Oct;52(10):1404-1404. [doi: 10.1176/appi.ps.52.10.1404]
14. Uchino BN. Social support and health: a review of physiological processes potentially underlying links to disease outcomes. J Behav Med 2006 Aug;29(4):377-387. [doi: 10.1007/s10865-006-9056-5] [Medline: 16758315]
15. Langford CP, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. J Adv Nurs 1997 Jan;25(1):95-100. [doi: 10.1046/j.1365-2648.1997.1997025095.x] [Medline: 9004016]
16. Uehara E. Dual exchange theory, social networks, and informal social support. Am J Sociology 1990 Nov;96(3):521-557. [doi: 10.1086/229571]
17. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
18. Rheingold H. The Virtual Community: Homesteading on the Electric Frontier. New York, USA: HarperPerennial; 1993.

19. Cacciamani G, Aron M, Gill I, Desai M. Oncological outcome according to attainment of pentafecta after robot assisted radical cystectomy among bladder cancer patients using KORARC database (730 multicenter robot radical cystectomy database). BJU Int 2020 Jul 18:-. [doi: 10.1111/bju.15222] [Medline: 32870582]
20. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
21. Wellman B. For a Social Network Analysis of Computer Networks:a Sociological Perspective on Collaborative Work and Virtual Community. In: Proceedings of the 1996 ACM SIGCPR/SIGMIS conference on Computer personnel research. 1996 Presented at: SIGCPR'96; April 11-13, 1996; New York, USA. [doi: 10.1145/238857.238860]
22. Walker HA, Thye SR, Simpson B, Lovaglia MJ, Willer D, Markovsky B. Network exchange theory: recent developments and new directions. Soc Psychol Q 2000 Dec;63(4):324. [doi: 10.2307/2695843]
23. Lee JY, Yang C, Hsu C, Wang J. A longitudinal study of leader influence in sustaining an online community. Ing Manag 2019 Mar;56(2):306-316. [doi: 10.1016/j.im.2018.10.008]
24. Yuqing R, Kraut R, Kiesler S. Applying common identity and bond theory to design of online communities. Organ Stud 2016 Jun 30;28(3):377-408. [doi: 10.1177/0170840607076007]
25. Gouldner AW. The norm of reciprocity: a preliminary statement. Am Sociol Rev 1960 Apr;25(2):161. [doi: 10.2307/2092623]
26. Dahlke DV, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015 Mar 27;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
27. Wasko ML, Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Q 2005;29(1):35. [doi: 10.2307/25148667]
28. Wu B, Jiang S, Chen H. Effects of individuals' motivations on communications in online health forums. Soc Behav Personal 2016 Mar 23;44(2):299-312. [doi: 10.2224/sbp.2016.44.2.299]
29. Chang MK, Cheung W, Tang M. Building trust online: interactions among trust building mechanisms. Inf Manag 2013 Nov;50(7):439-445. [doi: 10.1016/j.im.2013.06.003]
30. Raghavun K, Vassileva J. Visualizing Reciprocity to Motivate Participation in an Online Community. In: 5th IEEE International Conference on Digital Ecosystems and Technologies. 2011 Presented at: IEEE DEST'11; May 31-June 3, 2011; Daejeon, South Korea. [doi: 10.1109/DEST.2011.5936604]
31. Granitz NA, Koernig SK, Harich KR. Now it's personal. J Mark Educ 2008 Oct 13;31(1):52-65. [doi: 10.1177/0273475308326408]
32. Aiello LM, Barrat A, Schifanella R, Cattuto C, Markines B, Menczer F. Friendship prediction and homophily in social media. ACM Trans Web 2012 May 1;6(2):1-33. [doi: 10.1145/2180861.2180866]
33. Thelwall M. Homophily in MySpace. J Am Soc Inf Sci 2009 Feb;60(2):219-231. [doi: 10.1002/asi.20978]
34. Hogg MA, Terry DI. Social identity and self-categorization processes in organizational contexts. Acad Manage Rev 2000 Jan;25(1):121-140. [doi: 10.5465/amr.2000.2791606]
35. Nahapiet J, Ghoshal S. Social capital, intellectual capital, and the organizational advantage. AMR 1998 Apr;23(2):242-266. [doi: 10.5465/amr.1998.533225]
36. Takahashi N. The emergence of generalized exchange. Am J Sociol 2000 Jan;105(4):1105-1134. [doi: 10.1086/210400]
37. Katz JE, Rice RE. Social consequences of Internet use: Access, involvement and interaction. Cambridge, MA: MIT Press 2002:-. [doi: 10.7551/mitpress/6292.001.0001]
38. Joyce E, Kraut RE. Predicting continued participation in newsgroups. J Comp Mediated Comm 2006 Apr;11(3):723-747. [doi: 10.1111/j.1083-6101.2006.00033.x]
39. Watts DJ. The 'new' science of networks. Annu Rev Sociol 2004 Aug;30(1):243-270. [doi: 10.1146/annurev.soc.30.020404.104342]
40. Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. J Am Med Inform Assoc 2014 Oct;21(e2):e212-e218 [FREE Full text] [doi: 10.1136/amiajnl-2013-002282] [Medline: 24449805]
41. Fredrickson BL. The broaden-and-build theory of positive emotions. Philos Trans R Soc Lond B Biol Sci 2004 Sep 29;359(1449):1367-1378 [FREE Full text] [doi: 10.1098/rstb.2004.1512] [Medline: 15347528]
42. Yoo W, Namkoong K, Choi M, Shah DV, Tsang S, Hong Y, et al. Giving and receiving emotional support online: communication competence as a moderator of psychosocial benefits for women with breast cancer. Comput Human Behav 2014 Jan;30:13-22 [FREE Full text] [doi: 10.1016/j.chb.2013.07.024] [Medline: 24058261]
43. Yang F, Zhong B, Kumar A, Chow S, Ouyang A. Exchanging social support online: a longitudinal social network analysis of irritable bowel syndrome patients' interactions on a health forum. Journalism Mass Commun Q 2017 Sep 28;95(4):1033-1057. [doi: 10.1177/1077699017729815]
44. Tianmijiayuan. URL: http://bbs.tnbz.com/ [accessed 2020-09-11]
45. Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013 Apr;56(4):82-89. [doi: 10.1145/2436256.2436274]

46. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog. In: International Conference on Brain and Health Informatics. 2013 Presented at: BHI'13; October 29-31, 2013; Maebashi, Japan. [doi: 10.1007/978-3-319-02753-1_36]
47. Jiang S, Gao Q, Chen H, Roco MC. The roles of sharing, transfer, and public funding in nanotechnology knowledge-diffusion networks. J Assn Inf Sci Tec 2014 Jun 10;66(5):1017-1029. [doi: 10.1002/asi.23223]
48. Heaney MT. Multiplex networks and interest group influence reputation: an exponential random graph model. Soc Netw 2014 Jan;36:66-81. [doi: 10.1016/j.socnet.2012.11.003]
49. Yan L, Peng J, Tan Y. Network dynamics: how can we find patients like us? Ing Syst Res 2015 Sep;26(3):496-512. [doi: 10.1287/isre.2015.0585]
50. Robins G, Pattison P, Kalish Y, Lusher D. An introduction to exponential random graph (p*) models for social networks. Soc Netw 2007 May;29(2):173-191. [doi: 10.1016/j.socnet.2006.08.002]
51. Kolaczyk ED, Csárdi G. Statistical Analysis of Network Data With R. New York, USA: Springer; 2014.
52. Snijders TA, Pattison PE, Robins GL, Handcock MS. New specifications for exponential random graph models. Sociol Method 2016 Jun 23;36(1):99-153. [doi: 10.1111/j.1467-9531.2006.00176.x]
53. Barak A, Gluck-Ofri O. Degree and reciprocity of self-disclosure in online forums. Cyberpsychol Behav 2007 Jun;10(3):407-417. [doi: 10.1089/cpb.2006.9938] [Medline: 17594265]
54. Rozin P, Royzman EB. Negativity bias, negativity dominance, and contagion. Pers Soc Psychol Rev 2016 Dec 21;5(4):296-320. [doi: 10.1207/S15327957PSPR0504_2]
55. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci U S A 2009 Dec 22;106(51):21544-21549 [FREE Full text] [doi: 10.1073/pnas.0908800106] [Medline: 20007780]

Abbreviations

ERGM: exponential random graph model

Edited by G Eysenbach; submitted 31.01.20; peer-reviewed by X Song, D Walker, S Purkayastha; comments to author 12.06.20; revised version received 29.07.20; accepted 21.08.20; published 29.09.20.

Please cite as:
Liu X, Jiang S, Sun M, Chi X
Examining Patterns of Information Exchange and Social Support in a Web-Based Health Community: Exponential Random Graph Models
J Med Internet Res 2020;22(9):e18062
URL: http://www.jmir.org/2020/9/e18062/
doi:10.2196/18062
PMID:32990628

Original Paper

Accelerating Innovation in Health Care: Insights From a Qualitative Inquiry Into United Kingdom and United States Innovation Centers

Kathrin Cresswell[1], BSc, MSc, PhD; Robin Williams[2]; Narath Carlile[3], MD, MPH; Aziz Sheikh[1], FRSE, FMedSci

[1]Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[2]Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom

[3]Harvard Medical School, Boston, MA, United States

Corresponding Author:
Kathrin Cresswell, BSc, MSc, PhD
Usher Institute
The University of Edinburgh
Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 131 651 4151
Email: kathrin.beyer@ed.ac.uk

Abstract

Background: Digital health innovations are being prioritized on international policy agendas in the hope that they will help to address the existing health system challenges.

Objective: The aim of this study was to explore the setup, design, facilities, and strategic priorities of leading United Kingdom and United States health care innovation centers to identify transferable lessons for accelerating their creation and maximizing their impact.

Methods: We conducted qualitative case studies consisting of semistructured, audio-recorded interviews with decision makers and center staff in 6 innovation centers. We also conducted nonparticipant observations of meetings and center tours, where we took field notes. Qualitative data were analyzed initially within and then across cases facilitated by QSR International's NVivo software.

Results: The centers had different institutional arrangements, including university-associated institutes or innovation laboratories, business accelerators or incubators, and academic health science partnership models. We conducted interviews with 34 individuals, 1 group interview with 3 participants, and observations of 4 meetings. Although the centers differed significantly in relation to their mission, structure, and governance, we observed key common characteristics. These included high-level leadership support and incentives to engage in innovation activities, a clear mission to address identified gaps within their respective organizational and health system settings, physical spaces that facilitated networking through open-door policies, flat managerial structures characterized by new organizational roles for which boundary spanning was key, and a wider innovation ecosystem that was strategically and proactively engaged with the center facilitating external partnerships.

Conclusions: Although innovation in health care settings is unpredictable, we offer insights that may help those establishing innovation centers. The key in this respect is the ability to support different kinds of innovations at different stages through adequate support structures, including the development of new career pathways.

(J Med Internet Res 2020;22(9):e19644) doi:10.2196/19644

KEYWORDS

innovation; health information technology; health care

Introduction

Health systems internationally are facing unprecedented pressures to address the challenges associated with demographic shifts while improving quality and safety and decreasing cost [1]. Digital health innovations are increasingly seen by policy makers and funders as instrumental in addressing these challenges [2-6]. Significant strategic investments are being made in this area in the United Kingdom and elsewhere, including the establishment of national innovation agencies and

governmental city and regional development initiatives [7,8]. These are characterized by a range of different interpretations of the concept of innovation itself, but the majority focus on product innovation—the creation of new technological artifacts and the processes of bringing these to the market.

The creation of new technological artifacts through digital health innovation has, however, a checkered history with examples of substantial successes and dismal failures [9]. Alongside a large reservoir of potential innovations with many challenges to be addressed [10], there is a graveyard of innovations that failed completely at the outset or did not successfully scale-up [11]. This is partly due to the inherent difficulties in planning innovations in which the emergence of truly novel practice is hard to predict and technologies must satisfy a range of diverse requirements and needs [12,13]. Progress in health care innovation has been further hampered by uncertain pathways to the market, the lack of established methods for achieving success, and cumbersome processes of preparing innovations to meet exigencies of clinical governance and health service procurement [14-17].

Therefore, there is a need to better understand the innovation landscape, with a view to obtaining insights into factors that catalyze ideas and translate them into innovations that have the potential to improve outcomes for patients, providers, and health systems. There are important lessons to be learned from addressing the range of approaches adopted in different settings internationally, particularly those that have created local and regional innovation environments. These include the creation of the so-called *innovation hubs or centers*, which are collaborative, enabling spaces that bring together heterogeneous expertise from different sectors. We sought to investigate the setup, design, facilities, and strategic priorities of leading international health care innovation centers to identify transferable lessons for those seeking to accelerate or stimulate innovation within health care and identify current and common opportunities and challenges.

Methods

Permissions

We obtained institutional review board approval for this study from the Centre for Population Health Sciences at the University of Edinburgh, United Kingdom, on February 22, 2018. Each participant provided informed consent.

Design

We conducted a series of qualitative case studies exploring a range of United Kingdom and United States health care innovation center facilities that are considered by funders and innovators as examples of success.

Sampling

For our purposes, the definition of an innovation center was relatively broad as we wanted to capture the breadth of success factors across a range of contexts. Therefore, we defined an innovation center as an organizational entity that focused on incubating, developing, or accelerating new digital products for health care delivery and health promotion.

Sampling was informed by a recent mapping of leading health care innovation centers where centers that are viewed as successful are discussed [18]. The ability of centers to achieve impact and returns on investment was a key criterion to be included in our sample. We recruited centers that had been established at least 5 years before data collection (as a proxy indicator of success) to obtain insights into the challenges faced and sustainability.

To ensure maximum variation, we sampled a range of locations in the United Kingdom and the United States with a variety of foci, including academic centers, early discovery, and scale-up facilities [19]. These also included a mix of project- and product-based services and relatively new as well as established centers. Some had an emphasis on digital health products, whereas others did not exclusively focus on digital health.

Within each innovation center, we used purposive sampling to identify a diverse range of stakeholders who were involved in planning, procuring, developing, using, or managing innovation centers or associated facilities [20]. Participants comprised opinion leaders, system developers, innovators, managers, and users from various backgrounds (clinical, engineering, technology, managerial, commercial). Initial contacts were established by emailing senior innovation center leaders, and interview participants were snowball sampled through these contacts.

Data Collection

Data collection in each case study consisted of semistructured, in-depth one-to-one interviews or, if more convenient for interviewees, telephone interviews. Interviews were conducted with a topic guide (Textbox 1), exploring views on the setup, culture, and features of innovation approaches as well as expected and experienced benefits, experiences and lessons learned, perceived challenges, and potentially transferable lessons. Questions were informed by conceptual work led by one of our coauthors (AS) scoping where and how innovations in health care settings are succeeding [18].

All interviews and site visits were conducted by the same researcher (KC). Interviews were, with permission, digitally audio-recorded. However, audio-recording was not feasible in 5 interviews because these took place in noisy environments. In such instances, the researcher took extensive field notes. Recordings were then transcribed verbatim together with accompanying field notes.

We were opportunistic in developing our program of visits and observations, with the researcher joining center meetings and guided tours of physical spaces. At these, the researcher took field notes that were unstructured but involved recording the location, people, the topic of discussions, impressions on the environment, and any other emerging impressions on social dynamics.

Data generation ended when saturation was reached and no significantly new themes emerged from the concurrent data analysis [21].

Textbox 1. A sample topic guide.

Interviewee's background: current position, role in relation to the innovation center
Setup and facilities (technologies, networks, and managerial structures)
Expected and experienced benefits
Facilitators for and barriers to health care innovation
Challenges and lessons learned
Sustainability models
Anything else?
Anyone else we can speak to?

Data Analysis

Qualitative data collection and analysis were iterative, allowing emerging themes to be explored further while seeking disconfirming evidence [22]. Coding was informed by our extensive previous literature reviews on digital health interventions and earlier work on innovation environments [23,24]. Our coding framework was based on this extensive previous work, providing an overall initial coding structure. We did, however, also allow new themes to emerge and refined the framework accordingly. These new dimensions are discussed in detail in the Results section.

Interview notes, transcripts, and observation notes were uploaded onto the NVivo11 software and initially coded against the topic guide categories by case study (within-case analysis). As data analysis progressed, we identified new categories and rearranged codes and subcodes to present a holistic picture of innovation center strategies, stakeholders, and environments. In doing so, we combined a diverse range of interviewees, perspectives, facilities, and contexts. Detailed within-case analysis was followed by analysis across cases to identify overarching themes, similarities and differences between cases, and potential implications for other settings.

Results

We visited 6 international innovation centers in the United States and the United Kingdom, conducted interviews with 34 individuals and 1 group interview with 3 participants, and observed 4 center meetings. The characteristics of the centers are provided in Table 1, and the characteristics of the interviewees are provided in Table 2.

All the centers had different physical and organizational setups and facilities, depending on their primary purpose. However, we also identified some common threads across the settings. Overarching themes and subthemes are illustrated in Figure 1.

Table 1. Innovation center characteristics.

Center number	Primary purpose	High-level overview	Date established; how they started	Ongoing funding model	Outputs	Location
1	Four innovation centers with focus on drug development, innovation incubation, and research acceleration part of 1 university umbrella organization	Located on the university campus, some degree of networking across centers	2000, 2006, 2010, 2011	• Commercial funding • University funding • Fellowship and teaching model	• Technological products • Education • Networking	United States
2	Rapid start-up facilities for commercial companies and digital creative industries	Managed office space specializing in incubating digital health companies	2013; designed to promote growth for high-growth tech start-ups	• Tenancy (letting out space to companies)	• Technology acceleration and scale-up	United Kingdom
3	Corporate accelerator, variety of health-related specialties	University-affiliated, focus on evidence-based innovation	2009; to identify, develop, and scale evidence-based health care innovation	• Funding councils • University (operational budget) • Commercial • Philanthropy	• Technology acceleration and scale-up	United Kingdom
4	Academic health science partnership	Virtual network consisting of scientists, health care staff and organizations, support to establish relationships	2009; to translate research and innovation into benefits for patients and populations	• Funding councils	• Education and networking leading to digital innovation	United Kingdom
5	Coworking space	Managed office space specializing in establishing partnerships through coworking	2010; to create workspaces that help to create communities of practice	• Tenancy (letting out space to companies)	• Relationship building leading to digital innovation	United Kingdom
6	Digital health start-up center	Located on hospital premises, focus on helping stakeholders to find out if their product is (or could be made) commercially viable and provide support for the development process	2013; to drive internal innovation	• Health care organization operational budget	• Technologies that can be used in the operational hospital environment	United States

Table 2. Interviewee characteristics.

Participant number	Gender	Occupation and role in the center	Center number	Location
1	Male	Director	2	United Kingdom
2	Male	Designer	3	United Kingdom
3	Group interview (1 male, 2 females)	Director, 2 innovation managers	4	United Kingdom
4	Male	Director	3	United Kingdom
5	Male	Director	2	United Kingdom
6	Male	Advisor	3	United Kingdom
7	Male	Clinician	3	United Kingdom
8	Male	Innovation champion	3	United Kingdom
9	Female	Clinician	2	United Kingdom
10	Female	Strategy consultant	2	United Kingdom
11	Male	Academic	3	United Kingdom
12	Male	Innovation manager	6	United States
13	Male	Innovation analyst	6	United States
14	Female	Innovation manager	6	United States
15	Male	Clinician	6	United States
16	Male	Clinician	6	United States
17	Male	Engineer/designer	6	United States
18	Male	Director	6	United States
19	Female	Academic	6	United States
20	Male	Policy	6	United States
21	Male	Academic/clinician	6	United States
22	Male	Industry	6	United States
23	Male	Academic/clinician	6	United States
24	Male	Academic/clinician	6	United States
25	Male	Academic/clinician	6	United States
26	Male	Academic/clinician	1	United States
27	Male	Academic/clinician/director	1	United States
28	Male	Academic/clinician	1	United States
29	Male	Academic	1	United States
30	Female	Academic/director	1	United States
31	Female	Academic/codirector	1	United States
32	Male	Director	1	United States
33	Male	Academic/clinician/director	1	United States
34	Male	Academic/clinician	1	United States

Figure 1. Overview of findings.

Mission and Business Strategy

Different centers had various underlying values, infrastructures, and needs. However, all worked hard to create organizational cultures that placed innovation at the core of their activities. This involved actively engaging with external communities to promote shared knowledge (what has worked commercially and what the important problems within health care are) and the creation of easier pathways for opportunities to solve those problems:

> *Just generally just to have a buzz going on about innovation is happening and what's going on, and just bringing promotion internally as well as externally, so that we're open, the external community knows, like we are open for business.* [Participant 12, male, innovation manager, United States]

Common to all was also an effort to align with the various key local and national societal and health system challenges and existing technological and social infrastructures, and bring together various stakeholders involved in these. Centers had portfolios that combined more and less adventurous or disruptive innovation forms, allowing to focus on solving the most pressing real-world problems while still satisfying market needs and coordinating with other existing initiatives. The key here was perceived to be the alignment of commercial, clinical, and patient needs and values, as these could act as incentives for the various stakeholders who need to be involved:

> *…the product has to address a significant medical need, number one...the majority of times, that's going to require that it's going to be commercially attractive, right?* [Participant 29, male, academic, United States]

Defining a unique proposition to adopters that was not addressed elsewhere was seen as essential to create value and impact:

> *You've got to be very clear what it is that you're trying to do in terms of establishing a unique selling point and a unique position within the market.* [Participant 6, male, advisor, United Kingdom]

Activities frequently involved mapping key local, national, and international stakeholders of potential relevance to the center and aligning their motivations, values, and needs with activities. For example, there was often a focus on bringing together *communities of interest*, such as commercial sectors, academic settings, and health care professionals, thereby bridging the gap between the problem, idea, product, and use of the product in context.

To ensure value to patients, it was argued that the first stage of the innovation process should be to identify existing needs and thereafter identify the technology that might address those needs. This needs-based approach was also seen to help bring different stakeholders together as a fundamental starting point for further activities that facilitated aligning different viewpoints and incentive structures around clinical, patient, and economic needs:

> *…if you're trying to start with a technology and then hoping that what you come up with is going to be received positively, by all these stakeholders, your chances are pretty low. But if you can start with...understanding the stakeholder landscape, in the beginning then you have a better chance.* [Participant 32, male, director, United States]

Although stakeholders often had different expectations and needs in relation to timelines, which required a great deal of relationship building, the centers frequently took on this intermediary role and acted as connectors of otherwise disconnected worlds (although these again varied across the centers they included, for instance, academic health care and commercial domains):

> *I think our role has been making sure they get connected with the people who can help them, and so expanding across the network was not just them, they could not have done that by themselves.* [Participant 15, male, clinician, United States]

To be commercially viable, innovation center leadership had to balance a number of tensions. These included aligning organizational priorities with stakeholder motivations (which may both be subject to change over time), senior leadership

commitment to innovation while allowing a degree of local creativity, and some risk taking (eg, funding for *risky projects* while ensuring a steady stream of income):

> *...we have that appetite to take on a little bit of risk to work with these newer companies as long as there's alignment and we know that it's a good team, it's a good product, it's something that we will derive value from.* [Participant 18, male, director, United States]

Most centers had a mixture of short-term *risky* funding sources they had won or could give out (eg, seed funding, which was particularly relevant for early development and proof of concept of innovations) and also relied on more stable sources of income for security (eg, operational or research council funds). More stable funding was often associated with sustained investment to bring innovations to a point where they could survive in the market. Textbox 2 summarizes the various funding sources discussed by the participants.

Textbox 2. Summary of funding sources.

Seed funding for start-up
Organizational operations budget (to ensure alignment with operational objectives)
Grant funding (government, research councils)
Commercial funding (eg, venture capital firms)
Private foundations (eg, angel investors, philanthropic donations)
Tenancy (letting out space to companies)
Fellowship and teaching funding

Facilities and Managerial Structures

Physical space and buildings varied from one-room office spaces to whole buildings, where center staff were colocated with commercial companies, and centers that were located within a health care organization. A crucial feature of many centers was an *open-door policy*, meaning those with ideas who wanted to innovate could come in as a first point of contact:

> *Anyone...can come, schedule some time with one of our members and talk about and hash through the idea of where they are, what the next step is.* [Participant 16, male, clinician, United States]

If clinicians were identified as important stakeholders (which depended on the primary purpose of the center), colocation with clinical premises was important so that clinicians could make use of the facilities, attend events, and network.

Another important characteristic of the physical space was flexibility such that facilities could be used for multiple purposes (including individual working, hosting meetings, external events, and conferences). Spaces also promoted colocation and were adaptable to ensure responsiveness to changing stakeholders' needs over time. If centers consisted of whole buildings, these encouraged social contacts on an informal basis and provided ample opportunities for people to meet either in a planned manner or opportunistically, including cafes and pleasant outside spaces. In short, most spaces consisted of an enriching and engaging environment that staff wanted to spend time in and that they were proud to show off to external stakeholders:

> *Why this building is a very important investment for the hospital and, you know, why people are super-excited about it because basically for the first time you would have clinical and research right here where people could easily bump into each other and talk to each other.* [Participant 19, female, academic, United States]

The director of a center told us that "*the single two most important things are good Internet and good coffee*" (note, center visit).

Leadership often comprised a small, tight-knit team from diverse backgrounds, with commercial, managerial, and clinical knowledge, networks, and skills that aligned with the purpose and mission of the center, combined with an in-depth understanding of the organizational environment (although not all centers were embedded within health care institutions):

> *You have to get people who understand where the bodies are buried in the system you're trying to disrupt.* [Participant 1, male, director, United Kingdom]

Management structures tended to be relatively flat and informal, with strong high-level leadership support for innovation while still allowing team members a degree of creative flexibility. This meant that innovation centers were in some cases not bound by institutional regulations or tied to hierarchical structures that might limit new ways of working and building relationships. For instance, academic and clinical members were often relieved of some pressures associated with their other roles by means of secondments or protected time to innovate:

> *...I think it depends on willingness and whether the environment is conductive to do this. I mean, what would the university say if a professor says that, I want to spend x per cent of my time on this, for various values of x. Would they be sympathetic? Would they be questioning? Would they be outright negative? Do they allow someone to take big chunks of their time to leave for six months, one year, two years...and build something and then leave it to some other team and come back to the university?* [Participant 3, male, academic/clinician, United States]

Team attitude and culture seemed to play a particularly important role with a common drive to *get things done*, a

risk-taking attitude, and a focus on collectively solving problems. Staff members in many cases had an intrinsic commitment to innovation with a strong belief that this was the right way forward to solve health care challenges. Center teams were often proactively recruited by senior leadership. Here, skills and interpersonal capabilities were important:

> *I would say probably more a willingness to be able to talk to others and that personality to be a connector. That's more important...you saw it at the Apple Store and they didn't care that you had any Apple products or if you knew anything about technology. Those are things that they can teach you. But they care more about your people skills and then your personalities because that's very hard to teach...* [Participant 13, male, innovation analyst, United States]

Although certain characteristics were common (eg, managerial backgrounds), we observed a different combination of professional skills and backgrounds, depending on the primary objectives of the center. Many core staff members had experience working in different settings (eg, commercial, academic, clinical, managerial) and the ability to span boundaries, move between different worlds, and connect them (eg, entrepreneurial researchers or highly research-minded entrepreneurs).

The core staff tended to work strategically with third parties where specific skills or specialties needed to be brought in to support different aspects of the innovation journey. Creatively drawing on these was viewed as crucial.

The emergence of innovation centers as a relatively new development also demanded the creation of hybrid roles such as innovation strategy managers and innovation analysts who had no established career pathways and may therefore struggle to prove their value to the wider organizational setting:

> *...innovation managers, the people there to understand and think about describing their own value to their organisations and career pathing for the people here, for the innovation managers here and across all of the innovation team…hospitals don't really know how to value these people. The people in those roles often don't have a language to describe what they do and the value they bring to their institution, to researchers and so on.* [Participant 20, male, policy, United States]

Wider Infrastructural Considerations

Strategic networking with external stakeholders was a key activity across all centers we visited, frequently characterized by proactive efforts to build relationships with collaborators. These varied across centers, depending on the core mission, but often included academic institutions, policy makers, health care providers, patient organizations, current and future funders, and commercial organizations. In doing so, centers brought together stakeholders who would not typically meet in intensive time-limited interactions and events (such as accelerators, hackathons, challenges, conferences, sandbox events, and competitions):

> *So I think the events serve as a major platform for bringing people together and there are some events that we've done specifically targeting those kinds of, you know, interactions.* [Participant 19, female, academic, United States]

Publicizing interactions and events was key to these interactions with dedicated public relations support to promote positive messages and celebrate individual and group wins.

Most acknowledged that there was a critical need to align activities with the wider innovation ecosystem. Centers that were part of this study had placed themselves at the center of academic, commercial, and governmental networks. They were strategically placed in attractive cities that were easily reached through national and international travel networks. A vibrant commercial environment featured heavily and was purposefully aligned to leading universities and health care organizations, where relevant. In some instances, this created a fluid talent pool of people moving between sectors and bridging multiple communities:

> *There's probably less of that here because a lot of people will switch careers back and forth all the time, so the talent pool is pretty fluid. So you'll have a lot of people that will cross-pollinate between the different groups. So I think some of the culture tends to merge a little bit.* [Participant 17, male, engineer/designer, United States]

As such, the ecosystem became somewhat of a magnet that drew entrepreneurial spirits in and attracted a certain type of person, which, in turn, was seen to transform the ecosystem:

> *...everybody from everywhere wants to come here to make their money and be where the excitement is.* [Participant 33, male, academic/clinician/director, United States]

However, the mismatching timelines of different institutional stakeholders were frequently cited as barriers to innovation, with commercial partners needing to move quickly and academic and health care settings being averse to risk and therefore less equipped for moving fast owing to often deeply engrained bureaucratic procedures and hierarchical structures.

In addition, despite a general recognition of the importance of evaluation, many centers struggled with establishing measurable metrics that indicated value. Financial metrics are important for all types of organizations. Although start-ups, accelerators, and incubation centers tended to focus on the number of patents and companies created, university-affiliated centers tended to focus on improvements in the quality and safety of care and staff and patient experience.

A further challenge consisted of aggregating project metrics to make claims about the overall success of the center:

> *When we're talking about how we measure the programme as a whole something we've wrestled with is how do we then take all these disparate clinical metrics and make them into something you can come up with. Because we've tried to sell the programme as we're innovating, we're delivering better*

> *healthcare and we're delivering financial value.* [Participant 15, male, clinician, United States]

Many also struggled to allocate appropriate time and resources to investigate *failed* initiatives, although the importance of this activity was generally recognized:

> *...when we fail, we tell everybody...that we failed because there is so much more learning from that...* [Participant 30, female, academic/director, United States]

Another challenge was scaling, as it was perceived as one of the most unpredictable aspects of the innovation journey and required dedicated resources and expertise. Therefore, some stated there was a danger that innovation centers would "*create a thousand different solutions that the system does not want*" [Participant 5, male, director, United Kingdom].

Discussion

Summary of Findings

We collected qualitative data from a range of settings and identified some important common characteristics of the way different innovation centers approached innovation. Successful centers brought together various combinations of expertise and experience (including academic, technology, service delivery, professional, business, and regulatory) to promote innovations. In the context of myriad opportunities, they helped to identify pain points, stimulate ideas, and facilitate the development of promising avenues. They did so by traversing key segments of the innovation journey, from early high-risk, unproven potential to commercial investment appraisal.

Strengths and Limitations

Although efforts have been made to characterize features of particular successful innovation environments, attempts to reproduce these *critical success factors* in other settings have mostly failed to deliver, as outcomes are unpredictable and contingent on particular forms and contexts of innovation [25]. Our study provides a starting point for those wishing to navigate this challenging area, providing insights into stimulating innovation in high-risk health care environments.

However, innovation research and policy have been held back by a lack of agreed definitions of innovation, with many efforts and centers focusing on product innovation. Therefore, our empirical focus was somewhat limited, potentially neglecting other types of innovation (eg, organizational, service, and social innovation) and different innovation pathways [26-30].

This was an exploratory study that focused on United Kingdom and United States settings, as we wanted to establish factors that have been identified to promote innovation across contexts. Limitations include the modest number of cases and variations in the data collected within each center. Ideally, we would have sampled a wider range of centers in different geographical locations, including other countries, and with a similar focus to produce more comparable results. We would also have liked to recruit a more comparable sample of individual respondents (including a greater number of *on-the-ground* staff) and a more balanced representation of men and women (although this may reflect innovation center workforce trends), but we were limited by accessibility to centers and individuals. We also had to navigate complex approval procedures and a certain degree of trade-off between the depth and breadth of data collected.

In addition, this was a retrospective study of stakeholder perceptions; therefore, it will likely be subject to recall bias. Successful innovations are those in which various possible barriers were avoided and challenges negotiated along the way—a long-term process that is highly unpredictable and is therefore difficult to extrapolate inductively from one case to the next. Similarly, there are likely to be different support models depending on the stage of innovation and the age, size, and funding model of the center. For example, some factors, such as the ability to take risks, may be less realistic in smaller and less well-funded centers, and these centers may also have smaller teams and more restrictive environments than large well-funded centers (eg, old buildings that do not allow for colocation). It may be that in these environments other factors take more prominent roles than others. For example, if the building is not suitable for colocation, then staff may need to compensate by moving around and actively network within the wider ecosystem. There may also be the need to identify other factors that help a center stand out, for example, including a broad range of stakeholders in their activities such as patient and public representatives. Different stages and centers will likely require different roles and foci [31]. Longitudinal real-time ethnographic studies could help to address these issues. These should also seek to identify which innovation initiatives are most likely to be *successful* under what circumstances and develop a quantifiable set of indicators to guide future efforts.

Integration of Findings With the Current Literature

We observed some seemingly paradoxical requirements surrounding the support structures. On the one hand, there was a need for frameworks that encouraged diversity to promote ideas at the outset (eg, where different interdisciplinary groups could meet in short-duration projects). On the other hand, there was a strategic requirement to focus on sustained investment in particular areas to bring innovation to the point it could survive in the market [32]. This strategy seems appropriate for navigating uncertainty, given that many promising innovations are likely to fail along the journey [33].

Our study also shows that there is a clear need for new career pathways for organizations to support different forms of innovation and the need to develop roles and skills for different stages of the innovation journey. These include new hybrid roles of boundary spanners (connectors) who can bridge different worlds. Interpersonal skills and previous experience in context are key in this respect. In addition, successful innovators are often those who have been previously involved in a series of earlier *failed* innovations through which they have gained the resources, reputation, experience, knowledge, and linkages needed for eventual *success* [34].

Health care organizations could promote the exploitation of these skills through the creation of career pathways around new kinds of hybrid roles. For instance, staff need to be presented with opportunities to enhance their capabilities, knowledge, and links with multiple arenas through their involvement in a series

of projects. Innovation calls for access to diverse skills, which may include engineering, venture capitalism, entrepreneurship, financial services, investment communities, technology companies, academia, and clinical and professional services. Key intermediaries often internalize understanding or links to these kinds of expertise.

Both countries in which we carried out our research had firmly established national digital health system strategies, and these national drives resulted in national funding for related innovation, research, and innovation from which the centers benefited. However, there was a marked difference in vendor landscapes and health system competition (public sector vs competitive insurance market), which are likely to have impacted various innovation and acceleration efforts. For example, there may have been stronger drivers for economic gains and a stronger pool of professional expertise driven by the competitive insurance market.

Implications for Practice and Further Research Emerging From This Study

On the basis of this study, we developed a framework of key considerations that a leadership team may wish to consider when establishing a health care innovation center and revising short-, medium-, and long-term strategies. These are summarized in Table 3. This list is not intended to be exhaustive, and dimensions are likely to vary across localities. It should therefore be seen as a guide to structure thinking and not as a recipe for success.

There are also some important implications for informaticians. The key here will be establishing a needs-based approach to innovation that is driven by engaging a range of stakeholder communities and aligning their values. This may be achieved by exploring how different needs can be addressed through various forms of innovation before developing new technological artifacts.

Table 3. Key considerations when establishing and guiding a health care innovation center.

Priority weighing	Area	Key consideration
Priority 1	Need and opportunity	Alignment with key local and national societal and health system challenges
Priority 2	Leadership	Senior-level commitment to innovation and associated commitment to associated organizational changes
Priority 3	Strategic prioritization and incentives	Alignment of the center with organizational priorities in the short, medium, and long term and associated incentives
Priority 4	People	Valuing and promoting interpersonal skills and boundary-spanning capabilities
Priority 5	Culture	Flat and informal management structures allowing a degree of creative flexibility
Priority 6	Funding and resources	Diversity in funding sources allowing a mixture of stable and *risky* investments
Priority 7	Relationships	Allowing the organic development of communities of practice around specific challenges
Priority 8	Space	Easily accessible, pleasant, flexible, and conducive to formal and informal networking by a variety of parties

Conclusions

Although definite measures of success are difficult to establish, we have begun extracting some key considerations that can be used by planners and implementers to guide the establishment and maintenance of health care innovation centers. These include strategy and leadership that view innovation as an organizational priority, establishing organizational cultures and structures that allow experimentation and creative flexibility, and designing physical environments that facilitate networking and relationship building.

There is a clear need to consider different forms of innovation and how these require different kinds of organizational support structures, including establishing new career pathways for hybrid boundary spanners.

Acknowledgments

The authors gratefully acknowledge all participants' time and input. This work was funded by the Wellcome Trust-University of Edinburgh Institutional Strategic Support Fund 3. The views expressed are those of the authors and not necessarily those of the funders.

Conflicts of Interest

None declared.

References

1. Daar AS, Singer PA, Persad DL, Pramming SK, Matthews DR, Beaglehole R, et al. Grand challenges in chronic non-communicable diseases. Nature 2007 Nov 22;450(7169):494-496. [doi: 10.1038/450494a] [Medline: 18033288]

2. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the 'triple aim' of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856 [FREE Full text] [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
3. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
4. Building our Industrial Strategy. Government of UK. URL: https://www.gov.uk/government/consultations/building-our-industrial-strategy [accessed 2020-09-02]
5. Using Information Technology to Improve the NHS. Government of UK. URL: https://www.gov.uk/government/publications/using-information-technology-to-improve-the-nhs [accessed 2020-09-02]
6. Research Forum on DataScience and AI. World Innovation Summit for Health (WISH). URL: http://www.wish.org.qa/data-science-and-ai/ [accessed 2020-09-02]
7. Innovate UK. Government of UK. URL: https://www.gov.uk/government/organisations/innovate-uk [accessed 2020-09-02]
8. City Deals. Government of UK. URL: https://www.gov.uk/government/collections/city-deals [accessed 2020-09-02]
9. Gawande A. Why doctors hate their computers. New Yorker Nov 2018:12.
10. Fagerberg J, Srholec M. National innovation systems, capabilities and economic development. Res Policy 2008 Oct;37(9):1417-1435. [doi: 10.1016/j.respol.2008.06.003]
11. Hyysalo S. -. Health technology development and use: from practice-bound imagination to evolving impacts. Routledge; 2010.
12. Williams R. Why is it difficult to achieve e-health systems at scale? Inf Commun Soc 2016 Jan 11;19(4):540-550. [doi: 10.1080/1369118x.2015.1118521]
13. Herzlinger RE. Why innovation in health care is so hard. Harv Bus Rev 2006 May;84(5):58-66, 156. [Medline: 16649698]
14. Uyarra E, Edler J, Garcia-Estevez J, Georghiou L, Yeow J. Barriers to innovation through public procurement: a supplier perspective. Technovation 2014 Oct;34(10):631-645. [doi: 10.1016/j.technovation.2014.04.003]
15. Pollock N, Williams R. The business of expectations: how promissory organizations shape technology and innovation. Soc Stud Sci 2010 May 20;40(4):525-548. [doi: 10.1177/0306312710362275]
16. Phillips F, Garman A. Barriers to entrepreneurship in healthcare organizations. J Health Hum Serv Adm 2006;28(4):472-484. [Medline: 16583847]
17. Barnett J, Vasileiou K, Djemil F, Brooks L, Young T. Understanding innovators' experiences of barriers and facilitators in implementation and diffusion of healthcare service innovations: a qualitative study. BMC Health Serv Res 2011 Dec 16;11:342 [FREE Full text] [doi: 10.1186/1472-6963-11-342] [Medline: 22176739]
18. Bates DW, Sheikh A, Asch DA. Innovative environments in health care: where and how new approaches to care are succeeding. Health Aff (Millwood) 2017 Mar 1;36(3):400-407. [doi: 10.1377/hlthaff.2016.1311] [Medline: 28264940]
19. Marshall MN. Sampling for qualitative research. Fam Pract 1996 Dec;13(6):522-525. [doi: 10.1093/fampra/13.6.522] [Medline: 9023528]
20. Ritchie J, Lewis J, Nicholls C, Ormston R. editors. Qualitative research practice: A guide for social science students and researchers. Sage; 2013:A.
21. Fusch P, Ness L. Are we there yet? Data saturation in qualitative research. Qual Res 2015;20(9):1408-1416. [doi: 10.4135/9781412963909.n99]
22. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
23. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
24. Cresswell K, Cunningham-Burley S, Sheikh A. Creating a climate that catalyses healthcare innovation in the United Kingdom - learning lessons from international innovators. J Innov Health Inform 2017 Jan 25;23(4):882 [FREE Full text] [doi: 10.14236/jhi.v23i4.882] [Medline: 28346133]
25. van der Panne G, van Beers C, Kleinknecht A. Success and failure of innovation: a literature review. Int J Innov Mgt 2011 Nov 20;07(03):309-338. [doi: 10.1142/s1363919603000830]
26. Berry LL, Shankar V, Parish JT, Cadwallader S, Dotzel T. Creating new markets through service innovation. MIT Sloan management review 2006;47(2):56.
27. von Hippel E. Democratizing innovation: the evolving phenomenon of user innovation. JfB 2005 Mar;55(1):63-78. [doi: 10.1007/s11301-004-0002-8]
28. Chesbrough H, Vanhaverbeke W, West J. editors. Open innovation: Researching a new paradigm. Oxford University Press on Demand; 2006.
29. Mulgan G. The process of social innovation. Innovations 2006 Apr;1(2):145-162. [doi: 10.1162/itgg.2006.1.2.145]
30. Transformative Healthcare Technologies for 2050 Healthcare Technologies Information Day Birmingham Botanical Gardens 27 March 2019. The Engineering and Physical Sciences Research Council (EPSRC). URL: https://epsrc.ukri.org/files/funding/calls/2019/transformativehealthcareinformationday/ [accessed 2020-09-02]

31. Guidelines - Metrics & Milestones For Successful Incubator Development. The Aspen Institute. URL: https://assets.aspeninstitute.org/content/uploads/files/content/docs/resources/A%20White%20Paper_Metrics%20%20Milestones%20for%20Incubators.pdf [accessed 2020-09-02]
32. Spinardi, GrahamWilliams, Robin, (2005) ?. Chap. 3. Lyall, Catherine and Tait, Joyce (eds) New modes of governance : developing an integrated policy approach to science, technology, risk and the environment. Aldershot, Ashgate, pp 2005:45. [doi: 10.4324/9781315248042-3]
33. Tian X, Wang TY. Tolerance for failure and corporate innovation. Rev Financ Stud 2011 Dec 5;27(1):211-255. [doi: 10.1093/rfs/hhr130]
34. Rothwell R. Towards the fifth ‐ generation innovation process. Int Mark Rev 1994 Feb;11(1):7-31. [doi: 10.1108/02651339410057491]

Edited by G Eysenbach; submitted 26.04.20; peer-reviewed by E Maldonado, C Kaylor-Hughes, R Lee; comments to author 12.06.20; revised version received 16.06.20; accepted 07.07.20; published 25.09.20.

Please cite as:
Cresswell K, Williams R, Carlile N, Sheikh A
Accelerating Innovation in Health Care: Insights From a Qualitative Inquiry Into United Kingdom and United States Innovation Centers
J Med Internet Res 2020;22(9):e19644
URL: http://www.jmir.org/2020/9/e19644/
doi:10.2196/19644
PMID:32975524

Letter to the Editor

Comment on “Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine”

Jamal Rahmani[1], PhD; Roya Karimi[1], PhD; Yousef Khani[1], PhD; Siamak Sabour[1,2], MD, PhD

[1]Department of Clinical Epidemiology, School of Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[2]Safety Promotions and Injury Prevention Research Centre, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Corresponding Author:
Siamak Sabour, MD, PhD
Department of Clinical Epidemiology
School of Health and Safety
Shahid Beheshti University of Medical Sciences
Chamran Highway, Velenjak, Daneshjoo Blvd
Tehran, 198353-5511
Iran
Phone: 98 2122421814
Email: s.sabour@sbmu.ac.ir

Related Article:

Comment on: https://www.jmir.org/2019/5/e13260/

(J Med Internet Res 2020;22(9):e14944) doi:10.2196/14944

KEYWORDS

prediction; area under the curve; AUC; lung cancer

We read the recent article by Wang et al [1] with great interest. This paper was published in 2019 in the *Journal of Medical Internet Research*. The authors aimed to develop and validate a prospective risk prediction model to identify patients at risk of new incident lung cancer within the next 1 year in the general population. They used data from individual patient electronic health records (EHRs), which was extracted from the Maine Health Information Exchange network. The Extreme Gradient Boosting (XGBoost) algorithm was adopted to build the model, and the authors reported an area under the curve (AUC) of 0.88 (95% CI 0.87-0.88) for their model validation, according to a prospective cohort data. Finally, the authors concluded that their model was able to identify statewide, high-risk patients.

Risk prediction models are effectively useful due to their role in decision making. However, there are some methodological commentaries that we would like to mention. First, AUC is an appropriate measure for assessing discrimination. Discrimination is defined as the ability to distinguish events versus nonevents. However, it assumes that two persons are randomly selected—one who will develop the disease and one who will not. AUC assigns a higher probability of an outcome to the one who will develop the disease. A c-index value of 0.5 expresses a random chance; however, the usual c-index for a prediction model is 0.60 to 0.85. This range can be changeable under different conditions. What we should always consider about the AUC measure is that a high value of AUC discerns excellent discrimination, but it can also reflect a situation with limited relevance. This situation might arise because the variable is related to the diagnostic or early onset of the disease instead of prediction [2,3]. Furthermore, the receiver operating characteristic (ROC) would be a good tool for binary classification, but it is not instrumental for risk stratification. For risk stratification (low- and high-risk bins), the sensitivity in low and high specificity, and positive predictive value (PPV) in high-risk bins, are more discriminating parameters for the ability of the algorithm.

Second, there are several types of external validation such as validation in more recent patients (temporal validation), in other places (geographic validation), or by other investigators at other sites (fully independent validation). Having two exemplary data sets with huge sample sizes, it would be suggestible to test the above-mentioned external validity. Moreover, internal validation is a necessary part of model development. It determines the reproducibility of a developed prediction model for the derivative sample and prevents the over-interpretation of the data. Resampling techniques, such as cross-validation and bootstrapping, can be performed; bootstrap validation, in particular, appears to be the most attractive option for obtaining

stable optimism-corrected estimates [2]. Furthermore, it is of importance that the authors add the validation of data production in the real world after deployment, since it would be more revealing due to the unexpected data challenges encountered during real-time usage by clinical providers.

Third, a mistake that is very common occurs when referring to statistically significant *P* values. A *P* value depends on statistical, instead of clinical, logic; thus, researchers should consider judging outputs based on effect size, rather than *P* value.

A further common issue is missing data that can influence the model development. Missing data often follow a nonrandom pattern, where there is an explanation and cause behind it. If all missing values are removed, the cause and explanation will be lost, which may affect the conclusion and the model development. To generate the model, multivariable regression techniques usually use as a stepwise model (backward is more preferable), and concomitantly checking the Akaike information criterion can help us to decide if the model fits well enough.

Finally, it is important to investigate the interactions between variables in prediction studies. Developing a model, score, or index without considering interactions among variables may elicit changes to the prediction in the real world and lead to misleading messages [3-5].

Editorial notice: The corresponding author of "Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine" did not respond to our invitation to reply to this commentary.

Conflicts of Interest

None declared.

References

1. Wang X, Zhang Y, Hao S, Zheng L, Liao J, Ye C, et al. Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine. J Med Internet Res 2019 May 16;21(5):e13260 [FREE Full text] [doi: 10.2196/13260] [Medline: 31099339]
2. Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical Prediction Models Involving Radiological Parameters: Study Design and Statistical Methods. Korean J Radiol 2016;17(3):339-350 [FREE Full text] [doi: 10.3348/kjr.2016.17.3.339] [Medline: 27134523]
3. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009 Feb 23;338:b375. [doi: 10.1136/bmj.b375] [Medline: 19237405]
4. Sabour S. Prediction of preterm delivery using levels of VEGF and leptin in amniotic fluid from the second trimester: prediction rules. Arch Gynecol Obstet 2015 Apr;291(4):719. [doi: 10.1007/s00404-014-3568-y] [Medline: 25490880]
5. Sabour S, Ghassemi F. Predictive value of confocal scanning laser for the onset of visual field loss. Ophthalmology 2013 Jun;120(6):e31-e32. [doi: 10.1016/j.ophtha.2013.01.055] [Medline: 23732064]

Abbreviations

AUC: area under the curve
EHR: electronic health record
PPV: positive predictive value
ROC: receiver operating characteristic

Edited by A Moorhead; submitted 05.06.19; peer-reviewed by X Ling, J Liao; comments to author 06.08.19; revised version received 06.08.19; accepted 17.09.19; published 15.09.20.

Please cite as:
Rahmani J, Karimi R, Khani Y, Sabour S
Comment on "Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine"
J Med Internet Res 2020;22(9):e14944
URL: https://www.jmir.org/2020/9/e14944
doi:10.2196/14944
PMID:32930665

Letter to the Editor

Comment on "Designing Robust N-of-1 Studies for Precision Medicine: Simulation Study and Design Recommendations"

Reid D Landes[1], PhD

Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR, United States

Corresponding Author:
Reid D Landes, PhD
Department of Biostatistics
University of Arkansas for Medical Sciences
4301 W Markham St
Slot 781
Little Rock, AR, 72205
United States
Phone: 1 501 526 6714
Email: rdlandes@uams.edu

Related Article:

Comment on: https://www.jmir.org/2019/4/e12641/

(J Med Internet Res 2020;22(9):e16179) doi:10.2196/16179

KEYWORDS

sample size; misleading statements

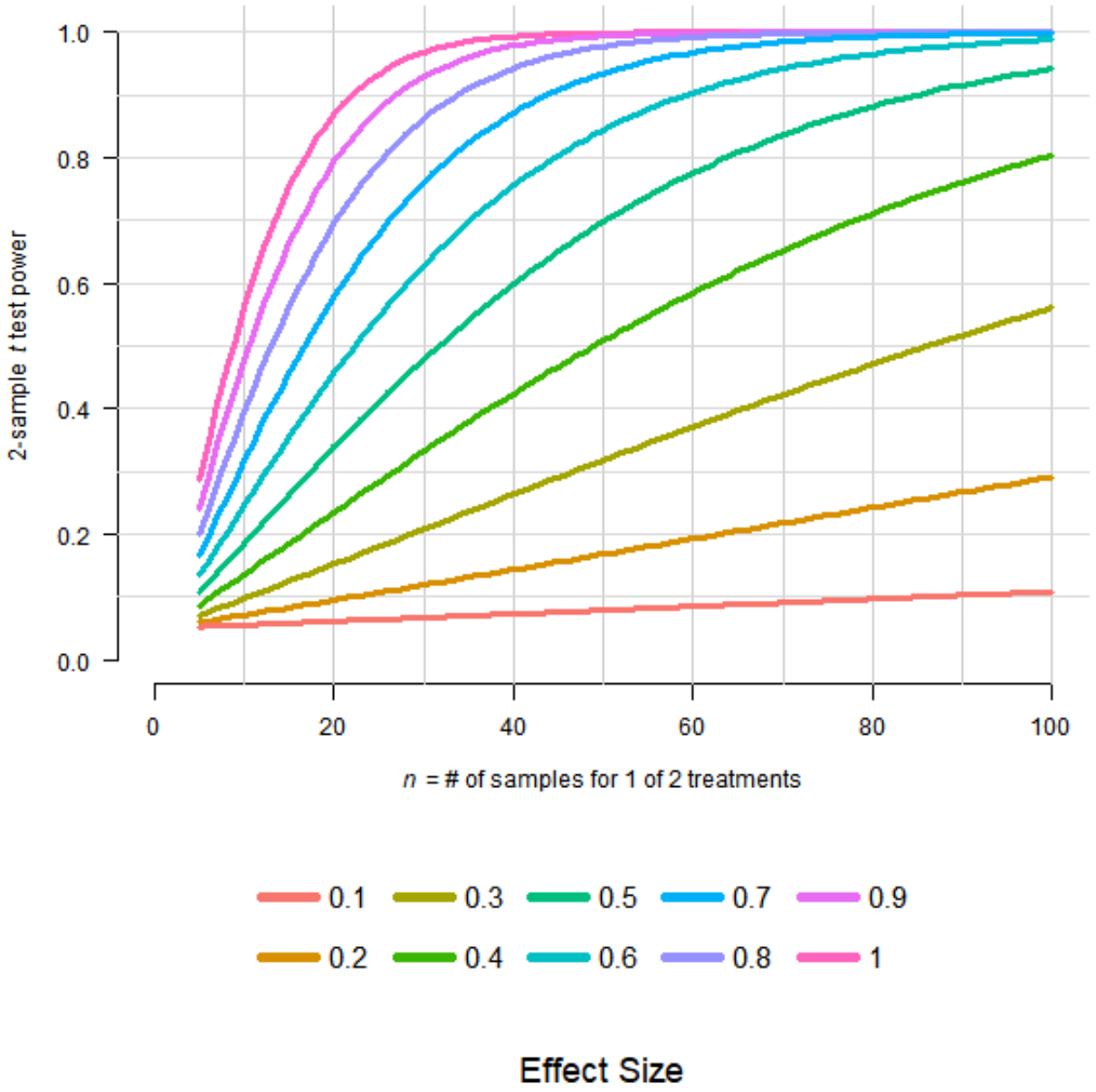
In "Designing robust N-of-1 studies for precision medicine: Simulation study and design recommendations" by Percha et al [1], the authors use misleading language when speaking about the required numbers of samples regarding results in Figure 4a. For example, they write on page 8:

> *In Figure 4a, we see that for effect sizes of 0.1, 0.2, and 0.3, more than 100 samples are needed to obtain a power of 0.8 (at a standard 5% significance level). For an effect size of 0.4, at least 100 samples are needed. For effect sizes of 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0, the numbers of samples needed to attain a power of 0.8 are approximately 65, 45, 35, 26, 21, and 18, respectively. [Figure 4]*

Since Figure 4a is exactly equivalent to power curves from a two-sample, equal-variance *t* test (see Figure 1; generating R code provided in Textbox 1), the numbers of samples are for *one* of the two treatments; thus, the total numbers of samples are doubled. An easy fix in most instances of the unclear language is to add "per treatment" after "samples." I provide a list of potential clarifying edits to the article's text below (but may have missed some instances):

- Figure 4c caption: "(ie, number of samples *per treatment*, with sampling rate fixed at 1 sample per time unit)"
- Figure 4a and 4b: the label for the horizontal axis should be "Number of samples *per treatment*"
- Page 8: "In Figure 4a, we see that for effect sizes of 0.1, 0.2, and 0.3, more than 100 samples *per treatment* are needed to obtain a power of 0.8 (at a standard 5% significance level). For an effect size of 0.4, at least 100 samples *per treatment* are needed. For effect sizes of 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0, the numbers of samples *per treatment* needed to attain a power of 0.8 are approximately 65, 45, 35, 26, 21, and 18, respectively."
- Page 9: "For an effect size of 0.5 and σ_p=0.0, 0.4, 0.8, 1.2, 1.6, 2.0, the numbers of samples *per treatment* needed to obtain a power of 0.8 are 61, 76, 89, 111, 135, and 176, respectively. For an effect size of 1.0, the numbers of samples *per treatment* needed are 20, 24, 28, 34, 43, and 53, respectively."
- Page 9: "For σ_p=0.0, 0.4, 0.8, 1.2, 1.6, 2.0 and α=0.1, the numbers of samples *per treatment* required are 36, 64, 110, 174, 228, and 250, respectively. For α=10.0, the numbers of samples *per treatment* required are only 20, 23, 28, 34, 42, and 53, respectively."

Figure 1. For effect sizes ranging from 0.1 to 1.0, power of a 0.05 level two-sample *t* test plotted by *n*, the number of samples in one treatment group. Total sample size is assumed to be *2n*.



Textbox 1. R code.

```
#--- R code for generating Figure 1.
library(pwr)
color <- c("#F8766D", "#D89000", "#A3A500", "#39B600", "#00BF7D", "#00BFC4",
          "#00B0F6", "#9590FF", "#E76BF3", "#FF62BC")

#--- Producing the power plot (upper portion of figure)
plot(50, 1, type = 'n', xlim = c(0,100), ylim = c(0,1), axes = FALSE,
    ylab = substitute(paste("2-sample ", italic('t'), " test power")),
    xlab = substitute(paste(italic('n'), " = # of samples for 1 of 2 treatments") ) )
axis(1, at = seq(0, 100, 20)); axis(2, at = seq(0, 1, .2), las = 2)
abline(v = seq(10, 100, 10), col = 'gray80')
abline(h = seq(.1, 1, .1), col = 'gray85')

#--- Filling in the power plot
for(.d in 10:1/10){
  tmp.power <- NULL
  for( .n in 5:100){
    #--- Computes power for a 2-sample t-test, each sample with n observations.
    p <- pwr.t.test(n = .n, d = .d, sig.level = .05, power = NULL,
                type = "two.sample", alternative = "two.sided")$power
    tmp.power <- c(tmp.power, p)
  }
  lines(5:100, tmp.power, lwd=3, col=color[.d*10])
}

#--- Producing the the legend (lower portion of figure)
plot(50, 1, type = 'n', xlim = c(0,100), ylim = c(.25, .75),
     axes = FALSE, xlab = 'Effect Size', ylab = '', cex.lab=1.5)
x <- 15
for(iter in seq(2,10,2)){
  mult <- (iter/2)
  .x <- x*mult
  segments( .x-3, 0.67, .x+3, lwd=5, col= color[iter-1])
  text(.x+3, .67, pos=4, (iter-1)/10, cex=1.25)
  segments( .x-3, 0.33, .x+3, lwd=5, col= color[iter] )
  text(.x+3, .33, pos=4, (iter)/10, cex=1.25)
}
```

Percha and colleagues have agreed to the above changes; these changes have been made to the original paper.

Conflicts of Interest

None declared.

Reference

1. Percha B, Baskerville EB, Johnson M, Dudley JT, Zimmerman N. Designing Robust N-of-1 Studies for Precision Medicine: Simulation Study and Design Recommendations. J Med Internet Res 2019 Apr 01;21(4):e12641 [FREE Full text] [doi: 10.2196/12641] [Medline: 30932871]

Edited by G Eysenbach, T Derrick; submitted 07.09.19; peer-reviewed by B Percha; accepted 10.08.20; published 15.09.20.

Please cite as:
Landes RD
Comment on "Designing Robust N-of-1 Studies for Precision Medicine: Simulation Study and Design Recommendations"
J Med Internet Res 2020;22(9):e16179
URL: https://www.jmir.org/2020/9/e16179
doi:10.2196/16179
PMID:32930671

Corrigenda and Addenda

Correction: Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models

Dianbo Liu[1,2*], PhD; Leonardo Clemente[1,2,3*], MSc; Canelle Poirier[1,2*], PhD; Xiyu Ding[1,4], MSc; Matteo Chinazzi[5], PhD; Jessica Davis[5], BSc; Alessandro Vespignani[5,6], PhD; Mauricio Santillana[1,2,4], PhD

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

[2]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[3]Tecnologico de Monterrey, Monterrey, Mexico

[4]Harvard TH Chan School of Public Health, Boston, MA, United States

[5]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, United States

[6]ISI Foundation, Turin, Italy

[*]these authors contributed equally

Corresponding Author:
Mauricio Santillana, PhD
Computational Health Informatics Program
Boston Children's Hospital
300 Longwood Avenue
Landmark 5th Floor East
Boston, MA, 02215
United States
Phone: 1 (617) 919 1795
Email: msantill@g.harvard.edu

Related Article:

Correction of: https://www.jmir.org/2020/8/e20285/

(J Med Internet Res 2020;22(9):e23996) doi:10.2196/23996

In "Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models" (J Med Internet Res 2020;22(8):e20285) the authors noted three errors.

The order of authors in the original article was listed as:

> *Canelle Poirier, Dianbo Liu, Leonardo Clemente, Xiyu Ding, Matteo Chinazzi, Jessica Davis, Alessandro Vespignani, Mauricio Santillana*

The correct order of authors is:

> *Dianbo Liu, Leonardo Clemente, Canelle Poirier, Xiyu Ding, Matteo Chinazzi, Jessica Davis, Alessandro Vespignani, Mauricio Santillana*

As well, degree information for authors Leonardo Clemente, Xiyu Ding, and Jessica Davis was listed incorrectly as "MD" in the original article. The correct degree for Leonardo Clemente and Xiyu Ding is "MSc"; the correct degree for Jessica Davis is "BSc".

The correction will appear in the online version of the paper on the JMIR Publications website on September 22, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 31.08.20; this is a non–peer-reviewed article;accepted 01.09.20; published 22.09.20.

Please cite as:
Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J, Vespignani A, Santillana M
Correction: Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models
J Med Internet Res 2020;22(9):e23996
URL: http://www.jmir.org/2020/9/e23996/
doi:10.2196/23996
PMID:32960774

Viewpoint

Digital Response During the COVID-19 Pandemic in Saudi Arabia

Marwah Hassounah[1], MBBS, MPH; Hafsa Raheel[1], MBBS, FCPS, MCPS; Mohammed Alhefzi[2,3], MD, SBCM–HI, MSc, FAIDH

[1]Prince Sattam Chair for Epidemiology and Public Health Research, Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[2]Preventive Medicine and Clinical Informatics, King Faisal Medical City for Southern Regions, Abha, Saudi Arabia

[3]Saudi Association for Health Informatics, Riyadh, Saudi Arabia

Corresponding Author:
Hafsa Raheel, MBBS, FCPS, MCPS
Prince Sattam Chair for Epidemiology and Public Health Research, Department of Family and Community Medicine
College of Medicine
King Saud University
Room No. 321/27, Third Floor
Medical Education Department
Riyadh,
Saudi Arabia
Phone: 966 556942926
Email: hafsaraheel@yahoo.com

Abstract

Background: The first case of COVID-19 in Saudi Arabia was confirmed on March 3, 2020. Saudi Arabia, like many other countries worldwide, implemented lockdown of most public and private services in response to the pandemic and established population movement restrictions nationwide. With the implementation of these strict mitigation regulations, technology and digital solutions have enabled the provision of essential services.

Objective: The aim of this paper is to highlight how Saudi Arabia has used digital technology during the COVID-19 pandemic in the domains of public health, health care services, education, telecommunication, commerce, and risk communication.

Methods: We documented the use of digital technology in Saudi Arabia during the pandemic using publicly available official announcements, press briefings and releases, news clips, published data, peer-reviewed literature, and professional discussions.

Results: Saudi Arabia's government and private sectors combined developed and launched approximately 19 apps and platforms that serve public health functions and provide health care services. A detailed account of each is provided. Education processes continued using an established electronic learning infrastructure with a promising direction toward wider adoption in the future. Telecommunication companies exhibited smooth collaboration as well as innovative initiatives to support ongoing efforts. Risk communication activities using social media, websites, and SMS text messaging followed best practice guides.

Conclusions: The Saudi Vision 2030 framework, released in 2017, has paved the path for digital transformation. COVID-19 enabled the promotion and testing of this transition. In Saudi Arabia, the use of artificial intelligence in integrating different data sources during future outbreaks could be further explored. Also, decreasing the number of mobile apps and merging their functions could increase and facilitate their use.

(J Med Internet Res 2020;22(9):e19338) doi:10.2196/19338

KEYWORDS

digital response; COVID-19; Saudi Arabia; digital health; containment; public health; pandemic; prevention

Introduction

The outbreak of SARS-CoV-2, emerging from the markets of Wuhan, led to the COVID-19 pandemic [1,2]. The current population-wide measures of home quarantine that were simultaneously applied worldwide to slow and prevent the spread of COVID-19 are unprecedented. The COVID-19 pandemic has caused disruption of daily services due to the community-wide mitigation measures taken by many countries. Due to the low likelihood of obtaining a vaccine in the near future, global efforts have vastly focused on social distancing and complete city and state lockdowns in many instances as the

only solutions to contain the pandemic [3]. These mitigation measures have necessitated the use of technology to maintain functions in all aspects of life.

The global experiences with the H1N1 influenza pandemic in 2009 [4] and Ebola virus in 2014 clearly indicated that timely and appropriate technology usage played a considerable role in controlling these pandemics [5-7]. A cloud computing tool for data collection and integration for confirmed cases of Middle East respiratory syndrome coronavirus (MERS-CoV), a GPS-based risk assessment tool [8], and Google Maps usage for the geographical representation of MERS-CoV cases worldwide are examples of the technological methods used to control the outbreak of Middle East respiratory syndrome (MERS) [8]. By establishing a national electronic surveillance system [9,10], Saudi Arabia also contributed to the global data pool of MERS-CoV information.

During the current COVID-19 pandemic, Saudi Arabia has been proactive in implementing disease containment measures and working to meet the community's needs and demands in a very short time [3]. It is currently estimated that 30,260,000 people in Saudi Arabia (89% of the population) use the internet, 96% of the population uses smartphones [11], and the majority of the population now has access to smartphones, laptop computers, desktop computers, and tablets; therefore, digital service provision is much easier than in the past and has aided the mitigation efforts established by the government.

Keeping in view the importance of quick and timely digital data sharing for policy actions, which is also emphasized by the World Health Organization (WHO) [12], our aim in this paper is to highlight how Saudi Arabia has used digital technology during the COVID-19 pandemic.

Methods

The authors documented Saudi Arabia's experience using publicly available official announcements, press briefings and releases, news clips, published data, peer-reviewed literature, and professional discussions. The searched information sources were in both English and Arabic languages. A literature search was conducted from March 20 to June 20, 2020. Each author collected, examined, and synthesized information on a designated sector; then, all the authors consolidated, discussed, and agreed on the final findings. The inclusion criteria for information were that the information depicted a prominent event during the COVID-19 pandemic response, included technology or digitalization, and was specific to Saudi Arabia. The findings are mainly presented in narrative form.

Results

Figure 1 and Table 1 visualize and summarize some of the applications and platforms used for various health sectors during the COVID-19 pandemic in Saudi Arabia.

Figure 1. Examples of digital apps available for various health care domains during the COVID-19 pandemic in Saudi Arabia.

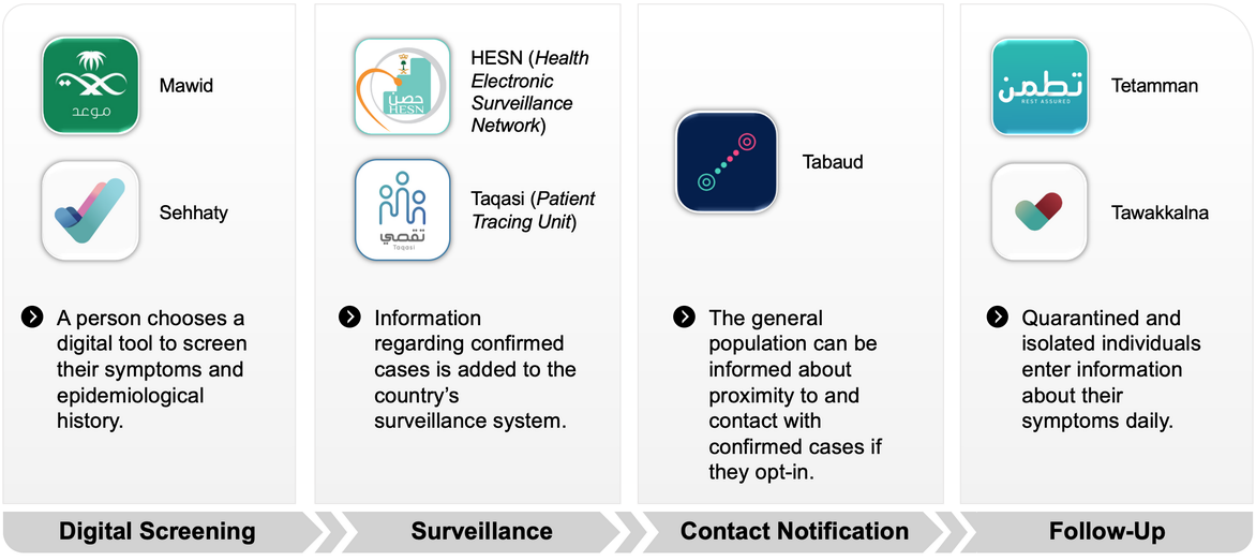


Table 1. Summary of telehealth applications available in Saudi Arabia during the COVID-19 pandemic.

Name	Type	Short description	Provider
Sehha [13]	Smartphone app (iOS/Android)	Teleconsultation (synchronous live video chat, nonsynchronous SMS text messaging).	Ministry of Health
Mawid [14]	Smartphone app (iOS/Android), web-based application	Symptom checker/appointment gateway to all Saudi Ministry of Health Services. Used as the main channel for the virtual COVID-19 screening/triaging by the Saudi Ministry of Health.	Ministry of Health
Anat [15]	Smartphone app (iOS/Android)	E-prescription[a] gateway; licensure of all health care professionals is checked with the Saudi Commission for Health Specialties.	Ministry of Health
Wasfaty [16]	Web-based	The official e-prescription gateway provided by the Ministry of Health.	Ministry of Health
Asefni [17]	Smartphone app (iOS/Android)	GPS-enabled requests for emergency services nationwide.	Saudi Red Crescent Authority
Cura [18]	Smartphone app (iOS/Android)	Teleconsultations (synchronous live video chat, nonsynchronous SMS text messaging, more specific subspecialties/for-profit).	Private
MayaClinic [19]	Smartphone app (iOS/Android)	Teleconsultations (nonsynchronous text messaging with health care providers).	Private
Nala [20]	Smartphone app (iOS/Android)	Teleconsultations (artificial intelligence-enabled chatbot provides decision support for the public, acts as an appointment gateway and nonsynchronous messaging with health care providers).	Private
Labayh [21]	Smartphone app (iOS/Android)	Teleconsultations (mainly provides psychology sessions and mental health services).	Private
80/20 Lifestyle [22]	Smartphone app (iOS/Android)	Remote patient engagement and lifestyle change recommendations.	Private
Virtual Medical Academy [23]	Web-based interactive academy	Videoconferencing events targeted to health care professionals.	Private
SCFHS Webinars [24]	Web-based seminars	Offers support services to health care professionals under their programs (Daem for residents and Emtenan for all health care professionals).	Saudi Commission for Health Specialties

[a]e-prescription: electronic prescription.

Health Sector Digitalization

Amid the COVID-19 pandemic, the Saudi Ministry of Health has implemented multiple informatics tools to provide public health information for individuals as well as the community.

Public Health Informatics Tools

In 2018, the Ministry of Health launched a national central health care appointment gateway through a mobile app and web-based application called Mawid, which translates to "Appointment" [14] (Figure 1). Soon after, in August 2019, the Sehhaty ("My Health") app was launched in the pursuit of a wide range of health promotional campaigns that target healthy lifestyles, using gamification and community-wide challenges [25]. However, both apps were updated to respond to the COVID-19 pandemic by introducing a symptom checker to enable people who suspect they have COVID-19 to directly book appointments at dedicated COVID-19 clinics [26] and drive-through mass testing locations around the Kingdom [27].

For COVID-19 surveillance, the Health Electronic Surveillance Network (HESN) has been mainly used as a reliable source of data for all COVID-19 laboratory tests in the Kingdom. The HESN serves as a national communicable disease surveillance platform. It was launched in 2012 and piloted during the largest public health event in Saudi Arabia: the annual pilgrimage season, or the Hajj [28]. Moreover, the Patient Tracing Unit (Taqasi) platform was implemented in March 2020 for the COVID-19 pandemic. Its purpose is to enhance and manage contact tracing around the Kingdom based on the laboratory results generated from the HESN.

Locally published preventative and clinical guidelines give directions for home isolation with documented daily follow-ups and for tracking symptoms for mild cases and contacts. To provide these functions, the National Health Emergency Operation Centre launched a smartphone app, Tetamman, which translates as "Rest Assured" [29]. In May 2020, the Ministry of Health announced that the Tetamman app will also be associated with a smart bracelet for individuals returning from abroad as well as those who are isolated in their homes [30].

Contact tracing has been termed as an essential epidemiologic tool for containing the COVID-19 outbreak and enforcing future plans for lifting lockdown safely. To achieve this, the Saudi

Data and Artificial Intelligence Authority (SDAIA) released two smartphone apps. The first is Tawakkalna, a GPS-enabled app to monitor and restrict individuals' movement during curfew hours with the capacity to issue permits for exceptions. The second app, Tabaud, whose name means "Distancing" [31], sends deidentified data to people who came in close contact with confirmed cases of COVID-19. The app follows the international Google and Apple guidelines on data privacy.

Health Care Delivery

The Saudi Ministry of Health (MOH), as the main health care provider in the Kingdom of Saudi Arabia, is looked upon as the main source of authentic and reliable health information for the Saudi population. Other channels of health care delivery include the Ministry of Defense, university teaching hospitals, and the private sector. Similarly, tertiary, secondary, and primary care facilities provide health care to both nationals and nonnationals. In 2011, the Saudi MOH agreed upon a vision to improve the standards, equitability, availability, and quality of health care in the Kingdom of Saudi Arabia by the use of electronic communication and information technology in this sector. The Vision 2030 National Transformation Program [32] health care strategic objectives for the years 2018 to 2020 aimed to increase access to care, improve quality, and promote the prevention of health risks. It highlights electronic health (eHealth) as an essential enabler to the health care transformation; hence, it tasks the National Health Information Centre with creating multisectoral coherent eHealth services. During the aforementioned community-wide measures to combat the spread of COVID-19, the government of Saudi Arabia and the private health care sector activated existing digital health solutions and produced new ones.

The MOH Call (937) Service Center was established to answer inquiries related to COVID-19. Moreover, one hospital initiated a remotely controlled robot for rounding and monitoring of intensive care unit patients [33].

For hospitals that had teleconsultation apps in place before COVID-19, whether well-established or in pilot phases, some activated their apps to serve patients who do not require in-person hospital visits, such as King Saud Medical City in the public sector and Dr Sulaiman Al Habib Medical Group in the private sector. The group messaging app WhatsApp remains the preferred messaging app in the Kingdom [11]; some hospitals and medical cities in the Eastern Region, such as Qatif Central Hospital and its primary care centers, initiated WhatsApp numbers to help patients register their medication refill requests, arrange for remote routine follow-ups, and inquire regarding their laboratory results.

Periods of pandemics have been shown to cause a surge in stress related to fear of the unknown and isolation. Hence, the literature highlights the need for establishment of psychological support to communities during such periods. To achieve this objective, the National Centre for Mental Health Promotion collaborated with the developers of a local mobile counselling app, Labayh, to provide free sessions for people experiencing anxiety and panic symptoms in the current situation [21]. In another instance, the Saudi Commission for Health Specialties (SCFHS) nationally launched a set of mental health support services for all health care professionals in the Kingdom under its Emtenan initiative as well as for residents in training (Daem) [24]. The SCFHS not only called its registered health care professionals but also sent them SMS text messages, enquiring about their safety and advising them to keep safe.

In the area of telepharmacy, the MOH and other tertiary health care facilities sent medications to patients' homes via courier companies or established telepharmacy services [15]. Furthermore, the MOH sent out SMS text reminders to all health care providers with active professional registration to use its electronic prescription (e-prescription) services in collaboration with private sector pharmacies. One example is the Anat mobile app [15], which enables providers to directly electronically prescribe medications to patients by credentialed and licensed providers. Other apps that have been active in recent years in Saudi Arabia are Wasfaty [16], translated as "My Prescription," which is the official gateway for e-prescriptions under the Ministry of Health's free services, and the Sehha [13] mobile teleconsultation app, which can provide patients with e-prescriptions via SMS text message following a medical consultation with a physician. Private telehealth services such as Cura [18] and Maya Clinic [19] are offering similar services either freely or with modest charges to support their COVID-19 efforts.

With the success of telemedicine services, King Salman bin Abdulaziz Al Saud of Saudi Arabia issued a royal decree to amend health professionals' practice regulations to allow telemedicine use for diagnostic and management purposes from the workplace and at home. This royal order also directs all relevant sectors to amend their regulations to accommodate this change [34].

Educational Sector Digitalization

According to Saudi national statistics [35], approximately 1,353,619 students are enrolled in 28 governmental and 34 private higher education institutes. Moreover, there are approximately 5000 schools in the Kingdom that provide secondary level education; these include both public and private sector institutions.

Electronic learning (e-learning) is not new in the Kingdom. Its first decade (1990-2000) in Saudi Arabia's education system was supported well by the evolution of computer technology and the World Wide Web [36]. By 2002, Saudi Arabia had established a national school e-learning platform with tailored electronic lessons [36]. The following years witnessed expansion and enhancement of e-learning in collaboration with international partners [36]. In 2017, as part of Vision 2030, the Ministry of Education (MOE) established the National Center for e-Learning [37]. This center serves to supervise and support eLearning in Saudi Arabia. The current COVID-19 pandemic poses immense challenges to maintaining continuity of educational services across the Kingdom. This challenge was most evident in the health educational sector due to the absence of a standard and unified method of eLearning and because educational methods depend on patient interactions.

It is already known that major universities in the Kingdom such as King Saud University, Taibah University, King Khalid

University, Qassim University, Islamic University of Madinah, Al-Baha University, and King Abdul-Aziz University are the most active e-learning university partners in the Kingdom. However, higher education institutions were challenged by the COVID-19 situation to continue tutoring and assessment of technical skills [38]. Hence, universities offered different methods of e-learning support depending on the course requirements and interim assessment needs [39].

Later, the Minister of Education congratulated higher educational institutes on their successful shift to distance learning since the COVID-19 outbreak. The universities reported that collectively, 1.2 million users were conducting 107,000 hours of web-based learning in more than 7600 virtual classes [40]. The MOE also directed higher education students and faculty to its website "Shams," an open education resource.

The aforementioned SCFHS is Saudi Arabia's accreditation and registration body for health care professionals. It offers a series of accredited educational webinars for continuing medical education hours. The SCFHS has collaborated with local and international platforms such as Virtual Medical Academy [23], UpToDate [41], and MDBriefcase [42]. The topics listed therein include a COVID-19 overview, physician burnout, a COVID-19 critical care crash course, and ethical issues during pandemics using COVID-19 as an example [43]. Other nonprofit public and private initiatives have followed the SCFHS's lead [43,44].

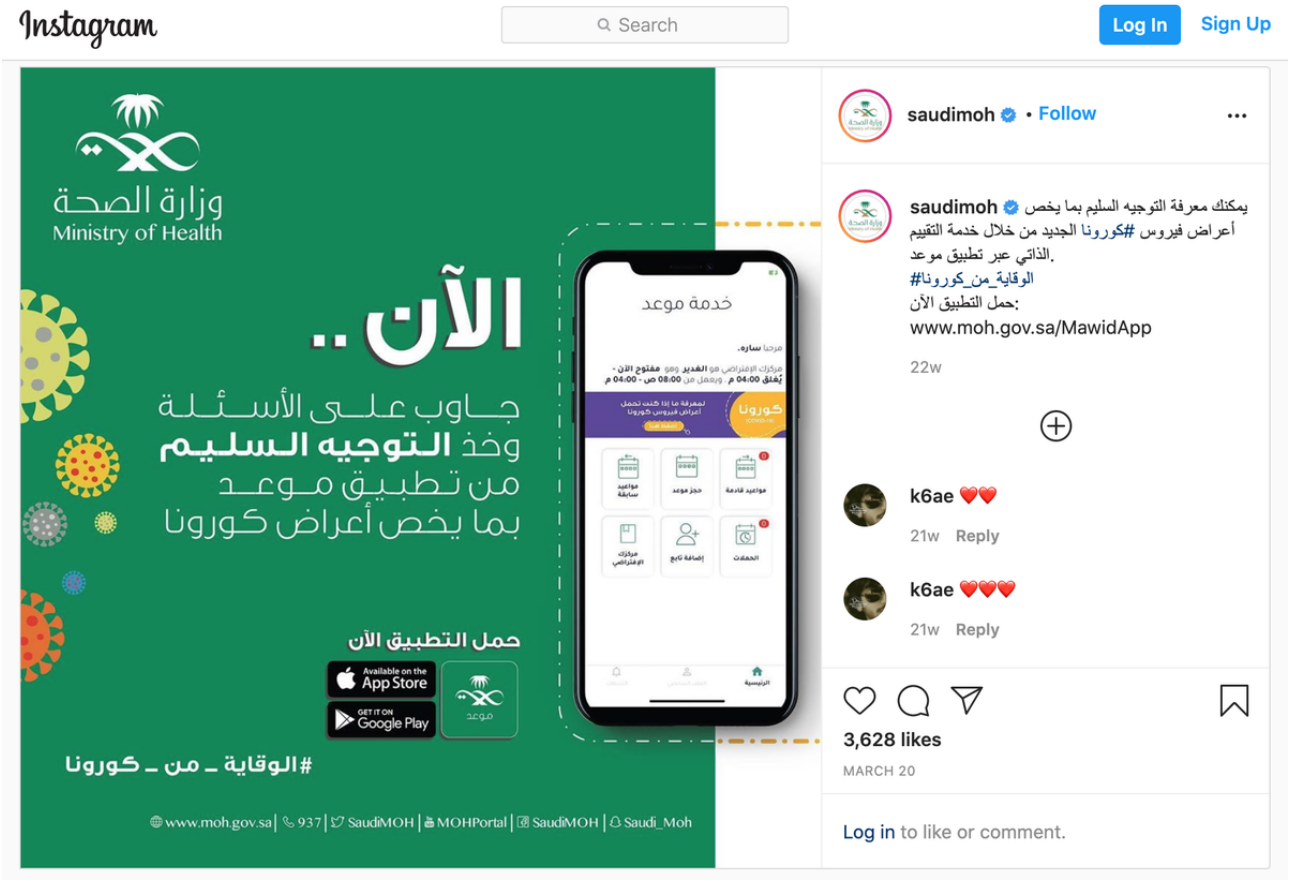
It is worth noting that all educational institutions, including higher education institutions, continued delivery of education during lockdown. Both public and private institutions used various two-way e-learning methods to continue teaching and student learning. This ranged from individual institute-based platforms such as Blackboard and McGraw-Hill Connect to common commercial platforms such as Zoom, Google Class, and FaceTime.

An interesting and unique step taken by the MOE was to shift public school education to its distant learning portals, namely Ein (translated as "Eye") and Vschool.sa [45]. Ein, which was launched by the Ministry of Education before the COVID-19 outbreak, features a television channel that broadcasts daily lessons based on the national curriculum [39]. During the COVID-19 pandemic, the Ein channel and a corresponding YouTube channel have been redirected to provide live tutoring of all school level subjects and lessons daily from 8:30 AM to 12 PM on weekdays [46]. This great effort was conducted by 127 teachers in 112 subjects. Ein also provides a website through which students can practice lesson exercises and communicate with their teachers [46]. The vschool.sa portal is unique to Saudi Arabia and is a unified learning portal by the Ministry of Education that complements Ein. It provides synchronized web-based tutoring, assessment tools, learning material, and apps for smartphone access [47].

Telecommunication, Commercial, and Miscellaneous Digital Services

The major telecom companies in Saudi Arabia, namely the Saudi Telecom Company (STC), Mobily, and Zain Saudi Arabia, have announced free-of-charge data services to the most used educational platforms as well as health and telehealth applications to facilitate the smooth delivery of e-learning as well as health care delivery during the pandemic. The expected high usage of internet services, which exceeded the current capacity by around 33%, was also supported by the Saudi Communication and Information Technology Commission (CITC), which developed related infrastructure to accommodate the sudden high demand [36]. In an unprecedented move, prior to the COVID-19 pandemic, the CITC had also launched a guide to inform consumers about the trusted available mobile apps that are officially registered within the commission [48]. Internet providers also enabled users to access the Ministry of Health and governmental educational websites without consuming their personal data. When a call is placed on an STC or Mobily mobile number, a voice recording plays that reiterates Ministry of Health messages to help prevent the spread of COVID-19. Telecom companies changed their network names to display a message saying "Stay Home" (Figure 2). All these measures contributed to health education and the awareness drive.

Figure 2. Instagram advertisement for the new symptom checker feature on the Saudi Ministry of Heath's scheduling mobile app.



One of the most notable initiatives in Saudi Arabia is "Move to Tech." This initiative was launched by the Saudi Ministry of Communications and Information Technology on March 10, 2020 [49]. It facilitates the use of current digital tools and the creation of new ones in response to COVID-19. This has increased the use of digital tools in several sectors, but mainly in education, the food industry and health care. Following this initiative, a COVID-19 Hackathon [50] was launched to provide innovative remote and virtual solutions to combat the pandemic.

The G20 International Economic Leaders' Summit, scheduled in March 2020 and hosted by Saudi Arabia, was required to "go digital" in light of the COVID-19 pandemic. The summit hosted 19 countries, the European Union, and the Central Bank Governors. Saudi Arabia initiated the G20 Extraordinary Virtual Leaders' Summit on COVID-19 using videoconferencing that was inclusive of all international delegates. It is worth mentioning that the platform used was built and coordinated by the SDAIA (Figure 3) [51,52].

Figure 3. Snapshot of the G20 Extraordinary Virtual Leaders' Summit on COVID-19, hosted by the Saudi King Salman Bin Abdulaziz, on March 26, 2020.

In 2018, the Saudi Red Crescent Authority launched a mobile app, "Asefni," which translates to "Save Me." The aim of this app was to facilitate emergency service requests with accurate GPS locations [17]. When strict local curfew and travel restrictions were imposed on all Saudi citizens, the app was updated to provide movement permits during the curfew order for individuals who required essential medical consultations. This permit is issued for each case only after an initial web-based assessment [53]. Similarly, Saudi Public Security launched an online portal, "Tanaqul," to receive requests for domestic land travel permits between cities for people with extenuating circumstances [54].

Electronic commerce, on the other hand, was a prosperous industry in the Kingdom even before COVID-19. Major retailers had established web-based ordering and home delivery services throughout the major cities for everyday grocery items, home essentials, and furniture. The community-wide quarantine has highlighted the role of these web-based commercial services in aiding the mitigation process.

Risk Communication Directed at the Public Through Social Media

The year 2011 witnessed a boom of social media and user-generated content in Saudi Arabia [55]. The current social media scene in Saudi Arabia with regard to the percentage of internet users shows that the most preferred and used social media platforms are YouTube, WhatsApp, Facebook, Instagram, and Twitter [11]. In April 2011, the Saudi Ministry of Health joined Twitter and successfully built its audience's trust over the years until it reached nearly 3 million followers in early 2020 [56,57]. Before the first confirmed case of COVID-19 in Saudi Arabia, the MOH used its website and social media platforms, including Twitter, Facebook, YouTube, Snapchat, Instagram, and TikTok, to distribute health education materials. Different formats were used, such as WhatsApp stickers for proper hygiene. The topics included what COVID-19 is, how it is transmitted, how to prevent getting it, and where it originated. As the pandemic progressed and new scenarios emerged, the literature was modified and expanded to accommodate these changes. It was also translated into other languages, including but not limited to English, Portuguese, French, Russian, Tagalog, Spanish, and Urdu, ensuring a wider spread of relevant information [58]. The MOH and other ministries also used SMS text messages in both English and Arabic to raise awareness and emphasize the practice of precautions. Regular messages were sent to all citizens in different languages.

After the first case of COVID-19 in Saudi Arabia was confirmed on March 3, 2020, the Twitter account of the official spokesperson of the MOH was activated to directly and quickly announce and respond to COVID-19 news [59]. Rumors and misinformation are an expected and organic part of risk communication. According to the WHO, the best practices to address rumors and misinformation in risk and crisis communication include prevention, monitoring, and strategies for approaching a rumor when it occurs [60]. The spokesperson of the MOH incorporated this WHO strategy of dissolving original rumors. On Twitter, he would retweet the rumor with a comment to directly spread the correct information [59]. Figure 4 shows a comment by the MOH spokesperson on a widely distributed tweet by a person who was concerned about a colleague at his school who had just come back from Iran and had shown symptoms of respiratory illness [61]. The spokesperson thanked him for his concern and reassured him that the MOH had reached out to the person of concern and tested him for the virus, and he was found to be negative. He concluded by asking all people to communicate similar concerns to the MOH call center.

Figure 4. Twitter reply by the Saudi Ministry of Health spokesperson to a tweet by a person who was concerned about a colleague who showed symptoms of respiratory illness after returning from Iran [62].

The MOH has also collaborated with other government and nongovernment health entities to establish the Prevention Ambassador Initiative. The initiative is a web-based course for the layperson that provides certification in baseline information on COVID-19 to help prevent and control the COVID-19 infodemic [63]. Another effort to contain internet rumors and misinformation is a Saudi Public Prosecution release stating that intentional spread of rumors about COVID-19 or sharing material that causes panic among the public is an electronic crime that can be punished with up to 5 years of imprisonment or a fine of SR 3 million (US $799,888.20) [64].

When curfews were established in a number of major cities, the Center for Government Communication launched a national social media campaign titled *Kollona Masool*, meaning "We Are All Responsible" [65]. The main message of the campaign was that people should stay at home as a patriotic duty to their country and fellow residents. This hashtag was widely used by officials and the general population, and government entities on Twitter changed their cover images to read *Kollona Masool*. At this point, global brands were separating their logos and Twitter user names to emphasize social distancing. Many Saudi government entities and community influencers did the same. Once the lockdown was gradually lifted, this campaign shifted to *Naoodo bi Hathar*, which means "Return Carefully."

Early in the pandemic, the Saudi Health Council and the National Health Information Center started the first Arabic web-based interactive map dedicated to COVID-19 [66] (Figure 5). The map regularly updates travel alerts, confirmed cases, treated cases, deaths, percentage of treated cases, and percentage of deaths. The map is available as an app on the Apple Store [67]. It has an artificial intelligence (AI)–enabled chatbot, "Bashayar," that offers simple guidance in simple language. It also has a pop-up news headline with the latest official Arabic news on COVID-19 and a dropdown menu with further resources, including educational material, statistics, graphs, sources, data sets, isolation hospitals, statuses of major local and international conferences, and other studies [66].

Figure 5. The Arabic language Corona Map provided by the National Health Information Center, under the Saudi Health Council.

As more cases were discovered, the National Centre for Disease Prevention and Control (NCDC), also known as Weqaya, launched the COVID-19 hub website, followed by the Ministry of Health's COVID-19 awareness website [68]. Here, Arabic guidelines and essential information could be found under broad audience categories: community and public, professionals and health care workers, and daily updates [69]. A month into COVID-19 containment measures, the Ministry of Health released a public link to its live local Arabic dashboard, COVID-19 Dashboard: Saudi Arabia [69]. Accessibility to credible facts in a timely manner are core WHO principles of risk communication [60].

Discussion

Principal Findings

The ways in which the world has attempted to respond digitally to COVID-19 have surely raised concerns about how the world may be transformed post–COVID-19. On one hand, the need for digital response has been significantly highlighted; on the other hand, major challenges associated with its usage have surfaced [62]. We have already witnessed historical success stories in which countries used technology such as electronic databases and Google Maps to curb the spread of outbreaks.

Even in the current pandemic, several success stories have surfaced in which technology usage helped save lives when global markets shut down and strict curfews were enacted. The United Kingdom, for instance, initiated a COVID-19 symptom tracker app that allows users to enter their symptoms for risk identification, referral, and follow-up [70]. The use of mobile data has been suggested to help identify people at greater risk of travel-related infectious disease and in directing mass screening efforts accordingly. Google and Apple [71] introduced a decision support tool that acts as a location checker and has been implemented in the Tabaud app. Location data gathered from smartphones is used by public health officials to track patterns of movement of quarantined or home-isolated individuals.

A recent article published in *The Lancet* [72] highlighted the use of AI in curbing COVID-19. Taiwan used AI to improve its national health insurance database and integrate it with its immigration and customs database to create Big Data for analytics and crossmatching of individuals. This system generated alerts during clinical visits based on travel history and clinical symptoms to aid case identification. It also used QR-code scanning and web-based reporting of travel history and health symptoms to classify travelers' infectious risks based on their flight origins and travel histories for the past 14 days. Persons with low risk (no travel to Level 3 alert areas) were sent a health declaration border pass to their phones via SMS text message for faster immigration clearance; those with higher risk were quarantined at home and tracked through their mobile phones to ensure they remained at home during the incubation period [73].

The use of telehealth and chatbots in the United States and Singapore has shown promising results in enabling remote triaging of care and providing rapidly accessible information; these measures enable the provision of care to patients without requiring them to leave their homes [74,75].

Saudi Arabia's digital response to the COVID-19 pandemic is noteworthy. The aforementioned digital tools of public health and health care services are on par with those used worldwide. A few areas still require more exploration, such as the use of AI. It may be desirable to connect all the governmental and nongovernmental apps created during the COVID-19 pandemic to effectively activate interoperability across different technologies. This can lead to the creation of large, continuously updated data sets, which can be later used for diagnosis, management, and policy implementation.

However, we do recommend decreasing the number of public health mobile apps available for use during a future outbreak. This is to decrease the burden on the end user, avoid confusion, and ensure better adherence. As mentioned previously, there are five applications for COVID-19 symptoms and history screening, follow-up of cases, and contact tracing. Last, it should

be ensured that digital location identifiers activated via these applications do not breach privacy and agreed-upon permissions, as both Apple and Google have raised concerns regarding adherence to Health Insurance Portability and Accountability Act (HIPPA) regulations [76].

Wuhan was the first city to implement complete lockdown and initiate the policy of "Suspend Classes Without Stopping Learning." Lessons from this policy show the importance of having a strong web-based teaching infrastructure, the necessity of building the capacity of teachers, and finding solutions to bridge up the information gap that may occur because of distance teaching [77]. Fortunately, Saudi Arabia already possessed public and private e-learning infrastructure at the time of the COVID-19 pandemic. Saudi universities conducted webinars and training to rapidly increase their faculty's capacity for e-learning [78]. In one of the MOE's COVID-19 webinars, a group of education experts found that the COVID-19 experience proved successful in breaking educators' psychological barriers to use technology and distant learning methods. They also highlighted the future potential of e-learning to enhance web-based question banks and electronic resources, further engage faculty, adopt remote administrative meetings, decrease costs, and improve outcomes [79]. The Minister of Education hinted that distance learning would be made part of the Kingdom's regular education system [80], as it appears to be the new norm.

The main challenge posed by the COVID-19 pandemic has been the provision of efficient, accurate, and timely information to populations at risk worldwide [81]. The experience of COVID-19 risk communication by the Saudi Ministry of Health was perceived as very useful by 72% and very satisfactory by 74% of a survey population of 3133 Twitter users [82]. Through previous evaluation and improvement efforts, as well as experience with MERS-CoV, the risk communication infrastructure for this pandemic had already been built. Previous literature showed the types and sources of information that people in Saudi Arabia were seeking during the MERS-CoV outbreak. One study showed that 40% of people preferred the internet as a source of information [83].

In 2017, Saudi Arabia underwent a WHO Joint External Evaluation for international health risk assessment, including risk communication. This evaluation documented the use of the MOH web-based social listening tool to monitor rumors and adapt messaging according to the audience [84]. Similarly, Finland used social media messages and email to thematically categorize the community response to COVID-19 and develop recommendations for evidence-based risk communication [85]. Here, we urge the Ministry of Health to document its risk communication experience with COVID-19 for future reference, decision making, and simulation training.

Despite Saudi Arabia's widespread usage of various technical platforms during the current pandemic, this experience of learning and sharing seems to be ongoing. The community shift toward digital solutions will unravel further challenges and advantages as we continue to control and mitigate the epidemic curve. The implications and impact of this shift are yet to be known and studied. Whether the emerged digital dynamics and new norms among different sectors should be continued after the end of community quarantine remains unanswered.

Limitations

We attempted to encompass all digital solutions and tools used during the COVID-19 outbreak in Saudi Arabia up to the time of manuscript revision; however, shortcomings are expected. The COVID-19 pandemic is a rapidly changing scenario with weekly updates. This paper also lists apps but does not evaluate them or check for user experiences. Moreover, the criteria for inclusion in this paper were subjective. The authors attempted to decrease the effect of this subjectivity using discussion and consensus.

Conclusion

Saudi Arabia has been working to digitally transform many of its sectors since the launch of the national agenda, Vision 2030, in 2017 [32]. The COVID-19 pandemic has expedited this transformation. It has tested the reliability of the country's digital infrastructure and has highlighted questionable gaps for decision makers. This has been a nationwide trial of Saudi citizens' acceptance and ability to use and engage with the digitalization of these services and communications. At this point, it is too early to evaluate the unique Saudi experience of population-wide digital solutions. Future research should further explore and analyze the successes and pitfalls, hindrances, and challenges of this digital experience for specific sectors, including institutions, employees, and consumers.

Acknowledgments

The authors are grateful to the Deanship of Scientific Research, King Saud University, for funding through the Vice Deanship of Scientific Research Chairs. The authors would like to acknowledge Mrs Sadia Javaid for her help in editing the manuscript.

Conflicts of Interest

None declared.

References

1. Q&A on coronaviruses (COVID-19). World Health Organisation. 2020 Mar 09. URL: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses [accessed 2020-08-24]

2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
3. Hilton T. Countries with coronavirus curfews: Saudi Arabia joins growing list. Alarabia. 2020 Mar 23. URL: https://english.alarabiya.net/en/features/2020/03/23/Countries-with-coronavirus-curfews-Saudi-Arabia-joins-growing-list [accessed 2020-08-24]
4. Past pandemics. World Health Organization Regional Office for Europe. URL: http://www.euro.who.int/en/health-topics/communicable-diseases/influenza/pandemic-influenza/past-pandemics [accessed 2020-07-24]
5. Cleary V, Balasegaram S, McCloskey B, Keeling D, Turbitt D. Pandemic (H1N1) 2009: setting up a multi-agency regional response centre--a toolkit for other public health emergencies. J Bus Contin Emer Plan 2010 Mar;4(2):154-164. [Medline: 20494880]
6. Wilson C, Jumbert MG. The new informatics of pandemic response: humanitarian technology, efficiency, and the subtle retreat of national agency. Int J Humanitarian Action 2018 May 30;3(1):8. [doi: 10.1186/s41018-018-0036-5]
7. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
8. Sandhu R, Sood SK, Kaur G. An intelligent system for predicting and preventing MERS-CoV infection outbreak. J Supercomput 2016;72(8):3033-3056 [FREE Full text] [doi: 10.1007/s11227-015-1474-0] [Medline: 32214655]
9. Middle East respiratory syndrome coronavirus (MERS-CoV) 2019. World Health Organization. 2019. URL: https://www.who.int/en/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov) [accessed 2020-08-24]
10. Saeed AAB, Abedi GR, Alzahrani AG, Salameh I, Abdirizak F, Alhakeem R, et al. Surveillance and Testing for Middle East Respiratory Syndrome Coronavirus, Saudi Arabia, April 2015-February 2016. Emerg Infect Dis 2017 Apr;23(4):682-685 [FREE Full text] [doi: 10.3201/eid2304.161793] [Medline: 28322710]
11. Digital 2019 Saudi Arabia. 2019 Jan 30. URL: https://www.slideshare.net/DataReportal/digital-2019-saudi-arabia-january-2019-v01 [accessed 2020-08-25]
12. Germann S. World Health Day: Digital and data's role in the COVID-19 response. Foundation Botinar. 2020 Apr 07. URL: https://www.fondationbotnar.org/world-health-day-digital-and-datas-role-in-the-covid-19-response/ [accessed 2020-08-25]
13. Sehha App. Apple App Store. URL: https://apps.apple.com/sa/app/%D8%B5%D8%AD%D9%87/id1205814003 [accessed 2020-08-25]
14. Mawid E Services 2020. Saudi Ministry of Health. URL: https://www.moh.gov.sa/en/eServices/Pages/cassystem.aspx [accessed 2020-08-25]
15. Al-Rabiah: Doctors can Issue Prescriptions via Anaat App. Saudi Ministry of Health. 2020 Mar 26. URL: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-03-26-003.aspx [accessed 2020-08-25]
16. Wasfaty. URL: https://wasfaty.sa [accessed 2020-08-25]
17. National Digital Transformation Annual Report 2019. Digital Saudi. URL: https://ndu.gov.sa/report/ndu-annual-report-en.pdf [accessed 2020-08-25]
18. Cura App. Cura Healthcare. URL: https://company.cura.healthcare/en/ [accessed 2020-05-20]
19. Maya Clinic. URL: http://www.mayaclinic.com/ [accessed 2020-08-25]
20. Nala – Talk to a doctor. Webpage in Arabic. URL: https://nala.ai/ [accessed 2020-08-25]
21. Labyah counseling sessions. Webpage in Arabic. Labyah. 2020. URL: https://www.labayh.net/login [accessed 2020-08-25]
22. 80/20 Lifestyle App. Webpage in Arabic. URL: https://80-20app.com/ [accessed 2020-08-25]
23. Virtual Medical Academy. Webpage in Arabic. URL: https://medicalacademy.org/portal/ [accessed 2020-08-25]
24. Daem Service 2020. Saudi Comission for Health Speacialities. URL: https://www.scfhs.org.sa/en/eservices/trainees/Pages/TrainingGuidanceServiceDesc.aspx [accessed 2020-06-20]
25. Al-Rabiah Highlights Updates of Transformation Programs and Initiatives 2020. Saudi Ministry of Health. 2019 Sep 10. URL: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2019-09-10-003.aspx [accessed 2020-08-25]
26. (Tetamman)Clinics at Healthcare Centers and Hospitals. Saudi Ministry of Health. 2020. URL: https://www.moh.gov.sa/en/HealthAwareness/Tataman-Clinics/Pages/default.aspx [accessed 2020-08-25]
27. Public Health: Expanding Testing. Saudi Ministry of Health. URL: https://www.moh.gov.sa/en/HealthAwareness/EducationalContent/PublicHealth/Pages/Expanded_Testing.aspx [accessed 2020-08-25]
28. Alotaibi B, Yezli S, Bin Saeed AAA, Turkestani A, Alawam A, Bieh K. Strengthening health security at the Hajj mass gatherings: characteristics of the infectious diseases surveillance systems operational during the 2015 Hajj. J Travel Med 2017 May 01;24(3):taw087 [FREE Full text] [doi: 10.1093/jtm/taw087] [Medline: 28355612]
29. MOH Launches New App (Tetamman)to Prevent COVID-19. Saudi Ministry of Health. 2020 Apr 11. URL: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-04-11-004.aspx [accessed 2020-05-20]
30. MOH: (Tetamman) Smart Bracelet a Must for Citizens Returning from Abroad. Saudi Ministry of Health. 2020 May 22. URL: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-05-22-002.aspx [accessed 2020-05-15]
31. Tabaud App. Saudi Data and Al Authority. URL: https://tabaud.sdaia.gov.sa/IndexEn [accessed 2020-08-25]
32. National Transformation Program Delivery Plan 2018-2020. Kingdom of Saudi Arabia Vision 2030. 2018. URL: https://vision2030.gov.sa/sites/default/files/attachments/NTP%20English%20Public%20Document_2810.pdf [accessed 2020-08-25]

33. "Android mobile use in the unit #Annaah_aharjh b #Ospy_ohd in #Almedinh__almnorh where these contribute technical monitoring of patients in #Jaihh_korona #Korona_aljdid". Tweet in Arabic. @MOH_Madinah. 2020 Mar 29. URL: https://twitter.com/moh_madinah/status/1244384199148175361 [accessed 2020-06-25]
34. Telemedicine Regulations in the Kingdom of Saudi Arabia. National Health Information Center (NHIC), part of the Saudi Health Council (SHC). URL: https://nhic.gov.sa/en/Initiatives/Documents/Saudi%20Arabia%20Telemedicine%20Policy.pdf [accessed 2020-08-25]
35. Statistical Yearbook 2018. Kingdom of Saudi Arabia General Authority for Statistics. URL: https://www.stats.gov.sa/en/46 [accessed 2020-04-14]
36. Aljaber A. E-learning policy in Saudi Arabia: Challenges and successes. Res Comp Int Educ 2018 May 17;13(1):176-194. [doi: 10.1177/1745499918764147]
37. National Center for e-Learning. Webpage in Arabic. URL: https://tinyurl.com/y2dwclu2 [accessed 2020-08-25]
38. Alkhalaf S, Drew S, AlGhamdi R, Alfarraj O. E-Learning System on Higher Education Institutions in KSA: Attitudes and Perceptions of Faculty Members. Procedia Soc Behav Sci 2012;47:1199-1205. [doi: 10.1016/j.sbspro.2012.06.800]
39. E-services. Saudi Ministry of Education. URL: https://www.moe.gov.sa/en/Pages/default.aspx [accessed 2020-08-25]
40. Gharman A. Universities define mechanism of final examination, student withdrawal is an option. Saudi Arabia News. 2020 Apr 04. URL: https://www.saudi24news.com/2020/04/universities-determine-the-mechanism-of-final-exams-student-withdrawal-is-available-saudi-arabia-news.html [accessed 2020-08-25]
41. UpToDate anywhere in Saudi Arabia. Saudi Ministry of Health. URL: https://www.scfhs.org.sa/MESPS/TrainingProgs/EduMatActiv/Documents/UTDA%20KSA%20Tip%20Sheet.pdf [accessed 2020-07-24]
42. Saudi Commission for Health Sciences eCore Courses. MDBriefcase. URL: http://www.mdbriefcase.net/SCFHS/Home.aspx [accessed 2020-08-25]
43. KFSH&RC webinar series. King Faisal Speacilist Hospital and Research Centre. URL: https://www.kfshrc.edu.sa/en/home/news/11097 [accessed 2020-07-24]
44. Rhythm Management During The COVID19 Pandemic Live Webinar. Saudi Heart Association. URL: https://saudi-heart.com/2020/04/12/rhythm-management-during-the-covid19-pandemic/ [accessed 2020-07-24]
45. Virtual School. Webpage in Arabic. Saudi Ministry of Education. URL: https://vschool.sa [accessed 2020-08-25]
46. Alyousef A. Offered by the Ministry of Education: multiple options in the virtual school. Al Jazirah. 2020. URL: http://www.al-jazirah.com/2020/20200315/ln28.htm [accessed 2020-08-26] [WebCite Cache ID http://www.al-jazirah.com/2020/20200315/ln28.htm]
47. Virtual Learning portal Vschool.sa for distant learning. Mohammad Shatta. URL: https://www.arab-box.com/vschool-sa/ [accessed 2020-06-25]
48. "Now you can find out the updated list of applications .. delivery through registered with the body of electronic platforms by visiting the following link .. #Klna_msaol #Khalik_balbat" Tweet in Arabic. @CITC_SA. 2020 Mar 28. URL: https://twitter.com/citc_sa/status/1243876903301394433?s=21 [accessed 2020-06-25]
49. Muawyaih AlAnsari. "Move to Tech" hashtag launches a phase of digital transformation. Webpage in Arabic. AlRiyadh Newspaper. 2020 Mar 11. URL: http://www.alriyadh.com/1809797 [accessed 2020-08-27]
50. COVID-19 Saudi Hackathon. Webpage in Arabic. URL: https://covid19saudihackathon.com [accessed 2020-08-25]
51. "The 2nd #G20 Digital Economy Task Force meeting was held virtually under the #G20SaudiArabia Presidency on 2 April 2020, focused on a collaborative response to #COVID19 & the challenges associated with the deployment of Emerging Technologies.". @g20org. 2020 Feb 04. URL: https://twitter.com/g20org/status/1245742567221858305?lang=en [accessed 2020-08-25]
52. Opening Remarks for the Custodian of the Two Holy Mosques during the Extraordinary G20 Leaders' Summit. g20.org. URL: https://g20.org/en/media/Documents/Virtual_Leaders_Summit_King_Salman_Opening_Remarks_EN.pdf [accessed 2020-08-25]
53. Saudi Red Crescent launches "Asefni" app to request emergency service. Arab News. 2018 May 21. URL: https://www.arabnews.com/node/1306831/saudi-arabia [accessed 2020-08-25]
54. Travelling Between Cities—Online Permission Portal—Saudi Arabia. Times of KSA. 2020 Apr 04. URL: https://timesofksa.com/travelling-between-cities-online-permission-portal-saudi-arabia [accessed 2020-08-25]
55. Radcliffe DB, Bruni P. State of Social Media, Middle East: 2018. University of Oregon Scholars' Bank 2019 Jan:2018. [doi: 10.7264/FT8N-W076]
56. Saudi Ministry of Health. URL: https://www.moh.gov.sa/en/Pages/Default.aspx [accessed 2020-08-25]
57. Saudi Ministry of Health Twitter Account. Twitter. URL: https://twitter.com/SaudiMOH [accessed 2020-06-25]
58. MOH Issues COVID-19 Awareness Guidelines in Many Languages. Saudi Ministry of Health. 2020 Mar 17. URL: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-03-17-001.aspx [accessed 2020-08-25]
59. The Ministry of Health denies WhatsApp's current message regarding Corona … and explains: No. Saudi News 24. 2020 Apr 11. URL: https://www.saudi24news.com/2020/04/the-ministry-of-health-denies-whatsapps-current-message-regarding-corona-and-explains-no.html [accessed 2020-08-25]

60. Emergency risk communication (ERC) 5-step capacity-building package. World Health Organization Europe. 2017. URL: https://www.euro.who.int/en/health-topics/health-emergencies/pages/whos-work-in-emergencies/risk-communication-in-emergencies/national-health-emergency-risk-communication-training-package [accessed 2020-08-25]
61. @spokesman_moh. 2020 Mar 03. URL: https://twitter.com/spokesman_moh/status/1234768772155744256 [accessed 2020-08-25]
62. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
63. Prevention Ambassador Initiative 2020. Activity AoED. URL: http://alweqayah.sa/home [accessed 2020-08-25]
64. AlHkamasani M. Three million riyals fine to promote rumors of corona. Riyadh Daily. 2020 Mar 04. URL: http://www.alriyadhdaily.com/article/65c66e4c5b9240769a0d74be72cc177e [accessed 2020-03-04]
65. Alhadhira AA. "We Are All Responsible" is a Saudi swarm intelligence leadership in the face of COVID-19. SJEMed 2020:37-44. [doi: 10.24911/SJEMed/72-1589573868]
66. Corona interactive map Kingdom Of Saudi Arabia. Saudi Ministry of Health. 2020. URL: https://coronamap.sa [accessed 2020-08-25]
67. Corona Map App. Apple App Store. URL: https://apps.apple.com/app/corona-map/id1503046302 [accessed 2020-08-25]
68. COVID-19 Awareness. Saudi Ministry of Health. URL: https://covid19awareness.sa/en/home-page [accessed 2020-06-25]
69. Information About Coronavirus Disease. Waqaya Saudi Centre for Disease Prevention and Control. URL: https://covid19.cdc.gov.sa [accessed 2020-08-25]
70. COVID-19 Symptom Study app. ZOE. URL: https://covid.joinzoe.com/ [accessed 2020-08-25]
71. COVID 19 Screening Tool. Apple App Store. URL: https://www.apple.com/covid19 [accessed 2020-08-19]
72. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Lancet Digit Health 2020 Apr;2(4):e166-e167. [doi: 10.1016/s2589-7500(20)30054-6]
73. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Mar 03;323(14):1341-1342 [FREE Full text] [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
74. Priya S. Singapore government launches COVID-19 chatbot. Singapore government launches COVID-19 chatbot, 2020. 2020 Feb 24. URL: https://www.opengovasia.com/singapore-government-launches-covid-19-chatbot/ [accessed 2020-08-25]
75. Rosenbaum E. Robotic medicine may be the weapon the world needs to combat the coronavirus. CNBC. 2020 Feb 26. URL: https://www.cnbc.com/2020/02/26/robotic-medicine-may-be-the-weapon-needed-to-combat-the-coronavirus.html [accessed 2020-08-25]
76. NAAG Urges Apple and Google to Take Further Steps to Protect Privacy of Users of COVID-19 Contact Tracing Apps. HIPPA Journal. 2020 Jun 18. URL: https://www.hipaajournal.com/national-association-of-attorneys-general-urges-apple-and-google-exercise-control-over-covid-19-contact-tracing-apps/ [accessed 2020-06-18]
77. Zhang W, Wang Y, Yang L, Wang C. Suspending Classes Without Stopping Learning: China's Education Emergency Management Policy in the COVID-19 Outbreak. JRFM 2020 Mar 13;13(3):55 [FREE Full text] [doi: 10.3390/jrfm13030055]
78. E-learning in Saudi Universities continues the education journey for students. Webpage in Arabic. Saudi Ministry of Education. URL: https://www.moe.gov.sa/ar/news/Pages/un-c-2020-75.aspx [accessed 2020-06-20]
79. E-learning experts praise the success of the Ministry of Education and Saudi universities in transitioning to distance learning. Saudi Ministry of Education. URL: https://www.moe.gov.sa/ar/news/Pages/el-2020-u43.aspx [accessed 2020-06-25]
80. Naar I. Distance learning could be strategic choice post-coronavirus crisis: Saudi minister. Al Arabiya. 2020 Apr 20. URL: https://english.alarabiya.net/en/coronavirus/2020/04/20/Distance-learning-could-be-strategic-choice-post-coronavirus-crisis-Saudi-minister [accessed 2020-08-25]
81. Mahmood SH, Hasan K, Colder Carras M, Labrique A. Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
82. Al Shalhoob A. A study report on: communication best practice in managing the COVID-19 crisis and building health awareness among individuals in the Saudi Community A case study of the Saudi ministry of health effort. Riyadh, Saudi Arabia: King Saud University; 2020. URL: https://barq-org.sa/554588 [accessed 2020-08-27]
83. Hoda J. Identification of information types and sources by the public for promoting awareness of Middle East respiratory syndrome coronavirus in Saudi Arabia. Health Educ Res 2016 Feb 26;31(1):12-23 [FREE Full text] [doi: 10.1093/her/cyv061] [Medline: 26612051]
84. Joint External Evaluation of IHR Core Capacities of the Kingdom of Saudi Arabia. World Health Organization. 2017. URL: https://www.who.int/ihr/publications/WHO-WHE-CPI-2017.25.report/en/ [accessed 2020-08-25]
85. Lohiniva A, Sane J, Sibenberg K, Puumalainen T, Salminen M. Understanding coronavirus disease (COVID-19) risk perceptions among the public to enhance risk communication efforts: a practical approach for outbreaks, Finland, February 2020. Eurosurveillance 2020 Apr 02;25(13):online. [doi: 10.2807/1560-7917.es.2020.25.13.2000317]

Abbreviations

AI: artificial intelligence
CITC: Communication and Information Technology Commission
eHealth: electronic health
e-learning: electronic learning
e-prescription: electronic prescription
HESN: Health Electronic Surveillance Network
HIPPA: Health Insurance Portability and Accountability Act
MERS: Middle East respiratory syndrome
MERS-CoV: Middle East respiratory syndrome coronavirus
MOE: Ministry of Education
MOH: Ministry of Health
SCFHS: Saudi Commission for Health Specialties
SDAIA: Saudi Data and Artificial Intelligence Authority
STC: Saudi Telecom Company
WHO: World Health Organization

Edited by G Eysenbach; submitted 15.04.20; peer-reviewed by A Bhagavathula, D Gunasekeran, A Benis; comments to author 11.06.20; revised version received 25.07.20; accepted 11.08.20; published 01.09.20.

Please cite as:
Hassounah M, Raheel H, Alhefzi M
Digital Response During the COVID-19 Pandemic in Saudi Arabia
J Med Internet Res 2020;22(9):e19338
URL: https://www.jmir.org/2020/9/e19338
doi:10.2196/19338
PMID:32790642

Original Paper

Efficacy of Hydroxychloroquine and Tocilizumab in Patients With COVID-19: Single-Center Retrospective Chart Review

Sohaib Roomi[1], MD; Waqas Ullah[1], MD; Faizan Ahmed[1], MD; Soban Farooq[2], MD; Usama Sadiq[1], MD; Asad Chohan[3], MD; Munnam Jafar[1], MD; Maryum Saddique[4], MD; Shristi Khanal[1], MD; Robert Watson[1], MD; Margot Boigon[1], MD

[1]Jefferson Health Abington, Abington, PA, United States

[2]King Edward Medical University, Lahore, Pakistan

[3]Newark Beth Israel Medical Center, Newark, NJ, United States

[4]Einstein Medical Center, Philadelphia, PA, United States

Corresponding Author:
Sohaib Roomi, MD
Jefferson Health Abington
1200 Old York Rd
Abington, PA
United States
Phone: 1 2674666985
Email: roomigujjar251@gmail.com

Abstract

Background: During the initial phases of the COVID-19 pandemic, there was an unfounded fervor surrounding the use of hydroxychloroquine (HCQ) and tocilizumab (TCZ); however, evidence on their efficacy and safety have been controversial.

Objective: The purpose of this study is to evaluate the overall clinical effectiveness of HCQ and TCZ in patients with COVID-19. We hypothesize that HCQ and TCZ use in these patients will be associated with a reduction in in-hospital mortality, upgrade to intensive medical care, invasive mechanical ventilation, or acute renal failure needing dialysis.

Methods: A retrospective cohort study was performed to determine the impact of HCQ and TCZ use on hard clinical outcomes during hospitalization. A total of 176 hospitalized patients with a confirmed COVID-19 diagnosis was included. Patients were divided into two comparison groups: (1) HCQ (n=144) vs no-HCQ (n=32) and (2) TCZ (n=32) vs no-TCZ (n=144). The mean age, baseline comorbidities, and other medications used during hospitalization were uniformly distributed among all the groups. Independent *t* tests and multivariate logistic regression analysis were performed to calculate mean differences and adjusted odds ratios with 95% CIs, respectively.

Results: The unadjusted odds ratio for patients upgraded to a higher level of care (ie, intensive care unit) (OR 2.6, 95% CI 1.19-5.69; *P*=.003) and reductions in C-reactive protein (CRP) level on day 7 of hospitalization (21% vs 56%, OR 0.21, 95% CI 0.08-0.55; *P*=.002) were significantly higher in the TCZ group compared to the control group. There was no significant difference in the odds of in-hospital mortality, upgrade to intensive medical care, need for invasive mechanical ventilation, acute kidney failure necessitating dialysis, or discharge from the hospital after recovery in both the HCQ and TCZ groups compared to their respective control groups. Adjusted odds ratios controlled for baseline comorbidities and medications closely followed the unadjusted estimates.

Conclusions: In this cohort of patients with COVID-19, neither HCQ nor TCZ offered a significant reduction in in-hospital mortality, upgrade to intensive medical care, invasive mechanical ventilation, or acute renal failure needing dialysis. These results are similar to the recently published preliminary results of the HCQ arm of the Recovery trial, which showed no clinical benefit from the use of HCQ in hospitalized patients with COVID-19 (the TCZ arm is ongoing). Double-blinded randomized controlled trials are needed to further evaluate the impact of these drugs in larger patient samples so that data-driven guidelines can be deduced to combat this global pandemic.

(J Med Internet Res 2020;22(9):e21758) doi:10.2196/21758

KEYWORDS

COVID-19; hydroxychloroquine; tocilizumab

Introduction

As of July 23, 2020, more than 4 million cases and 140,000 deaths from COVID-19 have been reported in the United States. There is currently no proven medical therapy for this disease except low-dose dexamethasone and remdesivir based on preliminary evidence with the mainstay of treatment being supportive care [1]. Multiple off-label and compassionate use therapies are currently being employed, targeting currently known pathophysiological mechanisms of this novel virus. Increasing social and economic devastation caused by COVID-19 has led the Federal Drug Administration (FDA) to issue emergency use authorizations (EUAs) for various drugs without proven benefits [2]. Although many of these drugs have revealed promising in vitro activity against the coronaviridae family, including SARS-CoV-2, the translation of these in vitro effects into clinical efficacy is a matter of debate. While, as physicians, we tend to assume that these drugs will do more good than harm when utilizing them as a last resort to severely ill patients, the fact remains that in the absence of randomized controlled trials, there is no way to reliably judge the impact of these medications. Fortunately, this situation is being remedied, with evidence emerging, initially from China, and more recently from trials in the United States and Europe.

Among others, hydroxychloroquine (HCQ) and the interleukin-6 (IL-6) inhibitor, tocilizumab (TCZ), became popular options to treat COVID-19. There is no concrete evidence supporting their use, and they were widely adopted across the world based on anecdotal data. Our hospital, following the guidelines of its parent enterprise, permitted the use of HCQ in COVID-19 patients who had respiratory insufficiency as indicated by low oxygen saturation. Similarly, TCZ was used for patients who met the criteria for cytokine release syndrome during the time frame of this study. The purpose of this study is to evaluate the overall clinical effectiveness of HCQ and TCZ in our hospital. We will compare our results to the preliminary results of the HCQ arm of the RECOVERY trial, which did not reveal a difference in 28-day mortality between the HCQ group and the usual care group [3]. In addition to mortality, we will evaluate other secondary endpoints and hypothesize that use of HCQ and TCZ will be associated with a reduction in the endpoints of an upgrade to the intensive care unit (ICU), need for invasive mechanical ventilation (IMV), acute renal failure necessitating dialysis, and reduction in D-dimer and C-reactive protein (CRP) on the 7th day of hospitalization.

Methods

Study Design and Participants

This retrospective cohort study included adult inpatients (≥18 years old) from Abington Hospital - Jefferson Health in the United States. All patients had a confirmed diagnosis of COVID-19 between March 1, 2020, and May 30, 2020. The study was approved by the Institutional Review Board, and the requirement for informed consent was waived by the Research Ethics Committee.

Data Collection

All COVID-19 patients who were admitted to the hospital between March 1, 2020, and May 30, 2020, were included. Data were extracted from electronic medical records (Sunrise) using a standardized data collection form. All authors contributed to data retrieval and an independent author adjudicated any difference in interpretation between the data extractors. SARS-CoV-2 was detected in respiratory specimens (nasopharyngeal or throat swabs) by real-time qualitative polymerase chain reaction (RT-qPCR). Routine blood work included complete blood count, serum electrolytes, renal function test, coagulation profile, serum ferritin, CRP, D-dimer level, lactate dehydrogenase, and myocardial enzymes (troponin T) on presentation to the hospital and on day 7 of hospitalization. Baseline comorbidities, including hypertension (HTN), diabetes mellitus (DM), chronic kidney disease (CKD), chronic obstructive lung disease (COPD), and coronary artery disease (CAD), were also recorded. The criteria of discharge from the hospital after recovery included resolution of fever, absence of symptoms for at least 1 day, and substantial clinical or radiological improvement.

Statistical Analysis

A chi-square (χ^2) test was used for comparison of categorical data and Fisher exact test was adopted if the expected count in more than 20% cells was less than 5. Continuous variables were presented as means and standard deviations while categorical variables were reported in percentages and proportions. To quantify the association between the dichotomous categorical variables, an unadjusted odds ratio (uOR) was obtained using a Cochran-Mantel-Haenszel method. To explore the risk factors and gauge the impact of potential effect modifiers (covariates) on our endpoints (in-hospital mortality, ICU upgrade, IMV, dialysis, and inflammatory marker level), binomial and multinomial logistic regression models were applied. The differences in the baseline comorbidities (DM, HTN, CAD, CKD, and COPD) and medication use (HCQ, TCZ, remdesivir, therapeutic anticoagulation, and steroids) were accounted for to obtain an adjusted odds ratio (aOR) for all outcomes. For normally and abnormally distributed continuous data, an independent sample *t* test and Mann-Whitney *U* test were used, respectively. A one-way analysis of variance (ANOVA) was used to compare differences in the mean of continuous variables for multiple in-hospital complications. A two-sided $\alpha<.05$ was considered statistically significant with corroborating inference from a 95% CI. Statistical analyses were performed using the SPSS software (version 25, IBM Corp).

Results

Demographics and Baseline Characteristics

Our study population consisted of 176 patients who were hospitalized and had a confirmed case of COVID-19 infection. All patients were divided into two comparison groups: (1) HCQ (n=144) vs no-HCQ (n=32) and (2) TCZ (n=32) vs no-TCZ (n=144), respectively. Table 1 and Figure 1 depict the underlying comorbidities and other medications used during hospitalization in both comparison groups. The mean age in years for the HCQ

and no-HCQ groups was 63.75 and 65.87 years, respectively (*P*=.55); for the TCZ and no-TCZ groups, it was 58.09 and 65.48 years, respectively (*P*=2.75). The most common underlying comorbidities in all the four groups were DM, HTN, CAD, CKD, and COPD. Common medications used during hospitalization included steroids, anticoagulants, HCQ, and TCZ. These underlying comorbidities and medications used during hospitalization were nonsignificantly different between the comparison groups (*P*≥.05). The detailed percentages of group-wise comorbidities and demographics are given in Table 1.

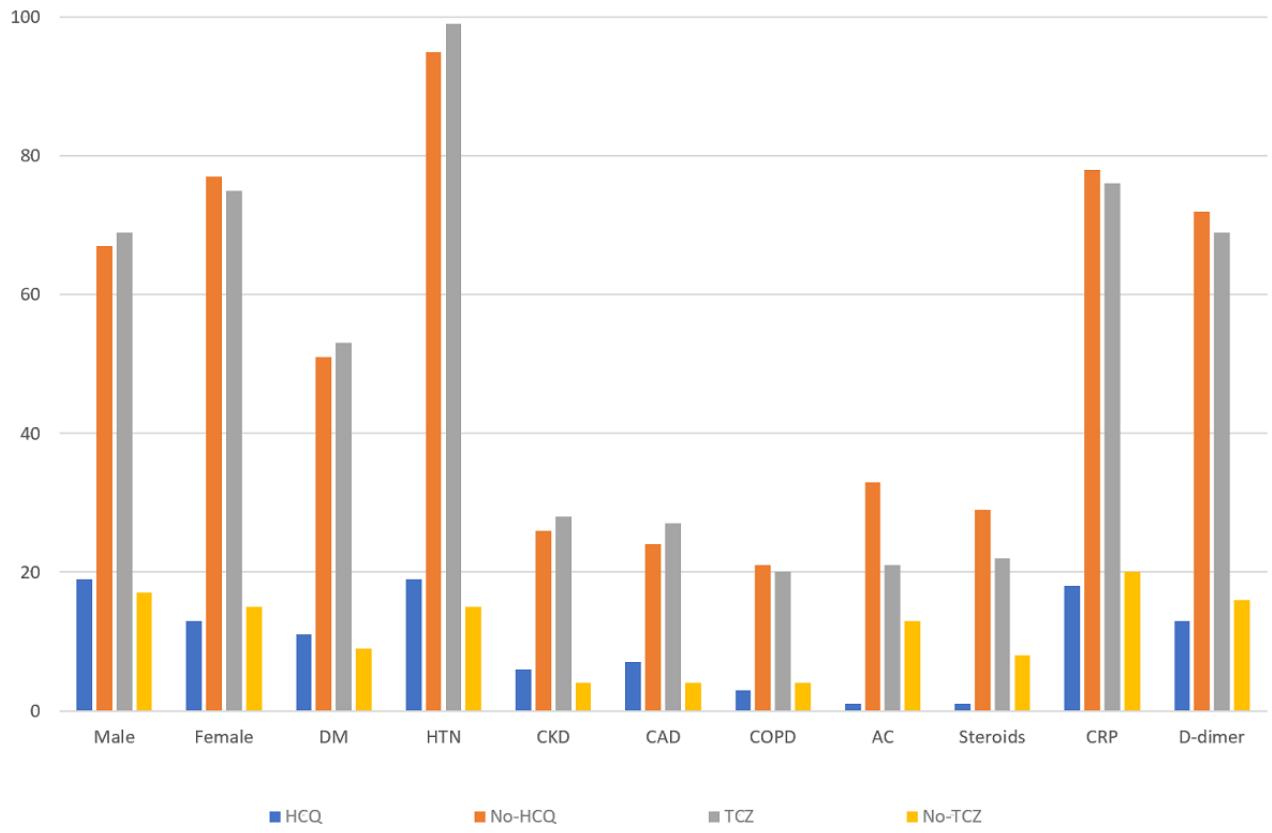
Table 1. Baseline characteristics of the included population across comparison groups.

Characteristic	No-HCQ[a]	HCQ	*P* value	No-TCZ[b]	TCZ	*P* value
Age (years), mean	65.87	63.75	.55	58.09	65.48	2.75
Sex			.17			.44
Male, n (%)	66 (80.50)	16 (19.50)		23 (27.70)	60 (72.30)	
Female, n (%)	75 (88.20)	10 (11.80)		13 (14.90)	74 (85.10)	
Diabetes mellitus			.72			.97
No, n (%)	92 (85.20)	16 (14.80)		23 (21.10)	86 (78.90)	
Yes, n (%)	49 (83.10)	10 (16.90)		13 (21.30)	48 (78.70)	
Hypertension			.24			.61
No, n (%)	48 (80.00)	12 (20.00)		14 (23.30)	46 (76.70)	
Yes, n (%)	93 (86.90)	14 (13.10)		22 (20.00)	88 (80.00)	
Coronary artery disease			.16			.78
No, n (%)	114 (82.60)	24 (17.40)		30 (21.60)	109 (78.40)	
Yes, n (%)	27 (93.10)	2 (6.90)		6 (19.40)	25 (80.60)	
Chronic kidney disease			.65			.56
No, n (%)	114 (83.80)	22 (16.20)		28 (20.30)	110 (79.70)	
Yes, n (%)	27 (87.10)	4 (12.90)		8 (25.00)	24 (75.00)	
Chronic obstructive pulmonary disease			.29			.56
No, n (%)	119 (83.20)	24 (16.80)		32 (21.90)	114 (78.10)	
Yes, n (%)	22 (91.70)	2 (8.30)		4 (16.70)	20 (83.30)	
Steroids			.39			.25
No, n (%)	115 (83.30)	23 (16.70)		32 (22.90)	108 (77.10)	
Yes, n (%)	26 (89.70)	3 (10.30)		4 (13.30)	26 (86.70)	
Anticoagulation			.32			.08
No, n (%)	115 (85.80)	19 (14.20)		25 (18.40)	111 (81.60)	
Yes, n (%)	26 (78.80)	7 (21.20)		11 (32.40)	23 (67.60)	

[a]HCQ: hydroxychloroquine.

[b]TCZ: tocilizumab.

Figure 1. Baseline comorbidities and medication use in the hydroxychloroquine (HCQ) group, tocilizumab (TCZ) group, and control groups. The x-axis represents sex, comorbidities, medications, C-reactive protein (CRP), and D-dimer level at presentation; the y-axis represents the percentage of subjects. DM: diabetes mellitus; HTN: hypertension; CKD: chronic kidney disease; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; AC: anticoagulation.



Odds Ratios of Outcomes

Table 2 and Figures 2 and 3 compare the above-mentioned outcomes between the TCZ group and the no-TCZ group. The unadjusted odds ratio for patients requiring an upgrade to the ICU was significantly higher in those who received TCZ compared to the control group (OR 2.6, 95% CI 1.19-5.69; *P*=.003). Similarly, patients who received TCZ had a significant reduction in CRP levels on day 7 of hospitalization compared to the control group (21% vs 56%, OR 0.21, 95% CI 0.08-0.55; *P*=.002). However, this reduction in the inflammatory markers did not translate into clinical benefits. There was no significant difference in the unadjusted odds of in-hospital mortality, IMV, acute renal failure necessitating dialysis, and discharge from the hospital after recovery between the two groups. The proportion of high D-dimer levels (>500 ng/dL) and elevated CRP (>100 ng/dL) on day 7 of hospitalization were also identical between the TCZ and no-TCZ groups. However, when we adjusted the observed odds ratios for baseline comorbidities, including DM, HTN, CKD, CAD, COPD, medications, use of anticoagulation at home, therapeutic anticoagulation during hospital stay, as well as steroid and HCQ use in the TCZ comparison group, the adjusted odds values were consistent with unadjusted odds ratios for all the outcomes. The exception was for an upgrade to medial ICU, where there was no difference between the TCZ group and the no-TCZ group. This is contrary to the unadjusted odds, which revealed more ICU upgrades in the TCZ group compared to the no-TCZ group (Table 2). The forest plots given in Figures 2 and 3 reveal the difference in unadjusted and adjusted odds between the TCZ group and the no-TCZ group.

Table 2. Tocilizumab (TCZ) regression analysis with outcome.

Outcome	TCZ, n	No TCZ, n	uOR[a] (95% CI)	*P* value	aOR[b] (95% CI)	*P* value
Invasive mechanical ventilation	47	31	1.94 (0.89-4.23)	.14	1.2 (0.49-2.9)	.67
Upgrade	50	28	2.6 (1.19-5.69)	.03	1.9 (0.80-4.5)	.14
Dialysis	6	6	1.13 (0.23-5.6)	.79	1.3 (0.21-8.3)	.76
Mortality	6	13	0.44 (0.97-1.99)	.43	0.28 (0.05-1.4)	.13
Discharge	25	38	0.54 (0.23-1.3)	.23	0.78 (0.28-2.1)	.64
D-dimer						
Day 1	50	52	0.93 (0.43-2.01	.99	0.7 (0.31-1.7)	.47
Day 7	77	64	1.92 (0.76-4.9)	.24	1.4 (0.52-4.2)	.45
C-reactive protein						
Day 1	63	55	1.38 (0.63-3.04)	.55	1.26 (0.54-2.93)	.59
Day 7	21	56	0.21 (0.08-0.55)	.002	0.17 (0.05-0.50)	.001

[a]uOR: unadjusted odds ratio.

[b]aOR: adjusted odds ratio.

Figure 2. Forest plot comparing unadjusted odds of outcomes between the tocilizumab (TCZ) and no-TCZ groups. CRP: C-reactive protein.

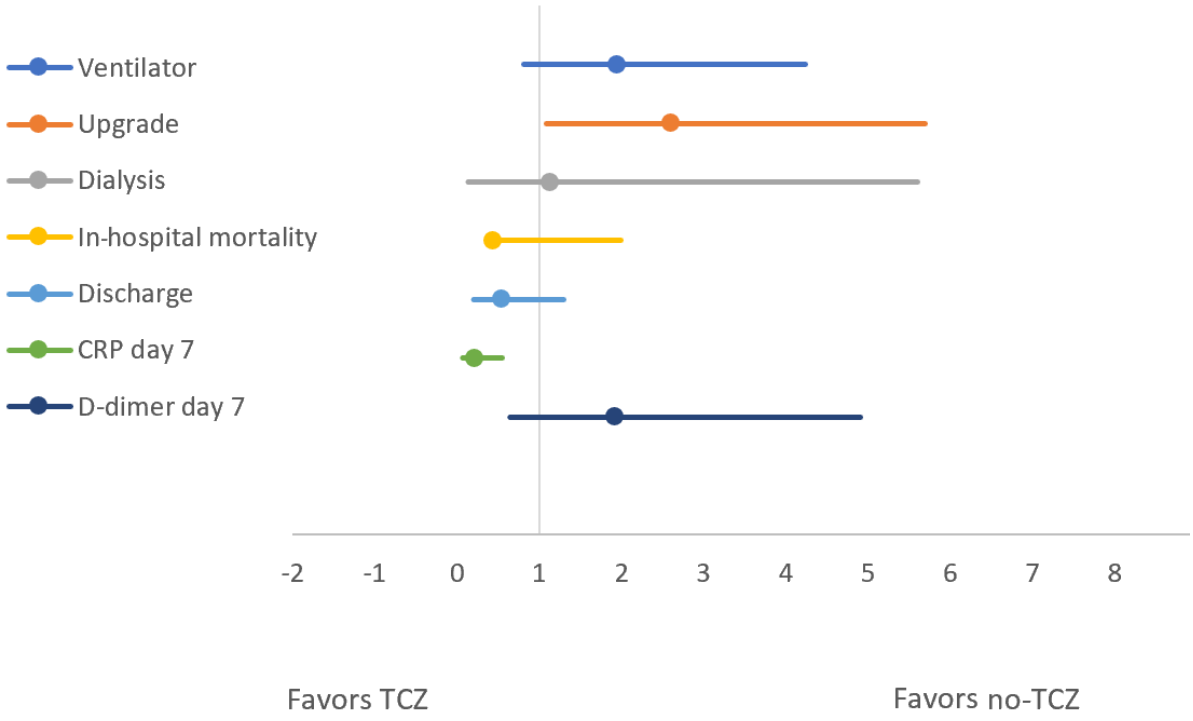
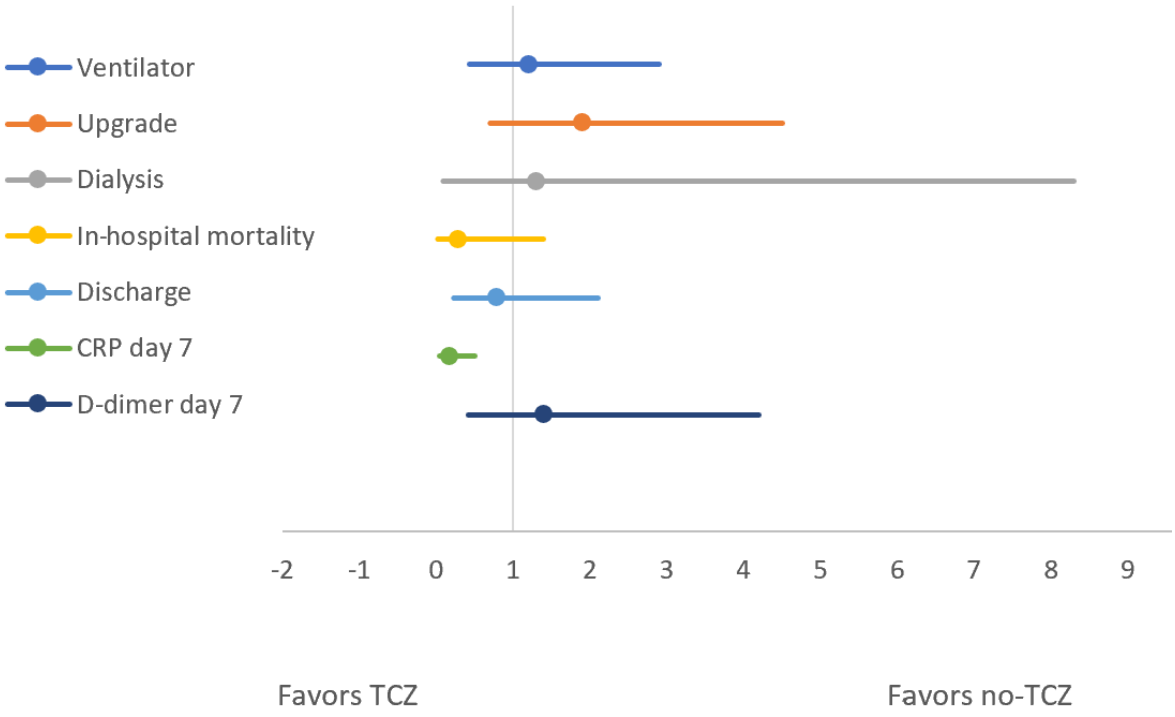


Figure 3. Forest plot comparing adjusted odds of outcomes between the tocilizumab (TCZ) and no-TCZ groups. CRP: C-reactive protein.



Similarly, Table 3 and Figures 4 and 5 compare the above-mentioned outcomes between the HCQ group and the no-HCQ group. The use of HCQ in patients with COVID-19 was not associated with a significant improvement in any of the outcomes. The unadjusted odds ratio of in-hospital mortality, upgrade to ICU, IMV, acute renal failure needing dialysis, or discharge after recovery were identical between patients receiving HCQ or not, respectively. Similarly, the proportion of high D-dimer and CRP levels on day 7 of hospitalization was not significantly different between the HCQ and no-HCQ groups. As with the TCZ comparison group, a multivariate regression analysis was used to adjust the observed odds ratios for baseline comorbidities and medications including TCZ in the HCQ comparison group. The adjusted odds values were consistent with unadjusted odds ratios for all the outcomes as having been depicted by similar forest plots in Figures 4 and 5.

Table 3. Hydroxychloroquine (HCQ) regression analysis with outcome.

Outcome	HCQ, n	No HCQ, n	uOR[a] (95% CI)	*P* value	aOR[b] (95% CI)	*P* value
Invasive mechanical ventilation	37	22	2.08 (0.84-5.14)	.16	1.2 (0.46-3.2)	.68
Upgrade	33	25	1.5 (0.63-3.59)	.48	0.9 (0.35-2.3)	.84
Dialysis	5	9	0.57 (0.12-2.02)	.57	0.34 (0.06-1.7)	.19
Mortality	13	6	2.28 (0.5-10.3)	.43	1.6 (0.33-7.9)	.54
Discharge	35	41	0.78 (0.36-1.7)	.68	1.15 (0.48-2.7)	.74
D-dimer						
Day 1	51	54	0.89 (0.37-2.10)	.95	0.8 (0.3-1.9)	.63
Day 7	69	33	4.47 (0.77-25)	.18	3.6 (0.59-22.7)	.16
C-reactive protein						
Day 1	55	62	0.75 (0.33-1.69)	.62	0.64 (0.27-1.51	.31
Day 7	49	29	2.42 (0.45-12.95	.50	2.0 (0.33-12.8)	.44

[a]uOR: unadjusted odds ratio.

[b]aOR: adjusted odds ratio.

Figure 4. Forest plot comparing unadjusted odds of outcomes between the hydroxychloroquine (HCQ) and no-HCQ groups. CRP: C-reactive protein.

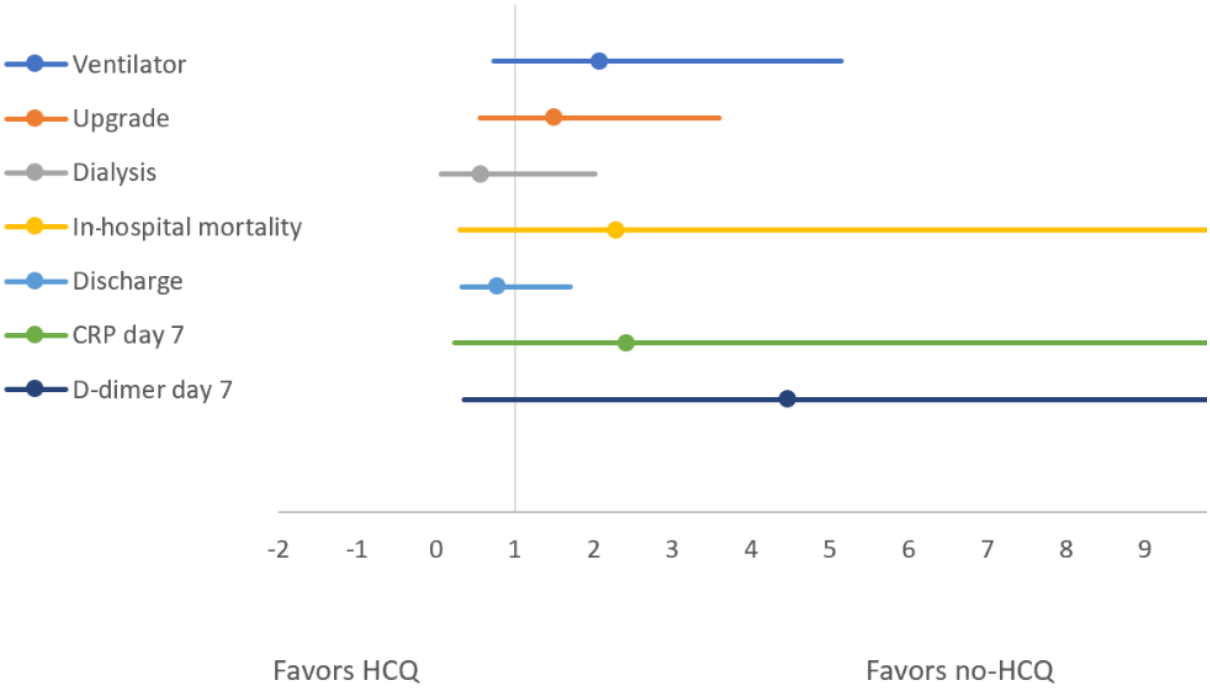
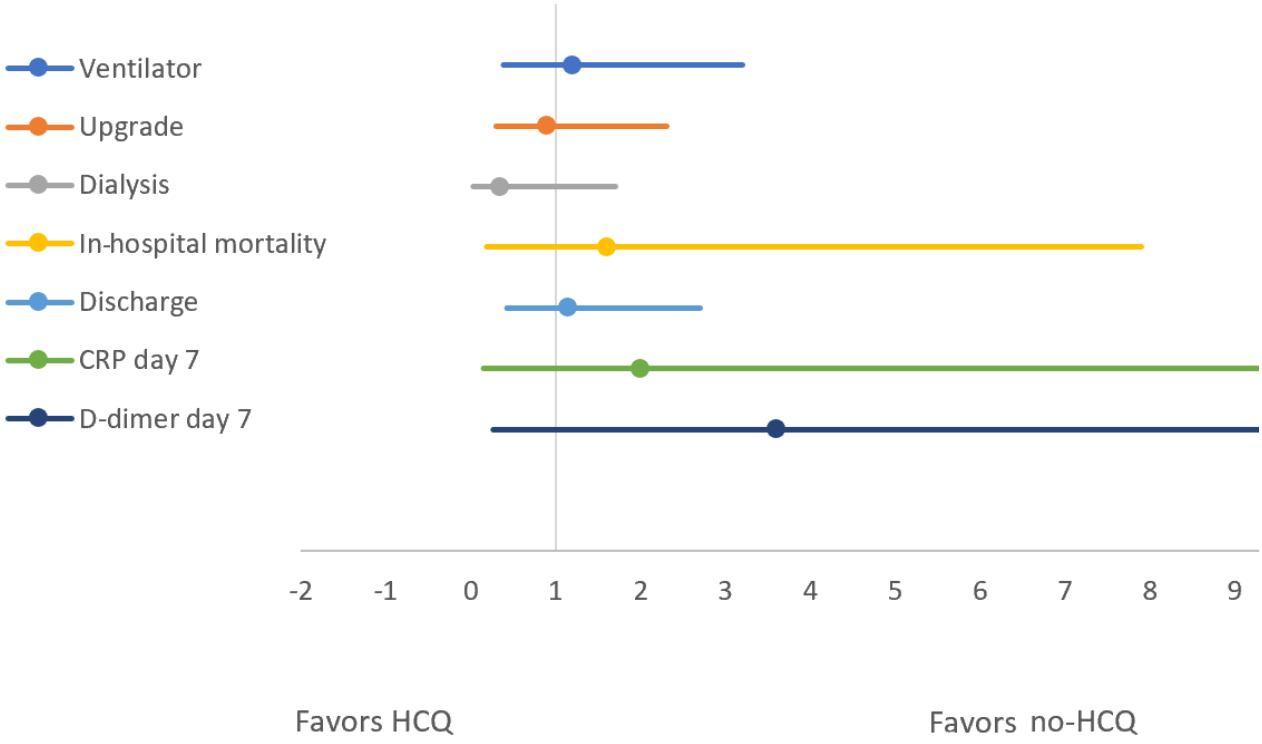


Figure 5. Forest plot comparing the adjusted odds of outcomes between hydroxychloroquine (HCQ) and no-HCQ groups. CRP: C-reactive protein.



Discussion

Principal Findings

The purpose of this study is to evaluate the overall clinical effectiveness of HCQ and TCZ in our hospital. Our results revealed that both TCZ and HCQ had no role in improving hard clinical outcomes in patients with COVID-19 admitted to the hospital. Compared to patients in the control group, those who received either of these medications did not show a significant reduction in the rate of in-hospital mortality, upgrade to ICU, IMV, reduction in acute renal failure to the point of needing dialysis, or discharge from the hospital after recovery. Although the patients who received TCZ appeared to have a higher rate

of ICU upgrade, this trend seemed to be driven by multiple comorbidities in the TCZ group, as evidenced by an identical adjusted odds ratio on multivariate analysis (Table 2; Figures 2 and 3). Partly contributing to this might be the higher use of TCZ in the sicker patients who fulfilled the criterion to receive the drug based on disease severity.

HCQ and TCZ, the major therapy for rheumatological diseases, have recently gained attention as one of the major cornerstone management approaches for COVID-19. HCQ is thought to work by inhibiting glycosylation of the host receptors, endosomal acidification, and proteolytic processing thereby, blocking viral entry into host cells [4-7]. TCZ, on the other hand is believed to counteract the misdirected immune response related to the COVID-19 cytokine storm [8]. Being a monoclonal antibody directed against IL-6, TCZ is thought to dampen the immune response and potentially reduce the adverse outcomes related to COVID-19.

A previous study by Xu et al [9] has shown a significant improvement in respiratory function (91% reduction in symptoms) and length of hospital stay in COVID-19 patients with a single dose of TCZ. However, that study was underpowered (n=21 patients) and had no control arm [9]. Similarly, Luo and colleagues [10] observed an 80% survival rate in patients receiving TCZ. Their study also was not followed up by a large-scale study and had several limitations. A recently published retrospective cohort study that included 544 patients admitted in different hospitals of Italy revealed that after adjustment for sex, age, recruiting center, duration of symptoms, and SOFA (sequential organ failure assessment) score, TCZ treatment was associated with a reduced risk of IMV or death (adjusted hazard ratio 0.61, 95% CI 0.40-0.92; *P*=.020) [11]. Our study consisted of 176 patients, and it demonstrated that there were no major clinical benefits to TCZ use in COVID-19 patients. A significant reduction of CRP levels on day 7 of hospitalization was observed in the TCZ group compared to the control group; yet this difference did not translate into clinical benefits in terms of a reduction in in-hospital mortality, medical ICU upgrade, or reduction in IMV (Figures 2 and 3). As mentioned in the study limitations below, a larger patient population and a randomized controlled design might have demonstrated a clinical benefit parallel to this reduction in CRP level. Among other ongoing randomized, double-blinded, controlled trials, the Oxford-based RECOVERY trial is also recruiting participants who meet the eligibility criteria into the TCZ arm of the trial [12].

Similarly, preliminary data from China reported that HCQ use was associated with a reduction in the viral load, duration of disease, and resolution of COVID-19 pneumonitis on imaging [13]. A small, nonrandomized, open-label French study consisting of 36 patients also reported significant reduction in the viral load in patients taking HCQ [14]. Major subsequent large-scale trials also reported its potential utility in reducing the need for IMV. However, the medical community was concerned regarding the potential cardiovascular adverse effects of off-label HCQ use. Despite all the controversies surrounding HCQ, its use prevailed in the earlier part of the pandemic, leading to stockpiling and shortage of HCQ in international markets, then followed by a swift decline in its use [15].

In our study, we systematically determined the impact of HCQ on the hard clinical outcomes in the COVID-19 cohort. Our mortality analysis showed a nonsignificant difference in the rate of in-hospital mortality in patients receiving HCQ group compared to those in the control group. It should be noted, however, that there was a two-fold higher risk of death in the HCQ arm. These findings are in line with a previous study by Magagnoli et al [16] that also reported a three times higher odds of death in patients receiving HCQ. Following this, another French study consisting of 181 patients with diagnosed COVID-19 pneumonitis reported that HCQ use was of no benefit [17]. In terms of mortality, the results of our study are similar to the recently published preliminary results of the RECOVERY trial in which patients were randomized between the HCQ group (n=1542) and usual care group (n=3132). There was no significant difference in 28-day mortality between the two groups (hazard ratio 1.11, 95 % CI 0.98-1.26; *P*=.10) [3]. Following these results, the FDA revoked the EUA for use of HCQ in COVID-19 patients on June 15, 2020 [18].

The most debilitating complication of SARS-CoV-2 infection is acute respiratory failure, necessitating the use of IMV and other concurrent resource-intensive tools in critical care units [15]. Previous studies have reported mixed results, showing 11% to 44% use of IMV in patients receiving HCQ and TCZ [13,19,20]. Magagnoli et al [17] included sicker patients, who were more likely to receive HCQ on compassionate grounds and hence were more prone to have adverse outcomes and death, calling into question its reliability. By contrast, our analysis adjusted the pooled estimate of IMV requirement in both HCQ and TCZ groups by identifying major potential confounders such as baseline comorbidities and other medications used during the hospital stay. By demonstrating a nonsignificant trend in all the above-mentioned outcomes, we recommend against the routine use of HCQ and TCZ in patients with COVID-19.

Limitations

The limitations of our study should be considered when interpreting the results. Due to the retrospective nonrandomized nature of the study, a causal relationship could not be established. Although the overall findings were adjusted for covariates including baseline comorbidities and medications, the impact of unmeasured confounders, such as initiation of several complementary therapies at the treating physician's discretion, could not be determined. Based on our clinical experience, the average duration of any therapy for COVID-19 was less than 7 days; therefore, we chose to use laboratory values from day 1 and day 7. However, given the variable frequency of laboratory specimen collection, it is not possible for us to ascertain if these values truly represented pre- and posttreatment values accurately in all cases. The patients who received TCZ were mainly selected based on the availability of the drug (which was in short supply intermittently during the time frame of our study), and these patients were sicker with lower PaO2/FiO2 ratios. Moreover, by excluding patients still in the hospital, the case fatality ratio in our study cannot reflect the true mortality of COVID-19. Our study did show a trend of beneficial events in terms of the point estimate of the pooled effect size. However, there was an overlap in the confidence

intervals and broad confidence intervals indicating that our study was underpowered to reach the level of significance. Although we adjusted the outcomes against demographics and underlying comorbidities, neither did we evaluate the contribution of underlying comorbidities to COVID-19 mortality via propensity score matching nor did our study evaluate the potentially harmful effects of these medications. We believe that a large-scale study will determine the true merits of these medications and will also reveal the potentially harmful outcomes of these medications. Many questions remain open, however. By adjusting the adult patients with the confirmed disease, we believe our population is the representative of the real-world cohort.

Conclusion

HCQ and TCZ use was not associated with a reduction in endpoints of in-hospital mortality, upgrade to medical ICU, need for IMV, acute renal failure necessitating dialysis, or discharge from the hospital. Although there was a significant reduction in CRP level on day 7 of hospitalization in the patients receiving TCZ, the lack of improvement in hard clinical outcomes suggests that large-scale randomized controlled trials are needed to evaluate the efficacy of these drugs.

Conflicts of Interest

None declared.

References

1. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. N Engl J Med 2020 Jul 17:e [FREE Full text] [doi: 10.1056/NEJMoa2021436] [Medline: 32678530]
2. Ison MG, Wolfe C, Boucher HW. Emergency Use Authorization of Remdesivir: The Need for a Transparent Distribution Process. JAMA 2020 May 14:e. [doi: 10.1001/jama.2020.8863] [Medline: 32407438]
3. Horby P, Landray M. Statement from the Chief Investigators of the Randomised Evaluation of COVid-19 thERapY (RECOVERY) Trial on hydroxychloroquine, 5 June 2020. recoverytrial.net. 2020 Jun 5. URL: https://www.recoverytrial.net/files/hcq-recovery-statement-050620-final-002.pdf [accessed 2020-06-05]
4. Zhou D, Dai S, Tong Q. COVID-19: a recommendation to examine the effect of hydroxychloroquine in preventing infection and progression. J Antimicrob Chemother 2020 Jul 01;75(7):1667-1670 [FREE Full text] [doi: 10.1093/jac/dkaa114] [Medline: 32196083]
5. Devaux CA, Rolain J, Colson P, Raoult D. New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? Int J Antimicrob Agents 2020 May;55(5):105938 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105938] [Medline: 32171740]
6. Yao X, Ye F, Zhang M, Cui C, Huang B, Niu P, et al. In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Clin Infect Dis 2020 Jul 28;71(15):732-739 [FREE Full text] [doi: 10.1093/cid/ciaa237] [Medline: 32150618]
7. Biot C, Daher W, Chavain N, Fandeur T, Khalife J, Dive D, et al. Design and synthesis of hydroxyferroquine derivatives with antimalarial and antiviral activities. J Med Chem 2006 May 04;49(9):2845-2849. [doi: 10.1021/jm0601856] [Medline: 16640347]
8. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. The Lancet 2020 Mar;395(10229):1033-1034. [doi: 10.1016/s0140-6736(20)30628-0]
9. Xu X, Han M, Li T, Sun W, Wang D, Fu B, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci U S A 2020 May 19;117(20):10970-10975 [FREE Full text] [doi: 10.1073/pnas.2005615117] [Medline: 32350134]
10. Luo P, Liu Y, Qiu L, Liu X, Liu D, Li J. Tocilizumab treatment in COVID-19: A single center experience. J Med Virol 2020 Jul 15;92(7):814-818 [FREE Full text] [doi: 10.1002/jmv.25801] [Medline: 32253759]
11. Guaraldi G, Meschiari M, Cozzi-Lepri A, Milic J, Tonelli R, Menozzi M, et al. Tocilizumab in patients with severe COVID-19: a retrospective cohort study. The Lancet Rheumatology 2020 Aug;2(8):e474-e484. [doi: 10.1016/s2665-9913(20)30173-9]
12. Wilkinson E. RECOVERY trial: the UK covid-19 study resetting expectations for clinical trials. BMJ 2020 Apr 28;369:m1626. [doi: 10.1136/bmj.m1626] [Medline: 32345591]
13. Gao J, Tian Z, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci Trends 2020 Mar 16;14(1):72-73 [FREE Full text] [doi: 10.5582/bst.2020.01047] [Medline: 32074550]
14. Gautret P, Lagier J, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. International Journal of Antimicrobial Agents 2020 Jul;56(1):105949. [doi: 10.1016/j.ijantimicag.2020.105949]

15. Peschken CA. Possible Consequences of a Shortage of Hydroxychloroquine for Patients with Systemic Lupus Erythematosus amid the COVID-19 Pandemic. J Rheumatol 2020 Jun 01;47(6):787-790 [FREE Full text] [doi: 10.3899/jrheum.200395] [Medline: 32269064]
16. Magagnoli J, Narendran S, Pereira F, Cummings TH, Hardin JW, Sutton SS, et al. Efficacy of Hydroxychloroquine and Tocilizumab in Patients With COVID-19: Retrospective Cohort Study. medRxiv 2020 Jun:e [FREE Full text] [doi: 10.1016/j.medj.2020.06.001]
17. Mahevas M, Tran V, Roumier M, Chabrol A, Paule R, Guillaud C, et al. No evidence of clinical efficacy of hydroxychloroquine in patients hospitalized for COVID-19 infection with oxygen requirement: results of a study using routinely collected data to emulate a target trial. medRxiv 2020 Jan:e [FREE Full text] [doi: 10.1101/2020.04.10.20060699]
18. Coronavirus (COVID-19) Update: FDA Revokes Emergency Use Authorization for Chloroquine and Hydroxychloroquine. US Food and Drug Administration. 2020. URL: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-chloroquine-and [accessed 2020-06-05]
19. Carlucci P, Ahuja T, Petrilli C, Rajagopalan H, Jones S, Rahimian J. Hydroxychloroquine and azithromycin plus zinc vs hydroxychloroquine and azithromycin alone: outcomes in hospitalized COVID-19 patients. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.05.02.20080036]
20. Rosenberg ES, Dufort EM, Udo T, Wilberschied LA, Kumar J, Tesoriero J, et al. Association of Treatment With Hydroxychloroquine or Azithromycin With In-Hospital Mortality in Patients With COVID-19 in New York State. JAMA 2020 May 11;323(24):2493 [FREE Full text] [doi: 10.1001/jama.2020.8630] [Medline: 32392282]

Abbreviations

ANOVA: analysis of variance
aOR: adjusted odds ratio
CAD: coronary artery disease
CKD: chronic kidney disease
COPD: chronic obstructive lung disease
CRP: C-reactive protein
DM: diabetes mellitus
EUA: emergency use authorization
FDA: Federal Drug Administration
HCQ: hydroxychloroquine
HTN: hypertension
ICU: intensive care unit
IL-6: interleukin-6
IMV: invasive mechanical ventilation
RT-qPCR: real-time qualitative polymerase chain reaction
TCZ: tocilizumab
uOR: unadjusted odds ratio

Edited by T Sanchez; submitted 24.06.20; peer-reviewed by MDC Miranda Duro, J Li, M Adly, A Dawood, L Lafon-Hughes; comments to author 17.07.20; revised version received 31.07.20; accepted 04.08.20; published 01.09.20.

Please cite as:
Roomi S, Ullah W, Ahmed F, Farooq S, Sadiq U, Chohan A, Jafar M, Saddique M, Khanal S, Watson R, Boigon M
Efficacy of Hydroxychloroquine and Tocilizumab in Patients With COVID-19: Single-Center Retrospective Chart Review
J Med Internet Res 2020;22(9):e21758
URL: https://www.jmir.org/2020/9/e21758
doi:10.2196/21758
PMID:32784192

Original Paper

Public Perceptions and Attitudes Toward COVID-19 Nonpharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data

Caitlin Doogan[1,2], MDataSci; Wray Buntine[2], PhD; Henry Linger[1], PhD; Samantha Brunt[3], MD

[1]Department of Human Centred Computing, Faculty of Information Technology, Monash University, Caulfield, Australia

[2]Department of Data Science and AI, Faculty of Information Technology, Monash University, Clayton, Australia

[3]Royal Perth Hospital, Perth, Australia

Corresponding Author:
Caitlin Doogan, MDataSci
Department of Data Science and AI
Faculty of Information Technology
Monash University
Wellington Rd
Clayton, 3800
Australia
Phone: 61 3 9903 1004
Email: caitlin.doogan@monash.edu

Abstract

Background: Nonpharmaceutical interventions (NPIs) (such as wearing masks and social distancing) have been implemented by governments around the world to slow the spread of COVID-19. To promote public adherence to these regimes, governments need to understand the public perceptions and attitudes toward NPI regimes and the factors that influence them. Twitter data offer a means to capture these insights.

Objective: The objective of this study is to identify tweets about COVID-19 NPIs in six countries and compare the trends in public perceptions and attitudes toward NPIs across these countries. The aim is to identify factors that influenced public perceptions and attitudes about NPI regimes during the early phases of the COVID-19 pandemic.

Methods: We analyzed 777,869 English language tweets about COVID-19 NPIs in six countries (Australia, Canada, New Zealand, Ireland, the United Kingdom, and the United States). The relationship between tweet frequencies and case numbers was assessed using a Pearson correlation analysis. Topic modeling was used to isolate tweets about NPIs. A comparative analysis of NPIs between countries was conducted.

Results: The proportion of NPI-related topics, relative to all topics, varied between countries. The New Zealand data set displayed the greatest attention to NPIs, and the US data set showed the lowest. The relationship between tweet frequencies and case numbers was statistically significant only for Australia (r=0.837, P<.001) and New Zealand (r=0.747, P<.001). Topic modeling produced 131 topics related to one of 22 NPIs, grouped into seven NPI categories: Personal Protection (n=15), Social Distancing (n=9), Testing and Tracing (n=10), Gathering Restrictions (n=18), Lockdown (n=42), Travel Restrictions (n=14), and Workplace Closures (n=23). While less restrictive NPIs gained widespread support, more restrictive NPIs were perceived differently across countries. Four characteristics of these regimes were seen to influence public adherence to NPIs: timeliness of implementation, NPI campaign strategies, inconsistent information, and enforcement strategies.

Conclusions: Twitter offers a means to obtain timely feedback about the public response to COVID-19 NPI regimes. Insights gained from this analysis can support government decision making, implementation, and communication strategies about NPI regimes, as well as encourage further discussion about the management of NPI programs for global health events, such as the COVID-19 pandemic.

(J Med Internet Res 2020;22(9):e21419) doi:10.2196/21419

KEYWORDS

COVID-19; SARS-CoV-2; topic modeling; nonpharmaceutical interventions; social media; public health; machine learning; social distancing; lockdown; face masks; infodemiology

Introduction

SARS-CoV-2, the novel virus causing COVID-19, was declared a pandemic by the World Health Organization (WHO) on March 11, 2020. SARS-CoV-2 has infected millions of people worldwide [1] and continues to threaten population health, as well as the socioeconomic and geopolitical positions, of many countries. In the absence of preventative and curative pharmaceutical treatments specific to COVID-19, governments are reliant on the success of strategic response programs to mitigate, delay, or suppress the transmission of SARS-CoV-2 [2]. Typically, these programs are dynamic regimes of nonpharmaceutical interventions (NPIs) [3]. NPIs are public health measures, aside from the use of pharmaceuticals or vaccinations, deployed by governments and health authorities to control community transmission of disease [3,4].

The restrictiveness of COVID-19 NPI regimes is dependent on a government's strategy to control the transmission of SARS-CoV-2 (eg, elimination versus suppression) [5]. However, for these strategies to be effective, the public must maintain adherence to the prescribed NPIs. Governments and health authorities need to ensure the ongoing public adherence to NPIs to gain control of viral transmission. To do so requires a greater understanding of the trends in the public's perceptions and attitudes toward these NPI regimes, as well as a means of determining why, and in what contexts, these adherence behaviors arise, decrease, and persist [6]. This level of understanding could assist governments in making more informed decisions about NPIs so that they may be made more acceptable to the public. Studies have demonstrated that NPI regimes with a higher degree of public acceptability attract a greater level of public adherence and, ultimately, reduce the rates of infection within a community [7,8].

Understanding public attitudes toward NPIs, together with epidemiological data, may also inform the optimal time when restrictions could be eased or removed, and importantly, how these adjustments are communicated to the public [9]. This knowledge will be imperative in the event these strategies may need to be implemented again to suppress any subsequent waves of infection during later phases of the COVID-19 pandemic.

Epidemiological modeling [10], surveys [2], and the analysis of past pandemics [11] have been the primary means to support decision making around COVID-19 NPI regimes. Several studies have assessed the success of such programs in countries that have passed the peak of infections [12,13], offering valuable lessons for other countries battling the virus [12]. However, a more comprehensive understanding of the public's perceptions and attitudes about NPIs would be of significant additional benefit. Moreover, the need for expedient implementation of NPIs requires rapid analysis. Social media analysis offers this opportunity due to the widespread use of these platforms, and the relative speed of data capture [14].

Recent studies have demonstrated that social media mining could provide information about the public response to specific health measures, such as suspension of sporting matches [15], cancellation of festivals [16], and provisions for vulnerable groups [17]. Our study can further contribute to this literature by demonstrating that social media mining can derive rapid insights about public perceptions and attitudes to NPI regimes.

Given the need for both timely responses to outbreaks [15] and more in-depth insights about trends in the public perception of NPIs, we chose to collect Twitter data, as it is accessible and of sufficient volume for the computational and qualitative analysis needed to provide information on attitudinal trends. Moreover, a comparative analysis of data drawn from multiple geographic regions may assist in the identification of factors that contribute to the understanding of the public acceptability and, ultimately, sustained adherence to NPI regimes.

In this study, we used a hybrid computational approach to analyze English language tweets related to COVID-19 NPIs across six anglophone countries. The aims of this study are to (1) identify which NPIs attract public attention in each of the countries and the extent to which they do so, (2) describe the perceptions and attitudes toward these NPIs and compare these between countries, (3) identify factors that may influence public perception and attitudes to NPI regimes.

Methods

This study adopts a hybrid approach that integrates computational and qualitative techniques to describe the public's perceptions and attitudes toward NPIs.

Data Collection and Preprocessing

Using the Twitter streaming API (application programming interface) service, we collected 2,587,625 tweets posted between January 1 and April 30, 2020. During the 121 days surveyed, COVID-19 was declared a pandemic, and each of the chosen countries had implemented NPI regimes to control the spread of the virus. We considered varying the collection start times for each country but felt there was no strong justification for doing so. The major onset of significant cases was within a week for all the chosen countries and, given that media coverage was international, the onset of tweeting predated the onset of cases in all countries. Tweets were retrieved using three hashtags related to COVID-19 in each country to collect an initial pool of data. We did not include hashtags such as #COVID-19, as this would not allow for an analysis targeted to one particular country. A frequency analysis of cocurrent hashtags informed the choice of secondary hashtags used for further retrieval of tweets. These hashtags, and the results of the frequency analysis, are reported in Multimedia Appendix 1.

Preprocessing of tweets included duplicate removal, lower casing, contraction expansion, and elongation reduction. Nonalphabetic characters, @usernames, #hashtags, URLs, the queried hashtag, and the name of the specific country were

removed. The Python package NLTK [18] was used for tokenization, part-of-speech tagging, lemmatization, and stop-word removal. Using the langid package [19], we identified and removed non-English tweets. Bigrams and trigrams were reviewed for inclusion. Tokens appearing less than 10 times in the data set or tweets containing less than 10 tokens were excluded from further analysis. In total, 777,869 tweets met the criteria for analysis.

Analysis of Public Attention

In the first instance, we determined if the attention of the public, as given by the number of daily tweets, was related to the number of daily confirmed cases. For each data set, we conducted a Pearson correlation analysis to determine if there was a statistically significant relationship between these variables. Graphical representations of these relationships are presented in Multimedia Appendix 2, and the results of this correlation analysis are presented in Multimedia Appendix 3.

Topic Modeling and Evaluation

Topic modeling [20] is used to determine hidden (latent) semantic structures that link documents through the identification of commonly co-occurring word sets. Ranked word sets, or topics, are summative representations of the documents in which they appear. We used MetaLDA [21], a topic model empirically demonstrated to perform better than the popular model Latent Dirichlet Allocation (LDA) [22] when modeling tweets that are sparse and noisy texts [23]. The choice of how many topics (k) to construct is a parameter of MetaLDA and affects the quality of the models produced. A quality model is one with coherent and interpretable topics. However, as the best k is not known, multiple models with different values for k (k = {30,40 … 130}) were constructed for each data set [24]. The values of k and the number of models constructed was determined by the size of the data set. The selection of quality models is informed by statistical and qualitative evaluation.

Statistical evaluation of models was based on the coherence measure normalized pointwise mutual information (NPMI) [25], calculated using the Palmetto package [26], and Glove2Vec embeddings [27]. Embeddings were trained on the Wikipedia corpus for all data sets except the Australian and New Zealand data sets, which were trained on a 150-million-word corpus of Australian news articles. We report these evaluations in Multimedia Appendix 4. Models with higher mean NPMI scores were selected for a qualitative review of their topics. The models with the most coherent and interpretable topics were further analyzed.

Comparative Analysis

Qualitative analysis of topics involved the construction of coding schema to guide the identification and labeling of NPI-related topics. This schema is displayed in Multimedia Appendix 5. The schema was primarily informed by the WHO advisory framework for influenza pandemic NPIs [11], supported by academic literature, and the specific COVID-19 NPI regimes being employed by the six countries at the time of analysis. In this process, we reviewed up to 100 of the most representative tweets per topic to determine if the topic was about NPIs and, if so, labeled with the relevant NPI. We also examined topics for inclusion when these topics referenced NPIs in related topics. Within the research team, we cross-checked the NPI labels assigned to topics [28]. Multimedia Appendix 4 displays the number of topics associated with NPIs for each country.

A comparative analysis of the public discussion of NPIs between countries was conducted. Tweets were recontextualized by reviewing embedded hypermedia, replies, and linked media. We supplemented our analysis by surveying COVID-19–related information provided by each country's government during the study period. Heatmaps were constructed to identify which NPIs attracted the public's attention for each country and the extent to which they did so. The proportion of each country's discussion of an NPI was calculated for each data set as the total number of tweets assigned to topics about an NPI, relative to the total number of tweets assigned to all topics about NPIs. To explore further, we constructed chord diagrams to visualize the relationships between discussions of NPIs in each country. These diagrams, shown in Multimedia Appendix 2, illustrate the number of tweets shared between categories of NPIs.

Results

Analysis of Public Attention

A Pearson correlation analysis showed that the number of daily COVID-19 cases and the daily number of tweets were strongly correlated only in the Australian (r=0.837, P<.001) and New Zealand (r=0.747, P<.001) data sets for the 121 days surveyed. The analysis was also conducted based on the date of the first confirmed case in each country. Again, a strong correlation was seen only for Australia (r=0.823, P<.001) and New Zealand (r=0.666, P<.001). The results of this analysis are shown in Multimedia Appendix 3.

Graphically, it appears that public attention dissipated over time, despite the number of cases continuing to rise exponentially in the surveyed countries except for Australia and New Zealand. These two countries brought their relatively small number of cases under control during the study period, which may account for the correlation between the decline in tweets and cases. Scaled visualizations of the number of tweets and cases per day for each data set are displayed in Multimedia Appendix 2.

Comparative Analysis

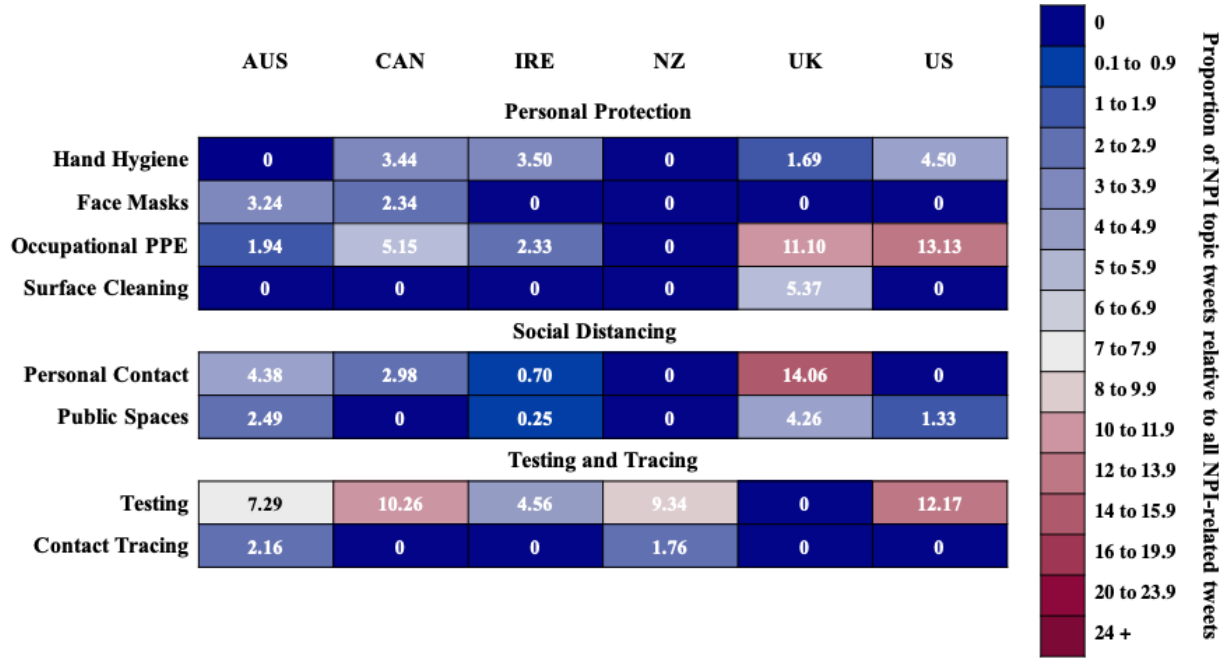
In total, 131 NPI-related topics discussed one of 22 NPIs that were identified. These NPIs were categorized into seven types of NPIs: Personal Protection (n=15), Social Distancing (n=9), Testing and Tracing (n=10), Gathering Restrictions (n=18), Lockdown (n=42), Travel Restrictions (n=14), and Workplace Closures (n=23). The qualitative details of these topics are in Multimedia Appendix 4.

The following discussion details the results of a comparative analysis of the public's perceptions and attitudes of NPIs between countries. We have divided the analysis and accompanying graphical representations into less restrictive and more restrictive NPIs based on their level of intrusiveness and economic cost. Exemplar tweets that illustrate public perceptions of NPIs are in Multimedia Appendix 6.

Analysis of Less Restrictive NPIs

Less restrictive NPIs include Personal Protection, Social Distancing, and Testing and Tracing. The proportions of tweets associated with these three categories are shown in Figure 1.

Figure 1. A heatmap of the proportions of tweets associated with less restrictive nonpharmaceutical interventions (NPIs) for each country. PPE: personal protective equipment.



Personal Protection

Four NPIs about Personal Protection were identified: Hand Hygiene, Face Masks, Occupational Personal Protective Equipment (PPE), and Surface Cleaning.

Tweets about hand hygiene were concentrated in the early period of each country's response. While the proportion of topics was relatively low, hand hygiene often featured in tweets discussing more prominent NPIs. Tweets from the United Kingdom referenced the National Health Service's (NHS) recommendation that hand washing length should equate with the amount of time it takes to sing the "Happy Birthday" song [29]. Tweets in the United States tended to be informative, explaining that soap and warm water was better than hand sanitizers to kill SARS-CoV-2 [30] and that people should stop touching their faces. In contrast, UK tweets expressed frustration at low handwashing compliance and public preference for alcohol-based hand sanitizers. Similarly, Canadian tweets promoted handwashing over the use of disposable gloves and considered those misusing gloves to be irresponsible.

Despite the Australian government's advice that face masks were not generally necessary for the public [31], Australian tweets perceived face masks to be an important defense against transmission. Canadian tweets expressed confusion over changing government advice regarding the wearing of face masks.

Except for tweets from New Zealand, the public perceived the shortage of occupational PPE to be the result of poor government decision making. Tweets from the United States expressed alarm at these shortages and frequently referenced ongoing media coverage of PPE shortages in outbreak areas. Perceptions diverged, however, as the supply of Chinese-sourced PPE increased in different countries. Irish tweets celebrated the highly publicized shipments, while Australian tweets expressed skepticism of the political motivations behind these deliveries. Canadian tweets voiced concern at the quality of Chinese PPE and lamented the lack of national capacity for PPE manufacturing.

In response to a shortage of cleaning products [32], Canadian and UK tweets encouraged surface cleaning, shared recipes for homemade disinfectants, and discussed practical measures to reduce surface transmission, including cleaning mobile phones.

Social Distancing

Two NPIs about Social Distancing were identified: Personal Contact and Public Spaces.

UK and Irish tweets discouraged handshaking and reported the rapid adoption of noncontact greetings, including elbow-bumping and foot tapping. However, UK tweets found the cessation of handshaking redundant in full-contact sports given the high degree of bodily contact. The UK discussion of personal contact was highly concentrated, with most tweets reminding others to be mindful of keeping their distance, particularly from the elderly.

Transmission in public spaces, such as on public transport, was perceived as a threat by the Australian, Irish, UK, and US public. US and UK tweets expressed dissatisfaction with the solutions authorities had imposed, such as reducing carriage capacity in London's underground rail services. Irish tweets complained about the challenges of social distancing, given the narrow footpaths in cities. Australian tweets supported social distancing but questioned the arbitrary nature of a 1.5 m (5 ft) specification rather than 2 m (6.5 ft) [33] seen in other countries. Australian and Irish tweets questioned the logic of social distancing when

other NPI implementations did not account for the practice, particularly in specific workplaces.

Testing and Tracing

Two NPIs about Testing and Tracing were identified: Testing and Tracing Apps.

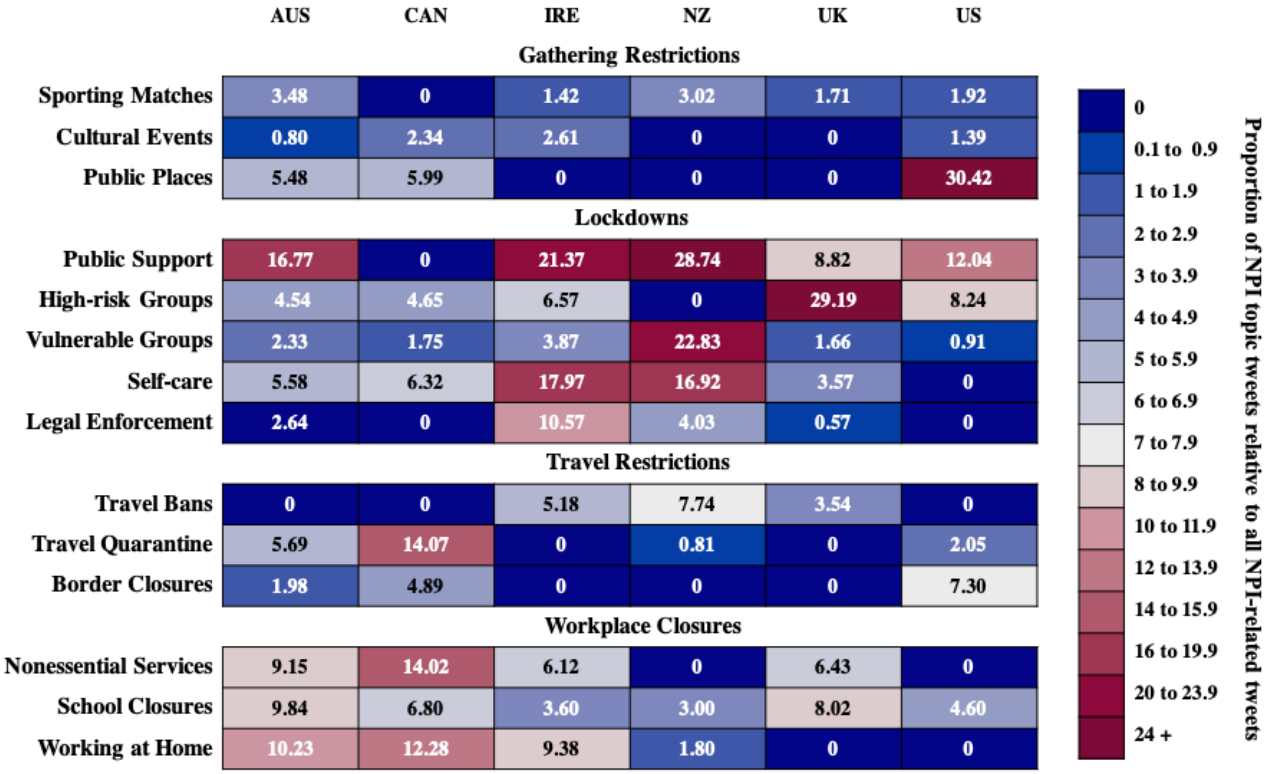
The highest proportions of topics about testing were from the Canadian and US data sets, where the public perception of testing efforts was negative. This was due to inconsistent testing criteria, testing backlogs, faulty testing kits, costs, and a lack of access to tests. Australian tweets viewed the reports of low community transmission to be a statistical manipulation, given testing at that time was restricted to returning travelers. Similarly, New Zealand and Irish tweets questioned the validity of case reports given the testing criteria.

The Australian and New Zealand public expressed concern about the privacy implications of contact tracing apps. The public understood the specifics of data collection, but tweets expressed distrust of tech companies that would be engaged to store the data. New Zealand tweets considered the implications of centralized data models and called for open source code and an independent privacy assessment.

Analysis of Restrictive NPIs

Categories of restrictive NPIs include Gathering Restrictions, Lockdowns, Travel Restrictions, and Workplace Closures. The proportions of tweets associated with these three categories are shown in Figure 2.

Figure 2. A heatmap of the proportion of tweets associated with restrictive nonpharmaceutical interventions (NPIs) for each country.



Gathering Restrictions

Three NPIs about Gathering Restrictions were identified: Sporting Matches, Cultural Events, and Public Places.

US tweets had a mixed response to the suspension of several major US sporting leagues announced on March 12, 2020. Though disappointed, many accepted the measure as necessary, while others perceived it to be an overreaction. The latter response could be explained by the low number of cases in several states at that time. Except for New Zealand, where the public looked forward to the resumption of elite sports, the suspension of sporting seasons was discussed with similarly disappointed acceptance. UK tweets expressed shock at the sudden cancellation of the English Premier League season. Australian tweets expressed anger toward politicians who were perceived to delay restrictions on mass gatherings based on their support for specific sporting teams.

Discussion about cultural events included religious gatherings. Tweets from Ireland displayed disbelief at the cancellation of Catholic Mass but quickly embraced the use of live streaming of services. Canadian tweets expressed solidarity and support for those celebrating religious holidays. Conversely, US tweets demonstrated anger toward specific religious figures or groups who continued to gather in large groups.

Irish and US tweets discussed the cancellation of St. Patrick's Day (March 17, 2020) festivities. Interestingly, the proportion of Irish tweets, which uniformly demanded the cancellation of festivities, were less than those seen in the US data set where the response was mixed. The cancellation of Australia's ANZAC day services (April 25, 2020) had a sobering effect, and while many encouraged the observation of services at home, an attitudinal shift was noted regarding the seriousness of the situation.

The US NPI topics on gatherings in public places accounted for 30.42% (23,099/75,938) of NPI-related tweets, the most intense of all NPIs. The majority of tweets passionately debated the implications for the 2020 US general election. The US public perceived the then-current campaign schedule as a threat to community health and urged postponement or postal voting to avoid increased transmission. Compulsory in-person voting in the Australian Queensland by-election (March 28, 2020) attracted scathing commentary from those who perceived the directive to be inconsistent with social distancing mandates.

Media coverage of beachgoers failing to socially distance in Florida and Sydney saw US and Australian tweets calling for beach closures. Additionally, Australian and Canadian tweets called for closures of running paths if social distancing did not improve.

Lockdowns

Five NPIs pertained to Lockdowns: Public Support, High-Risk Groups Vulnerable Groups, Self-Care, and Legal Enforcement.

There was broad support for lockdowns across countries. Before the implementation of national lockdowns in Australia, Ireland, and the United Kingdom, the public showed increasing frustration with the delayed implementation of these lockdowns. UK tweets called for tighter restrictions as a means of deterring noncompliant behaviors and protecting the NHS from being overwhelmed. Earlier Australian tweets, comparing their incremental introduction of restrictions to the New Zealand lockdown [34], indicated a preference for a complete lockdown. New Zealand tweets were calm and accepting. “Be kind” was consistently seen in these tweets, reflecting the New Zealand government’s “Unite against COVID-19” campaign messages [35].

The message to stay at home was promoted in the US and UK data sets with a sense of urgency. These tweets pleaded for people to “stay home and save lives.” UK tweets posted before the implementation of a lockdown perceived staying home as a civic duty. Australian tweets encouraged people to stay at home, but acknowledged people needed some flexibility to do so.

A large proportion of UK tweets expressed concern for high-risk community members (64,063/219,485, 29.19%), such as the elderly, and their protection was seen to be a community responsibility. Both UK and Australian tweets applauded grocery chains that reserved times for elderly or disabled individuals to make purchases. Discussion of outbreaks in aged care facilities differed between countries. Canadians blamed the outbreaks on inadequate PPE and cross-facility rostering of the staff, whereas Australian tweets blamed staff who attended work when ill, and Irish tweets blamed the broader public for not adhering to restrictions. Both the Irish and US tweets reported sadness and disappointment that their efforts to quarantine high-risk family members in aged care homes were ineffective.

The proportion of New Zealand tweets about vulnerable groups was the largest of all data sets (10,844/47,500, 22.83%). The discussion was mainly about the protection of the Māori people from the pandemic, which was perceived by the public as a national responsibility and called for culturally appropriate response plans.

The vulnerability of individuals in prison and immigration detention was discussed in all data sets except New Zealand. UK tweets appeared sympathetic to those in prison but also perceived early prison release programs as a risk to public safety. Canadian tweets supported only the release of those on remand for nonviolent crimes. US tweets did not agree with early prison release programs. Australian, UK, Canadian and Irish tweets aimed to raise awareness of the vulnerability of both refugees and those in immigration detention. Canadian and Irish tweets discussed the risk of transmission in homeless and women's shelters and amongst itinerant communities. The requisitioning of hotels to facilitate self-isolation and increased hygiene was supported.

Aside from the United States, tweets from all countries demonstrated a positive attitude toward staying well and discussed the importance of self-care, nutrition, routines, social connection, and exercise. Australian tweets self-reported mental health symptoms but indicated proactive management. In contrast, pessimistic UK tweets discussed the lack of funding for mental health services.

Attitudes toward increased police and military powers to enforce restrictions were perceived differently within the Irish data set. While the majority supported increased police presence in public, their role in enforcing lockdown received some criticism. Similarly, there was debate over a recent challenge in the Irish high court over the enforcement of these restrictions. The Australian mainstream media labeled the legal enforcement of restrictions as draconian. However, this perception was not consistent with the public, who mostly supported legal enforcement. The exception, however, was for the issuing of “frivolous” noncompliance fines, which were perceived as overzealous and unnecessary. New Zealand tweets demonstrated a positive attitude toward police and their approach to enforcement. UK tweets were mainly supportive of military enforcement and anxious to see these measures undertaken.

Travel Restrictions

Three NPIs about Travel Restrictions were identified: Travel Bans, Travel Quarantine, and Border Closures.

New Zealand tweets were highly supportive of the government's decision to place a travel ban on arrivals from mainland China who were not residents of New Zealand. Referencing New Zealand’s travel ban, UK and Irish tweets were outraged that similar measures had not yet been enacted.

US tweets called for increased quarantine measures for cruise ships to the standard applied to air travel. Australian tweets displayed outrage that passengers of the Ruby Princess cruise ship, from which 22 deaths and 700 cases were reported [36], were told only to self-quarantine. Both the Australian and Canadian public lacked confidence in people’s adherence to self-quarantine rules. Australian tweets supported the forcible quarantining of return-travelers, although many asked for further clarification about this decision.

Australian tweets supported the closure of international borders [37] but discussed difficulties for Australian residents returning home. Both Australian and US tweets called for internal border closures, praising leaders who took this measure. US and Canadian tweets called for the US-Canadian border to be closed due to the rapid rise in US cases.

Workplace Closures

Three NPIs about Workplace Closures were identified: Nonessential Services, School Closures, and Working at Home.

Australian, Irish, UK, and US tweets called for the closure of nonessential businesses. Australian and UK tweets reasoned that keeping businesses open encouraged noncompliant behaviors. Irish tweets perceived the closure of schools to be illogical when the hospitality industry remained open.

Parents concerned about the health of their children accounted for the majority of Australian, Canadian, Irish, and UK tweets about school closures. Homeschooling and curriculum continuity were discussed, with many seeking advice about accessing online learning materials, and the implications for high-school and university exams. Australian tweets were frustrated by the discrepancies between federal and state positions on school closures and the contradictory health advice about transmission in children.

Australian tweets highlighted the positive aspects of working at home and expressed surprise at the ease of transition to online meetings. Canadian tweets were contemplative of the adaptions but embraced their new arrangements as time progressed.

Discussion

Principal Findings

In this study, we found that Twitter offers a means by which governments and health authorities can gain rapid feedback about public perceptions and attitudes on NPIs. Topic modeling was used to identify seven categories of NPIs discussed in tweets from the six selected countries. A comparative analysis of NPI topics showed that less restrictive NPIs were broadly supported in all data sets with much of the public encouraging adherence to these restrictions. However, public attitudes toward restrictive NPIs differed between countries.

Four characteristics of NPI regimes were common to all countries and identified as potential predictors of public adherence to NPIs: timeliness of implementation, NPI campaign strategies, inconsistent information, and enforcement strategies.

The timeliness of the implementation of restrictive NPIs influenced public attitudes. Prolonged, staggered, or delayed implementation of restrictive NPIs was met by an angry and fearful public response and demands for restrictions to be increased, as seen in Australian, UK, and Irish tweets. Conversely, tweets from New Zealand, where there was a sudden and total implementation of highly restrictive NPIs, including lockdowns and workplace closures, showed overwhelming support for the regime when enacted despite the low number of cases. These observations indicate that delayed implementation of restrictions may heighten the public's sense of uncertainty about potentially increased restrictions.

Uncertainty may also be a result of inconsistencies between government recommendations, medical expertise, and global health organizations' advice. Information about NPIs not offered by governments was shared broadly on Twitter. For example, many tweets discouraged face touching despite there being limited government messaging regarding this behavior across the six countries. Information asymmetry and inconsistency had a greater effect when NPIs were more intrusive. We observed that Australian and UK public perceptions of face masks were inconsistent with government advice. Public attitudes toward face masks were mixed. Despite them not being recommended by either country's governments at the time, they were mainly regarded as appropriate by the public. However, many debated their necessity and efficacy, often supporting their position with scientific literature, references to other governments, health organizations' recommendations, and media articles. The public's access to alternative information via Twitter may impact their confidence in national NPI regimes when it conflicts with advice from government and health bodies.

Another influencing characteristic was the style of government NPI campaign strategies. Both the UK and New Zealand governments adopted strategies that fostered a strong sense of collective action. A key point of difference was the emphasis on unity, clarity, and empathy, as seen in New Zealand's "Unite against COVID-19" campaign with the public expressing positive attitudes toward NPIs. Campaign messages were fashioned into popular hashtags, including #BeKind, #BreaktheChain, and #StayInYourBubble. Significantly, New Zealand tweets emphasized empathy and kindness, a key element of the government's campaign strategy.

Conversely, the UK government adopted an instructive campaign to "stay home, protect the NHS, save lives." The public perceived it to be their collective responsibility to protect health care workers and those most at risk. However, the delayed implementation of more restrictive NPIs resulted in negative attitudes toward the government. These results indicate that effect- and emotion-based campaigns are potentially more compelling in maintaining adherence to NPIs, but their success is reliant on other factors.

The enforcement of social distancing and lockdowns appeared not to impact the public attitudes toward restrictions. This may be because there was already overwhelming support for these measures. Compliance in the United Kingdom has been linked to people's intrinsic motivation to obey the law [38]. However, we observed that noncompliant behaviors were unintentional and a result of a misunderstanding of the rules. As suggested by previous studies [7], these findings suggest that the ambiguous and dynamic nature of NPIs and the way they are communicated were factors that contributed to comprehension.

Strengths and Limitations

This study has a number of limitations. First, the demographics of Twitter users may not necessarily represent the populations of each of the selected countries [39]. Second, the choice to include only English tweets means that nonanglophones were not represented. However, our methods are consistent with medical research that makes use of topic modeling of tweets and is often restricted to the official language of the country

[39]. Furthermore, we expand upon previous approaches by undertaking a structured qualitative analysis of topic document collections, which are further contextualized through the review of related hypermedia. This hybrid approach provides the depth of insight offered by qualitative methods, with the speed and scalability of computational techniques.

Conclusions

The effectiveness of COVID-19 NPI regimes is dependent on ongoing public adherence. Given this, it is necessary to understand public perceptions and attitudinal trends toward these NPIs, as well as why and in what contexts these behaviors arise and persist [6]. As such, our study was motivated by the need to inform an understanding of such trends to support government communication strategies, as well as the planning and implementation of more effective NPI regimes in later phases of the COVID-19 pandemic.

We undertook a hybrid computational analysis of the public discussion of COVID-19 NPIs across six countries. As detailed above, four characteristics of NPIs regimes were identified as potential predictors of public adherence. The outcome of our analysis is that the widespread public acceptance of NPI regimes is predicated on the public's understanding, timeliness of implementation, the ability of governments to clearly communicate and justify the complexity of the regime, and importantly, their ability to implement the regime without ambiguity or undue enforcement measures.

Ongoing analysis of social media offers governments and health authorities insights into how their programs are heard as well as a critical perspective of their communication strategies. Such feedback should be integrated to produce a more effective public health response to the ongoing pandemic as well as future disease outbreaks. Ongoing and expanded analyses of social media will contribute to a richer understanding of attitudinal and behavioral drivers to inform public health strategies.

Acknowledgments

CD is supported by an Australian Government Research Training Program (RTP) Stipend, RTP Fee-Offset Scholarship, and an Australian Government Defence Science and Technology Group Postgraduate Research Scholarship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hashtags used for tweet retrieval.
[DOCX File , 45 KB - jmir_v22i9e21419_app1.docx]

Multimedia Appendix 2
Graphical analysis of the frequency of tweets per day against the number of confirmed cases per day, and chord diagrams of NPI topic category tweet co-occurrences.
[DOCX File , 1110 KB - jmir_v22i9e21419_app2.docx]

Multimedia Appendix 3
Results of the Pearson correlation analysis of the relationship between the number of confirmed cases and tweets per day.
[DOCX File , 19 KB - jmir_v22i9e21419_app3.docx]

Multimedia Appendix 4
NPI topic word sets, topic model evaluations, and the frequency of topics per NPI for each country.
[DOCX File , 41 KB - jmir_v22i9e21419_app4.docx]

Multimedia Appendix 5
NPI topic coding schema.
[DOCX File , 29 KB - jmir_v22i9e21419_app5.docx]

Multimedia Appendix 6
Exemplar tweets that illustrate the public discussion of NPIs.
[DOCX File , 20 KB - jmir_v22i9e21419_app6.docx]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-05-02]

2. Betsch C, Wieler LH, Habersaat K. Monitoring behavioural insights related to COVID-19. The Lancet 2020 Apr;395(10232):1255-1256. [doi: 10.1016/s0140-6736(20)30729-7]
3. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
4. Ferguson N, Laydon D, Nedjati-Gilani G. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team. 2020. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [accessed 2020-08-21]
5. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health 2020 Jul;5(7):e375-e385 [FREE Full text] [doi: 10.1016/S2468-2667(20)30133-X] [Medline: 32502389]
6. Qualls N, Levitt A, Kanade N, Wright-Jegede N, Dopson S, Biggerstaff M, CDC Community Mitigation Guidelines Work Group. Community Mitigation Guidelines to Prevent Pandemic Influenza - United States, 2017. MMWR Recomm Rep 2017 Apr 21;66(1):1-34 [FREE Full text] [doi: 10.15585/mmwr.rr6601a1] [Medline: 28426646]
7. Sharma R, Kaur S, Sodhi A. Knowledge, behaviour change, and anticipated compliance regarding non-pharmaceutical interventions during pandemic of influenza A H1N1 in Delhi. Lung India 2012 Oct;29(4):341-346 [FREE Full text] [doi: 10.4103/0970-2113.102817] [Medline: 23243348]
8. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. The Lancet Respiratory Medicine 2020 May;8(5):434-436. [doi: 10.1016/s2213-2600(20)30134-x]
9. Malecki K, Keating JA, Safdar N. Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. Clin Infect Dis 2020 Jun 16 [FREE Full text] [doi: 10.1093/cid/ciaa758] [Medline: 32544242]
10. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health 2020 May;5(5):e261-e270 [FREE Full text] [doi: 10.1016/S2468-2667(20)30073-6] [Medline: 32220655]
11. Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza. World Health Organization. 2019. URL: https://www.who.int/influenza/publications/public_health_measures/publication/en/ [accessed 2020-08-21]
12. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]
13. Viner RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. The Lancet Child & Adolescent Health 2020 May;4(5):397-404 [FREE Full text] [doi: 10.1016/S2352-4642(20)30095-X] [Medline: 32272089]
14. El Tantawi M, Al-Ansari A, AlSubaie A, Fathy A, Aly NM, Mohamed AS. Reach of Messages in a Dental Twitter Network: Cohort Study Examining User Popularity, Communication Pattern, and Network Structure. J Med Internet Res 2018 Sep 13;20(9):e10781 [FREE Full text] [doi: 10.2196/10781] [Medline: 30213781]
15. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
16. Han X, Wang J. Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China. IJGI 2019 Apr 09;8(4):185. [doi: 10.3390/ijgi8040185]
17. Jimenez-Sotomayor MR, Gomez-Moreno C, Soto-Perez-de-Celis E. Coronavirus, Ageism, and Twitter: An Evaluation of Tweets about Older Adults and COVID-19. J Am Geriatr Soc 2020 Aug 09;68(8):1661-1665 [FREE Full text] [doi: 10.1111/jgs.16508] [Medline: 32338787]
18. Bird S, Klein E, Loper E. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit. Massachusetts, United States: O'Reilly Media, Inc; 2009.
19. Lui M, Baldwin T. langid.py: An Off-the-shelf Language Identification Tool. In: Proceedings of the ACL 2012 System Demonstrations. 2012 Presented at: Proceedings of the ACL system demonstrations; July 2012; Jeju Island, Korea p. 25-30 URL: https://www.aclweb.org/anthology/P12-3005.pdf
20. Boyd-Graber J, Hu Y, Mimno D. Applications of Topic Models. Foundations and Trends® in Information Retrieval 2017;11(2-3):143-296. [doi: 10.1561/1500000030]
21. Zhao H, Du L, Buntine W, Liu G. Leveraging external information in topic modelling. Knowl Inf Syst 2018 May 12;61(2):661-693. [doi: 10.1007/s10115-018-1213-y]
22. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J of Mach Learn Res 2003;3:993-1022 [FREE Full text] [doi: 10.1145/2488388.2488514]
23. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. In: WWW '13: Proceedings of the 22nd International Conference on World Wide. 2013 Presented at: WWW '13: Proceedings of the 22nd International Conference on World

Wide; May 2013; Rio de Janeiro, Brazil p. 22 URL: https://dl.acm.org/doi/10.1145/2488388.2488514 [doi: 10.1145/2488388.2488514]
24. Lau J, Newman D, Baldwin T. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. 2014 Presented at: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics; April 2014; Gothenburg, Sweden URL: https://www.aclweb.org/anthology/E14-1056/ [doi: 10.3115/v1/e14-1056]
25. Bouma G. Normalized (pointwise) mutual information in collocation extraction. In: From Form to Meaning: Processing Texts Automatically. 2009 Presented at: Biennial GSCL Conference; 2009; Tubingen p. 31-40 URL: https://pdfs.semanticscholar.org/1521/8d9c029cbb903ae7c729b2c644c24994c201.pdf
26. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. 2015 Presented at: WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining; February 2015; Shanghai, China. [doi: 10.1145/2684822.2685324]
27. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. 2014 Presented at: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar URL: https://www.aclweb.org/anthology/D14-1162/ [doi: 10.3115/v1/d14-1162]
28. Barry CA, Britten N, Barber N, Bradley C, Stevenson F. Using reflexivity to optimize teamwork in qualitative research. Qual Health Res 1999 Jan;9(1):26-44. [doi: 10.1177/104973299129121677] [Medline: 10558357]
29. Public information campaign focuses on handwashing. Department of Health and Social Care. 2020 Mar. URL: https://www.gov.uk/government/news/public-information-campaign-focuses-on-handwashing [accessed 2020-05-22]
30. Tuladhar E, Hazeleger W, Koopmans M, Zwietering M, Duizer E, Beumer R. Reducing viral contamination from finger pads: handwashing is more effective than alcohol-based hand disinfectants. J Hosp Infect 2015 Jul;90(3):226-234. [doi: 10.1016/j.jhin.2015.02.019] [Medline: 25936671]
31. Information on the use of surgical masks. Department of Health, Australian Government. 2020 Mar. URL: https://www.health.gov.au/sites/default/files/documents/2020/03/coronavirus-covid-19-information-on-the-use-of-surgical-masks.pdf [accessed 2020-05-01]
32. Bogart N. 'Nobody was ready': Retail Council of Canada urges consumers to stop panicking as grocers stripped bare. CTV News. 2020. URL: https://www.ctvnews.ca/health/coronavirus/nobody-was-ready-retail-council-of-canada-urges-consumers-to-stop-panicking-as-grocers-stripped-bare-1.4851901 [accessed 2020-05-19]
33. Coronavirus Disease 2019 (COVID-19): Being prepared. Public Health Agency of Canada. 2020 Mar. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/being-prepared.html [accessed 2010-05-20]
34. Ardern J. New Zealand moves to COVID-19 Alert Level 3, then Level 4 in 48 hours. New Zealand Government. 2020 Mar. URL: https://www.beehive.govt.nz/release/new-zealand-moves-covid-19-alert-level-3-then-level-4-48-hours [accessed 2020-05-19]
35. Unite against COVID-19. New Zealand Government. 2020 Apr. URL: https://uniteforrecovery.govt.nz/ [accessed 2020-05-05]
36. Cockburn P. How the coronavirus pandemic would look in Australia if Ruby Princess had never docked. ABC News. 2020 Apr. URL: https://web.archive.org/web/20200423064948/https://www.abc.net.au/news/2020-04-23/coronavirus-across-australia-if-ruby-princess-never-docked/12172314 [accessed 2020-04-25]
37. Morrison S. Press Conference: Australian Parliament House. web.archive.org. 2020 Mar. URL: https://web.archive.org/web/20200412121759/https://www.pm.gov.au/media/press-conference-australian-parliament-house-act-10 [accessed 2020-04-24]
38. Kooistra EB, Reinders Folmer C, Kuiper ME, Olthuis E, Brownlee M, Fine A, et al. Mitigating COVID-19 in a Nationally Representative UK Sample: Personal Abilities and Obligation to Obey the Law Shape Compliance with Mitigation Measures. SSRN Journal 2020:e [FREE Full text] [doi: 10.2139/ssrn.3598221]
39. Griffis H, Asch DA, Schwartz HA, Ungar L, Buttenheim AM, Barg FK, et al. Using Social Media to Track Geographic Variability in Language About Diabetes: Analysis of Diabetes-Related Tweets Across the United States. JMIR Diabetes 2020 Jan 26;5(1):e14431 [FREE Full text] [doi: 10.2196/14431] [Medline: 32044757]

Abbreviations

API: application programming interface
LDA: Latent Dirichlet Allocation
NHS: National Health Service
NPI: nonpharmaceutical intervention
NPMI: normalized pointwise mutual information
PPE: personal protective equipment
RTP: Research Training Program
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.06.20; peer-reviewed by Y Liu, A Ferrario; comments to author 07.07.20; revised version received 05.08.20; accepted 06.08.20; published 03.09.20.

Please cite as:
Doogan C, Buntine W, Linger H, Brunt S
Public Perceptions and Attitudes Toward COVID-19 Nonpharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data
J Med Internet Res 2020;22(9):e21419
URL: https://www.jmir.org/2020/9/e21419
doi:10.2196/21419
PMID:32784190

Original Paper

Willingness to Use Home Collection Methods to Provide Specimens for SARS-CoV-2/COVID-19 Research: Survey Study

Eric William Hall[1], MPH, PhD; Nicole Luisi[1], MPH, MS; Maria Zlotorzynska[1], PhD, MPH; Gretchen Wilde[1], MPH; Patrick Sullivan[1], PhD, DVM; Travis Sanchez[1], MPH, DVM; Heather Bradley[2], MHS, PhD; Aaron J Siegler[3], MHS, PhD

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[2]Department of Epidemiology & Biostatistics, School of Public Health, Georgia State University, Atlanta, GA, United States

[3]Department of Behavioral, Social and Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Aaron J Siegler, MHS, PhD
Department of Behavioral, Social and Health Education Sciences
Rollins School of Public Health
Emory University
1518 Clifton Road
Atlanta, GA, 30322
United States
Phone: 1 404 712 9733
Email: asiegle@emory.edu

Abstract

Background: Innovative laboratory testing approaches for SARS-CoV-2 infection and immune response are needed to conduct research to establish estimates of prevalence and incidence. Self-specimen collection methods have been successfully used in HIV and sexually transmitted infection research and can provide a feasible opportunity to scale up SARS-CoV-2 testing for research purposes.

Objective: The aim of this study was to assess the willingness of adults to use different specimen collection modalities for themselves and children as part of a COVID-19 research study.

Methods: Between March 27 and April 1, 2020, we recruited 1435 adults aged 18 years or older though social media advertisements. Participants completed a survey that included 5-point Likert scale items stating how willing they were to use the following specimen collection testing modalities as part of a research study: home collection of a saliva sample, home collection of a throat swab, home finger-prick blood collection, drive-through site throat swab, clinic throat swab, and clinic blood collection. Additionally, participants indicated how the availability of home-based collection methods would impact their willingness to participate compared to drive-through and clinic-based specimen collection. We used Kruskal-Wallis tests and Spearman rank correlations to assess if willingness to use each testing modality differed by demographic variables and characteristics of interest. We compared the overall willingness to use each testing modality and estimated effect sizes with Cohen *d*.

Results: We analyzed responses from 1435 participants with a median age of 40.0 (SD=18.2) years and over half of which were female (761/1435, 53.0%). Most participants agreed or strongly agreed that they would be willing to use specimens self-collected at home to participate in research, including willingness to collect a saliva sample (1259/1435, 87.7%) or a throat swab (1191/1435, 83.1%). Willingness to collect a throat swab sample was lower in both a drive-through setting (64%) and clinic setting (53%). Overall, 69.0% (990/1435) of participants said they would be more likely to participate in a research study if they could provide a saliva sample or throat swab at home compared to going to a drive-through site; only 4.4% (63/1435) of participants said they would be less likely to participate using self-collected samples. For each specimen collection modality, willingness to collect specimens from children for research was lower than willingness to use on oneself, but the ranked order of modalities was similar.

Conclusions: Most participants were willing to participate in a COVID-19 research study that involves laboratory testing; however, there was a strong preference for home specimen collection procedures over drive-through or clinic-based testing. To increase participation and minimize bias, epidemiologic research studies of SARS-CoV-2 infection and immune response should consider home specimen collection methods.

(J Med Internet Res 2020;22(9):e19471) doi:10.2196/19471

KEYWORDS

COVID-19; SARS-CoV-2; specimen collection; survey; research; public health; infectious disease; virus; test

Introduction

The first case of the novel coronavirus SARS-CoV-2 in the United States was identified on January 20, 2020 [1]. By April 8, the number of reported cases in the United States had surpassed 400,000 [2]. Over that same time period, roughly 2 million specimens had been tested for SARS-CoV-2 RNA in laboratories across the country [3]. While increased laboratory capacity and the opening of drive-through facilities has increased testing access and case identification, these strategies remain insufficient to support the population-based research required to characterize the epidemiologic nature of this outbreak because testing is focused in clinical settings and on people with symptoms of disease. To develop a better understanding of the exposure, disease, and recovery process associated with SARS-CoV-2 infection, infectious disease researchers have called for innovative testing approaches and a rapid scaleup in the number of persons tested [4].

Self-specimen collection for testing has been successfully used in HIV and sexually transmitted infection (STI) research for well over a decade [5-8]. A review of 25 HIV testing studies found that across multiple specimen methods (finger prick, oral swabs), self-collection results had the same diagnostic accuracy as clinician-collected specimens, with no differences in the proportions of invalid results [9]. Another review comparing participant self-collection versus clinician collection for gonorrhea and chlamydia also found high performance (>90% sensitivity and specificity) for self-collected specimens [10]. In one of our previous studies, 93% of participants were able to successfully complete multiple specimen collections, and 85% preferred self-collection of specimens at home to a standard office visit [11]. In the present analysis, we aimed to assess the willingness of adults to use different specimen collection methods on themselves or their children as part of a COVID-19 research study. We hypothesized that modalities for home specimen collection would be preferred over clinic-based specimen collection.

Methods

Recruitment

Participants were recruited through web-based social media advertisements on Facebook, Snapchat, and Twitter from March 27, 2020 to April 1, 2020. Internet users who clicked on the advertisements were taken to a consent module and short screener to determine eligibility. Eligible respondents were adults aged ≥18 years. On the last day of recruitment, we oversampled Hispanic and Black respondents with targeted ads to increase the racial and ethnic diversity of the sample. Eligible participants completed a web-based survey that collected data on their demographics, current knowledge of COVID-19, stigma related to COVID-19, and relevant symptoms over the last 24 hours. We used cookie-based duplicate protection, which restricts respondents from completing the survey more than once from the same browser on the same device. Participants were not compensated for their participation.

Next, the participants answered a series of 5-point Likert scale items (1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree) about their willingness to use different specimen collection modalities to test for SARS-CoV-2 infection as part of a research study. Participants who indicated having children aged <18 years in their household were also asked about their willingness to use the same modalities to collect specimens from their child as part of a research study. The modalities included home collection of a saliva sample, home throat swab collection, home finger prick blood collection, drive-through site throat swab collection, clinic throat swab collection, and clinic blood collection. The questions indicated that all specimens collected at home would be mailed to a central laboratory for testing. The definitions provided to participants for each testing modality are reported in Table 1.

Finally, the participants were asked how the availability of a home specimen collection method to test for COVID-19 that used either a saliva sample or throat swab would impact their willingness to participate in a research study compared to a drive-through sample collection site and a clinic sample site. For these questions, possible answers included “more likely to participate in a research study,” “about the same likelihood to participate in a research study,” and “less likely to participate in a research study.”

Table 1. Definitions of specimen collection testing modalities used in a web-based survey to assess willingness to participate in a COVID-19 research study in the United States in March 2020.

Testing modality	Survey definition
Home saliva sample	A home saliva sample would involve you spitting in a tube and sending it to a certified laboratory.
Home throat swab	A home throat swab would involve you using a throat swab and sending it to a certified laboratory.
Home blood collection	A home blood test would involve using an automated finger prick device, collecting a blood sample on a specimen card, and mailing in a prepaid mailer to a certified laboratory.
Drive-through site throat swab	A drive-through site for throat swab would involve your traveling to a drive-through facility in your car to have a health care worker collect the swab.
Clinic throat swab	A laboratory throat swab would involve your traveling to a laboratory facility in a clinic or private laboratory to have a health care worker collect the swab.
Clinic blood collection	A laboratory blood test would involve your traveling to a laboratory facility in a clinic or a private lab to have blood drawn, similar to a usual doctor's visit.

Statistical Analysis

All analysis was performed using RStudio v1.1.453. To present a complete description of these data, we summarized the participants' willingness to use each testing modality by calculating both the mean (SD) and median (IQR). A stigma index score was calculated by summing the number of stigma-related items the participant indicated as true (maximum=4). Similarly, we calculated a knowledge index score by tabulating the number of correct responses to the knowledge items (maximum=14). In the methodological literature, there is an ongoing debate about whether parametric or nonparametric statistical methods should be used for Likert-type data [12,13]; therefore, we explored the data with both methods. First, we used nonparametric Kruskal-Wallis tests to assess if willingness to use each testing modality differed by categorical demographic variables and nonparametric Spearman rank correlation coefficients to assess if willingness differed by ordinal characteristics (eg, income, education, likelihood of currently having COVID-19, stigma index score, knowledge index score, and number of symptoms in the past 24 hours). Second, we used parametric statistical methods to facilitate interpretation of the main findings, using Cohen *d* to estimate the effect size of the overall willingness to use each testing modality. Cohen *d* reports the estimated difference in mean values in terms of SD [14]. For example, a Cohen *d* of 1 indicates that the mean of one group is one standard deviation away from than the mean of the comparison group. All *P* values were adjusted for multiple tests using the Bonferroni-Holm method.

Results

A total of 4593 respondents started the eligibility screener. Of these, 12 (0.3%) were removed for duplicate IP addresses, 1260 (27.4%) did not meet the eligibility criteria, and 1886 (41.1%) failed to complete the primary outcome survey questions, resulting in an analytic dataset of 1435 (31.2%) survey responses. The demographic characteristics are summarized in Table 2. Over half the participants (761/1435, 53.0%) were female, and the mean age was 40.0 years (SD 18.2). Many participants were non-Hispanic White (587/1435, 40.9%) or Hispanic (548/1435, 38.2%), and most had either completed a college degree (629/1435, 43.8%) or attended some college, associate degree, or technical school (382/1435, 26.6%). Over one-quarter of respondents (385/1435, 26.8%) reported children aged <18 years in their household and answered survey questions about their willingness to use different specimen collection modalities for SARS-CoV-2 testing to collect specimens from their children.

Figure 1 displays the participants' stated willingness to use different specimen collection modalities for SARS-CoV-2 testing on themselves and their children. Overall, the large majority of participants agreed or strongly agreed that they would be willing to use a home specimen collection method to obtain a saliva sample (1259/1435, 87.7%) or a throat swab (1191/1435, 83.1%) from themselves as part of a research study. More than half the participants agreed or strongly agreed that they would be willing to acquire a home specimen collection finger prick blood sample (928/1435, 64.7%), visit a drive-through site to provide a throat swab (914/1435, 63.7%), or visit a clinic to provide a blood sample (812/1435, 56.6%) or a throat swab (762/1435, 53.1%). In a separate question about relative preference between multiple specimen collection modalities, 990/1435 participants (69.0%) said they would be more likely to participate in a research study if they could collect a saliva sample or throat swab at home compared to going to a drive-through site. Similarly, 1023/1435 participants (71.3%) stated that they would be more likely to participate in a research study with specimens to be collected at home compared to a study with specimens collected at a clinic. Of the 1435 participants, only 63 (4.4%) and 82 (5.7%) reported that using a home specimen collection process would make them less likely to participate in a research study compared to sample collection at a drive-through site or a clinic, respectively.

Relative to the participants' willingness to participate in research themselves, their willingness to have their children participate in research was lower for each specimen collection modality (Figure 1). The proportion of participants willing to use each modality to collect specimens from their children ranged from 291/385 (75.6%) who were willing to perform home collection of a saliva sample to 124/334 (37.1%) who were willing to take their child to a clinic for a blood sample. However, the ranked orders of the participants' willingness to use each testing modality on themselves and on their children were similar.

For most comparisons, the stated willingness to use each specimen collection modality did not differ by demographic group, stigma index score, or presence of current COVID-19 symptoms (Table 2, Table 3, and Table 4). A notable exception was that younger participants were slightly less willing to obtain a home-collected throat swab (adjusted *P*=.049) or visit a drive-through site to provide a throat swab (adjusted *P*=.047). While there was no difference in willingness to use home collection saliva samples or throat swabs, participants who thought it was somewhat likely, likely, or very likely that they currently had COVID-19 had moderately higher willingness to visit a drive-through site (adjusted *P*=.01) or a clinic (adjusted *P*=.003) to provide a throat swab.

Table 2. Demographic characteristics of the study participants (internet-using adults aged >18 years in the United States in March 2020) and their stated willingness to use home saliva sample and throat swab specimen collection testing modalities on themselves as part of a COVID-19 research study (N=1435). All survey questions were 5-point Likert scale items where 1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree. Kruskal-Wallis tests and Spearman rank correlation coefficients were used to assess response differences by characteristic.

Characteristic	n (%)	Home: saliva sample			Home: throat swab		
		Mean (SD)	Median (IQR)	*P* value[a]	Mean (SD)	Median (IQR)	*P* value
Overall	1435 (100.0)	4.5 (0.9)	5 (4-5)	N/A[b]	4.4 (1.9)	5 (4-5)	N/A
Gender				>.99			>.99
Female	761 (53.0)	4.5 (0.9)	5 (4-5)		4.3 (1.0)	5 (4-5)	
Male	536 (37.4)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Other	36 (2.5)	4.6 (1.0)	5 (5-5)		4.5 (1.0)	5 (4.75-5)	
Age (years)				.14			.049
18-29	560 (39.0)	4.4 (1.0)	5 (4-5)		4.2 (1.1)	5 (4-5)	
30-49	391 (27.2)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
50-64	289 (20.1)	4.6 (0.7)	5 (4-5)		4.5 (0.9)	5 (4-5)	
≥65	194 (13.5)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Race/ethnicity				.06			>.99
Hispanic	548 (38.2)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Asian/Pacific Islander	52 (3.6)	4.6 (0.7)	5 (4-5)		4.5 (0.8)	5 (4-5)	
Non-Hispanic Black	158 (11.0)	4.3 (1.0)	5 (4-5)		4.2 (1.1)	5 (4-5)	
Non-Hispanic White	587 (40.9)	4.6 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Other	90 (6.2)	4.3 (1.1)	5 (4-5)		4.1 (1.3)	5 (4-5)	
Education				>.99			>.99
College, postgraduate, or professional school	629 (43.8)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Some college, associate's degree, or technical school	382 (26.6)	4.6 (0.8)	5 (4-5)		4.5 (0.9)	5 (4-5)	
High school/GED[c]	175 (12.2)	4.4 (1.0)	5 (4-5)		4.3 (1.1)	5 (4-5)	
Did not finish high school	27 (1.9)	4.6 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Annual income (US $)				.44			>.99
<24,000	294 (20.5)	4.5 (0.9)	5 (4-5)		4.3 (1.1)	5 (4-5)	
24,000 to <50,000	276 (19.2)	4.5 (1.0)	5 (4-5)		4.4 (1.1)	5 (4-5)	
50,000 to <75,000	203 (14.1)	4.6 (0.9)	5 (4-5)		4.5 (0.9)	5 (4-5)	
≥75,000	268 (18.7)	4.7 (0.8)	5 (5-5)		4.5 (0.9)	5 (4-5)	
Don't know	91 (6.3)	4.5 (0.8)	5 (4-5)		4.3 (0.9)	5 (4-5)	
How likely do you think it is you have COVID-19 now?				>.99			>.99
Very unlikely	356 (24.8)	4.4 (1.1)	5 (4-5)		4.3 (1.2)	5 (4-5)	
Unlikely	661 (46.1)	4.5 (0.9)	5 (4-5)		4.3 (1.0)	5 (4-5)	
Somewhat likely	324 (22.6)	4.5 (0.8)	5 (4-5)		4.5 (0.9)	5 (4-5)	
Likely/very likely	81 (5.6)	4.5 (0.8)	5 (4-5)		4.5 (0.9)	5 (4-5)	
Stigma index score				.70			>.99
0	722 (50.3)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
1-2	525 (36.6)	4.5 (1.0)	5 (4-5)		4.4 (1.0)	5 (4-5)	
≥3	106 (7.4)	4.3 (1.1)	5 (4-5)		4.3 (1.2)	5 (4-5)	
Knowledge index score				.02			>.99

Characteristic	n (%)	Home: saliva sample			Home: throat swab		
		Mean (SD)	Median (IQR)	*P* value[a]	Mean (SD)	Median (IQR)	*P* value
<12	337 (23.5)	4.4 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
12-13	655 (45.6)	4.5 (1.0)	5 (4-5)		4.4 (1.1)	5 (4-5)	
14	342 (23.8)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
Symptoms				.71			>.99
1 or more symptoms	747 (52.1)	4.5 (0.9)	5 (4-5)		4.4 (1.0)	5 (4-5)	
None	688 (47.9)	4.4 (1.0)	5 (4-5)		4.3 (1.1)	5 (4-5)	

[a]*P* values were adjusted for multiple comparisons using the Bonferroni-Holm method.

[b]N/A: not applicable.

[c]GED: General Education Development.

Figure 1. Stated willingness to use testing modalities as part of a COVID-19 research study by internet-using adults aged ≥18 years in the United States in March 2020 on themselves (top) and on their children (bottom). All specimen collection modalities involved testing the specimens in a central laboratory.

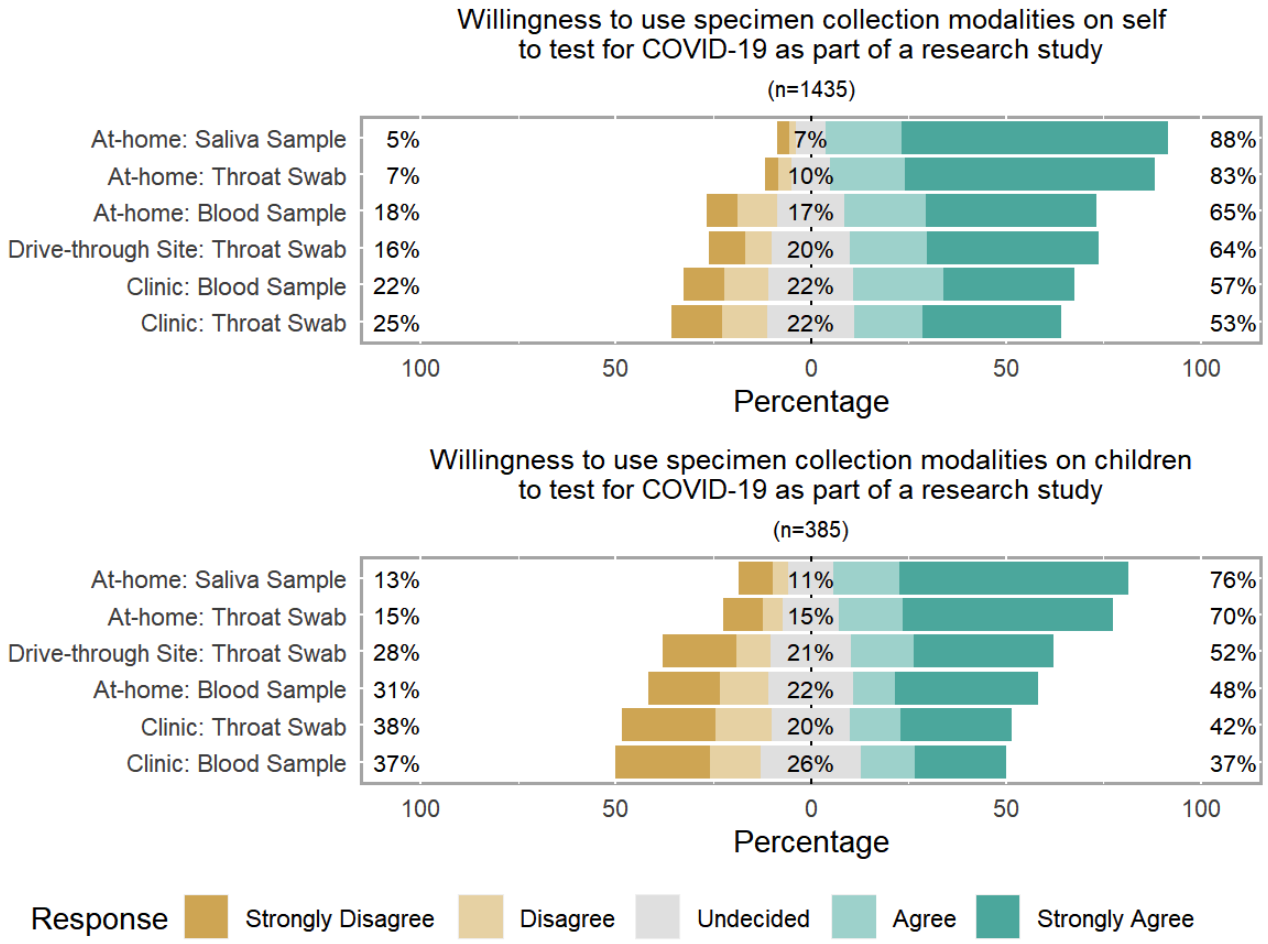


Table 3. Stated willingness of internet-using adults aged ≥18 years in the United States in March 2020 to use drive-through and clinic throat swab specimen collection testing modalities on themselves as part of a COVID-19 research study (N=1435). All survey questions were 5-point Likert scale items where 1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree. Kruskal-Wallis tests and Spearman rank correlation coefficients were used to assess response differences by characteristic.

Characteristic	Drive-through site: throat swab			Clinic: throat swab		
	Mean (SD)	Median (IQR)	*P* value[a]	Mean (SD)	Median (IQR)	*P* value
Overall	3.8 (1.3)	4 (3-5)	N/A[b]	3.5 (1.4)	4 (3-5)	N/A
Gender			>.99			.57
Female	3.8 (1.3)	4 (3-5)		3.4 (1.4)	4 (2-5)	
Male	3.9 (1.3)	4 (3-5)		3.6 (1.4)	4 (3-5)	
Other	3.7 (1.4)	4 (3-5)		3.6 (1.2)	3.5 (3-5)	
Age (years)			.047			>.99
18-29	3.7 (1.4)	4 (3-5)		3.5 (1.4)	4 (3-5)	
30-49	3.9 (1.3)	4 (3-5)		3.5 (1.4)	4 (3-5)	
50-64	4 (1.3)	4 (3-5)		3.4 (1.4)	4 (2-5)	
≥65	3.9 (1.2)	4 (3-5)		3.6 (1.3)	4 (3-5)	
Race/ethnicity			>.99			.99
Hispanic	3.9 (1.3)	4 (3-5)		3.6 (1.4)	4 (3-5)	
Asian/Pacific Islander	3.7 (1.4)	4 (3-5)		3.6 (1.4)	4 (2.75-5)	
Non-Hispanic Black	3.7 (1.4)	4 (3-5)		3.5 (1.5)	4 (2-5)	
Non-Hispanic White	3.9 (1.3)	4 (3-5)		3.4 (1.4)	3 (2-5)	
Other	3.6 (1.4)	4 (3-5)		3.4 (1.4)	3 (2.25-5)	
Education			.69			>.99
College, postgraduate, or professional school	3.8 (1.3)	4 (3-5)		3.4 (1.4)	4 (2-5)	
Some college, associate's degree, or technical school	4 (1.2)	5 (3-5)		3.6 (1.4)	4 (3-5)	
High school/GED[c]	3.6 (1.4)	4 (3-5)		3.5 (1.4)	4 (3-5)	
Did not finish high school	3.8 (1.3)	4 (3-5)		3.5 (1.5)	4 (2.5-5)	
Annual income (US $)			>.99			.82
<24,000	3.8 (1.3)	4 (3-5)		3.7 (1.4)	4 (3-5)	
24,000 to <50,000	3.9 (1.3)	4 (3-5)		3.5 (1.4)	4 (3-5)	
50,000 to <75,000	3.9 (1.2)	4 (3-5)		3.5 (1.4)	4 (3-5)	
≥75,000	3.9 (1.3)	5 (3-5)		3.4 (1.5)	3 (2-5)	
Don't know	3.6 (1.3)	4 (3-5)		3.4 (1.3)	3 (3-5)	
How likely do you think it is you have COVID-19 now?			.001			.003
Very unlikely	3.6 (1.5)	4 (3-5)		3.3 (1.5)	3 (2-5)	
Unlikely	3.8 (1.3)	4 (3-5)		3.5 (1.4)	4 (3-5)	
Somewhat likely	4 (1.2)	4.5 (3-5)		3.7 (1.3)	4 (3-5)	
Likely/very likely	4 (1.3)	5 (3-5)		3.7 (1.4)	4 (3-5)	
Stigma index score			>.99			.70
0	3.9 (1.2)	4 (3-5)		3.6 (1.3)	4 (3-5)	
1-2	3.8 (1.4)	4 (3-5)		3.4 (1.4)	3 (2-5)	
≥3	3.7 (1.4)	4 (3-5)		3.5 (1.5)	4 (2-5)	
Knowledge index score			>.99			>.99

Characteristic		Drive-through site: throat swab			Clinic: throat swab		
		Mean (SD)	Median (IQR)	*P* value[a]	Mean (SD)	Median (IQR)	*P* value
	<12	3.9 (1.3)	4 (3-5)		3.6 (1.4)	4 (3-5)	
	12-13	3.8 (1.3)	4 (3-5)		3.5 (1.4)	4 (2-5)	
	14	3.8 (1.3)	4 (3-5)		3.4 (1.4)	4 (2-5)	
Symptoms				.21			.65
	1 or more symptoms	3.9 (1.3)	4 (3-5)		3.6 (1.4)	4 (3-5)	
	None	3.7 (1.4)	4 (3-5)		3.4 (1.4)	4 (2-5)	

[a]All *P* values were adjusted for multiple comparisons using the Bonferroni-Holm method.

[b]N/A: not applicable.

[c]GED: General Education Development.

Table 4. Stated willingness of internet-using adults aged ≥18 years in the United States in March 2020 to use home and clinic blood sample specimen collection testing modalities on themselves as part of a COVID-19 research study (N=1435). All survey questions were 5-point Likert scale items where 1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree. Kruskal-Wallis tests and Spearman rank correlation coefficients were used to assess response differences by characteristic.

Characteristic	Home: blood sample			Clinic: blood sample		
	Mean (SD)	Median (IQR)	*P* value[a]	Mean (SD)	Median (IQR)	*P* value
Overall	3.8 (1.3)	4 (3-5)	N/A[b]	3.6 (1.3)	4 (3-5)	N/A
Gender			>0.99			>0.99
Female	3.8 (1.3)	4 (3-5)		3.5 (1.3)	4 (3-5)	
Male	3.9 (1.3)	4 (3-5)		3.7 (1.3)	4 (3-5)	
Other	4.1 (1.3)	5 (4-5)		3.6 (1.2)	4 (3-5)	
Age (years)			<0.001			>0.99
18-29	3.5 (1.4)	4 (2-5)		3.5 (1.4)	4 (3-5)	
30-49	4.0 (1.2)	4 (3-5)		3.7 (1.3)	4 (3-5)	
50-64	4.1 (1.1)	5 (3-5)		3.6 (1.3)	4 (3-5)	
≥65	4.0 (1.2)	4 (3-5)		3.7 (1.3)	4 (3-5)	
Race/ethnicity			>0.99			>0.99
Hispanic	3.6 (1.4)	4 (3-5)		3.7 (1.2)	4 (3-5)	
Asian/Pacific Islander	3.8 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
Non-Hispanic Black	3.6 (1.4)	4 (3-5)		3.5 (1.4)	4 (2-5)	
Non-Hispanic White	3.9 (1.3)	4 (3-5)		3.5 (1.3)	4 (3-5)	
Other	3.7 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
Education			>0.99			>0.99
College, postgraduate, or professional school	4.0 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
Some college, associate's degree, or technical school	3.9 (1.2)	4 (3-5)		3.4 (1.2)	4 (3-4)	
High school/GED[c]	3.7 (1.3)	4 (3-5)		3.5 (1.4)	4 (2-5)	
Did not finish high school	3.9 (1.2)	4 (3-5)		3.7 (1.3)	4 (3-5)	
Annual income (US $)			0.01			0.65
<24,000	3.9 (1.3)	4 (3-5)		3.7 (1.2)	4 (3-5)	
24,000 to <50,000	4.0 (1.2)	4 (3-5)		3.5 (1.3)	4 (3-5)	
50,000 to <75,000	3.8 (1.3)	4 (3-5)		3.7 (1.3)	4 (3-5)	
≥75,000	4.2 (1.1)	5 (4-5)		3.7 (1.3)	4 (3-5)	
Don't know	3.5 (1.4)	4 (3-5)		3.5 (1.3)	3 (3-5)	
How likely do you think it is you have COVID-19 now?			>0.99			>0.99
Very unlikely	3.9 (1.3)	4 (3-5)		3.4 (1.4)	4 (2-5)	
Unlikely	3.8 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
Somewhat likely	3.8 (1.3)	4 (3-5)		3.7 (1.2)	4 (3-5)	
Likely/very likely	3.5 (1.5)	4 (3-5)		3.6 (1.5)	4 (3-5)	
Stigma index score			>0.99			>0.99
0	3.9 (1.3)	4 (3-5)		3.7 (1.3)	4 (3-5)	
1-2	3.8 (1.4)	4 (3-5)		3.5 (1.4)	4 (2-5)	
≥3	3.7 (1.4)	4 (2-5)		3.4 (1.4)	4 (2-5)	
Knowledge index score			>0.99			>0.99

Characteristic	Home: blood sample			Clinic: blood sample		
	Mean (SD)	Median (IQR)	*P* value[a]	Mean (SD)	Median (IQR)	*P* value
<12	3.8 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
12-13	3.9 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
14	3.8 (1.3)	4 (3-5)		3.5 (1.4)	4 (3-5)	
Symptoms			>0.99			>0.99
1 or more symptoms	3.8 (1.3)	4 (3-5)		3.6 (1.3)	4 (3-5)	
None	3.9 (1.3)	4 (3-5)		3.5 (1.3)	4 (3-5)	

[a]*P* values were adjusted for multiple comparisons using the Bonferroni-Holm method.

[b]N/A: not applicable.

[c]GED: General Education Development.

The mean willingness rating (Likert 5-point scale) ranged from 3.5 (clinic throat swab, median 4) to 4.5 (home saliva sample, median 5; Table 5). Compared to the participants' willingness to use a home saliva sample, there was a medium effect size in willingness to use a home test blood sample (Cohen *d*=0.568; 95% CI 0.510-0.627) and willingness to use a drive-through throat swab (Cohen *d*=0.567; 95% CI 0.507-0.627). There was a large effect size in willingness to use a clinic for either a throat swab (Cohen *d*=0.802; 95% CI 0.732-0.872) or a blood sample (Cohen *d*=0.776; 95% CI 0.706-0.847) compared to using a home test saliva sample. A similar pattern was seen in comparisons between willingness to use different testing modalities for children.

Table 5. Stated willingness of internet-using adults aged ≥18 years in the United States in March 2020 to use specimen collection modalities on themselves and their children as part of a COVID-19 research study and relative effect sizes. All survey questions were 5-point Likert scale items where 1=strongly disagree, 2=disagree, 3=undecided, 4=agree, and 5=strongly agree. All *P*<.001.

Specimen collection modality	n (%)	Mean (SD)	Median (IQR)	Cohen *d*	95% CI
Willing to use on oneself (N=1435)					
Home: saliva sample	1435 (100.0)	4.5 (0.9)	5 (4-5)	Reference	N/A[a]
Home: throat swab	1435 (100.0)	4.4 (1.9)	5 (4-5)	0.114	0.085 to 0.144
Home: blood sample	1434 (99.9)	3.8 (1.3)	4 (3-5)	0.568	0.510 to 0.627
Drive-through site: throat swab	1435 (100.0)	3.8 (1.3)	4 (3-5)	0.567	0.507 to 0.627
Clinic: throat swab	1435 (100.0)	3.5 (1.4)	4 (3-5)	0.802	0.732 to 0.872
Clinic: blood sample	1434 (99.9)	3.6 (1.3)	4 (3-5)	0.776	0.706 to 0.847
Willing to use on one's children (n=385)					
Home: saliva sample	385 (100.0)	4.1 (1.3)	5 (4-5)	Reference	N/A
Home: throat swab	385 (100.0)	4.0 (1.3)	5 (3-5)	0.113	–0.029 to 0.254
Home: blood sample	334 (87.8)	3.4 (1.5)	3 (2-5)	0.454	0.388 to 0.520
Drive-through site: throat swab	385 (100.0)	3.4 (1.5)	4 (2-5)	0.517	0.373 to 0.660
Clinic: throat swab	385 (100.0)	3.1 (1.5)	3 (2-5)	0.742	0.596 to 0.889
Clinic: blood sample	334 (87.8)	3.0 (1.5)	3 (2-4)	0.851	0.768 to 0.933

[a]N/A: not applicable.

Discussion

Principal Findings

Response to the SARS-CoV-2 epidemic in the United States has been hampered by insufficient testing both for diagnosing persons and for public health assessments to describe the epidemiology of infection. Critical shortages of reagents, other supply chain issues, and lack of availability of health care workers have led to gaps in testing, and it is critical to diversify testing methods. Options include alternative specimens to be tested and specimen collection locations. Further, it is important to understand which testing options are best suited to which purposes (eg, clinical care, population research, and screening versus diagnosis). Self-collection of specimens at home has proven to be an acceptable approach in other infectious disease testing, and it could play an important role in the response to the SARS-CoV-2 epidemic in the United States.

Results from this study indicate that a large majority of adults would be willing to participate in a research study about SARS-CoV-2 infection or immune experience by collecting

specimens at home and mailing them to a laboratory for testing. There was some preference for certain specimen collection modalities (saliva samples and throat swabs were preferred over blood samples); however, these differences were largely driven by preference for remote home specimen collection methods versus methods that would require visits to clinical or laboratory locations for testing. Testing location and specimen collection preferences were consistent across demographic groups and other characteristics of interest. These results are similar to literature reports that indicate that home specimen collection options are preferred to clinic-based testing methods for HIV and STI screening [11,15,16].

These findings are important for the design of forthcoming epidemiologic studies of SARS-CoV-2 infection. Cross-sectional and cohort follow-up studies will depend on testing biological specimens to accurately measure disease prevalence, incidence, and recovery among participants. Our results indicate that study designs that use home specimen collection experience increased participation and higher retention compared to study designs that involve traveling to a drive-through site or clinic. The high willingness to use home specimen collection methods across demographic groups and other subgroups suggests that studies incorporating home specimen collection may be less susceptible to participation bias than designs requiring the collection of biological specimens in clinical settings.

The use of home specimen collection methods can also help ensure that research activities do not have a negative effect on adherence to current health guidance. Home specimen collection can be incorporated within the context text of social distancing guidelines [17], which can enable study participants to maximize their individual contributions to slowing the spread of COVID-19 disease. Similarly, reducing the contact between study participants and health care workers improves health care worker safety by reducing their risk of exposure. Further, such options can reduce the overall burden on health care providers and clinics that may not have the capacity to collect specimens as part of ongoing research. For instance, the current shortage of personal protective equipment among health care workers in the United States has been well documented [18], and the incorporation of home specimen collection can ensure that these resources are preserved for use in clinical care.

Willingness of people to use home specimen collection kits for research offers promising opportunities to conduct representative and timely research about SARS-CoV-2 infection and immune response. However, it is important to recognize that the collection of home specimens will also require rigorous testing to ensure that the laboratory results obtained provide accurate indications of SARS-CoV-2 infection and immune response, the kits are safe for participants to use, and the specimens are sufficiently robust to maintain validity after the process of return shipping to laboratories for analysis. It is important to distinguish between home collection of specimens that are mailed back to laboratories for analysis from the separate field of home testing, in which participants collect their own specimen, apply it to a test device, and interpret the results at home.

One reason that the use of home specimen collection may be associated with high willingness to participate in research related to SARS-CoV-2 infection is that the virus is highly infectious, and there is meaningful risk associated with entering clinical settings where people with symptoms of COVID-19 are congregated. Unlike testing as part of research studies for less contagious infectious diseases or for other types of disease, potential research participants who do not have symptoms may be particularly reluctant to report to clinical locations for screening. Relatedly, if home collection of specimens is developed and validated, it will be possible to conduct epidemiological studies that can both reach people in diverse geographic areas, including rural areas, and allow research to be conducted without exposing participants to potential harms associated with going to research sites that may result in their exposure to SARS-CoV-2.

Limitations

Our study results have several limitations. First, we assessed willingness using Likert scale items, which limits the ability to determine the magnitude of preference for one test modality over the other for any individual. However, Likert data are especially well suited for assessing the direction of preference, which is a relevant outcome given our desire to understand the potential impact of offering home specimen collection in research settings. Second, our recruitment methods targeted social media users, and our convenience sample may not be representative of all US adults. Third, opinions regarding testing for SARS-CoV-2 may change over time, given the rapidly shifting nature of public perception regarding the epidemic, and updates are merited to ensure that participant preferences remain stable. Finally, we know that there have historically been disconnects between expressed willingness to use self-testing or at-home specimen collection options for infectious diseases and the actual uptake of these highly acceptable devices [19]. However, compared to historical examples of the introduction of at-home tests before the advent of telemedicine, testing for SARS-CoV-2 infection with specimens collected at home for research or clinical purposes may be more acceptable given the broad availability of telehealth clinical services. These services may be used to provide support for participants who have questions about collecting specimens at home or to observe the self-collection of specimens at home until data are developed to document the sufficiency and quality of specimens collected at home.

Conclusions

Large scale population-based research and testing is needed to provide the epidemiologic data necessary to guide our public health response to the COVID-19 pandemic. Home specimen collection strategies should be considered to achieve the highest levels of participant engagement and retention, reduce the burden of specimen collection in overloaded health care settings, and reduce potential exposure of research participants to SARS-CoV-2 in research settings.

Acknowledgments

This work was supported by the US National Institute of Allergy and Infectious Diseases (3R01AI143875-02S1) and the Emory Center for AIDS Research (P30AI050409). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936. [doi: 10.1056/nejmoa2001191]
2. Coronavirus Disease 2019 (COVID-19): Cases in the U.S. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed 2020-04-08]
3. Our most up-to-date data and annotations. The COVID Tracking Project. URL: https://covidtracking.com/data [accessed 2020-04-08]
4. Guest J, Del Rio C, Sanchez T. The Three Steps Needed to End the COVID-19 Pandemic: Bold Public Health Leadership, Rapid Innovations, and Courageous Political Will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
5. Spielberg F, Levine RO, Weaver M. Self-testing for HIV: a new option for HIV prevention? Lancet Infect Dis 2004 Oct;4(10):640-646. [doi: 10.1016/s1473-3099(04)01150-8]
6. Wiesenfeld HC, Lowry DL, Heine RP, Krohn MA, Bittner H, Kellinger K, et al. Self-collection of vaginal swabs for the detection of Chlamydia, gonorrhea, and trichomoniasis: opportunity to encourage sexually transmitted disease testing among adolescents. Sex Transm Dis 2001 Jun;28(6):321-325. [doi: 10.1097/00007435-200106000-00003] [Medline: 11403188]
7. Gilbert M, Hottes TS, Kerr T, Taylor D, Fairley CK, Lester R, et al. Factors associated with intention to use internet-based testing for sexually transmitted infections among men who have sex with men. J Med Internet Res 2013 Nov 14;15(11):e254 [FREE Full text] [doi: 10.2196/jmir.2888] [Medline: 24240644]
8. Hottes TS, Farrell J, Bondyra M, Haag D, Shoveller J, Gilbert M. Internet-based HIV and sexually transmitted infection testing in British Columbia, Canada: opinions and expectations of prospective clients. J Med Internet Res 2012 Mar 06;14(2):e41 [FREE Full text] [doi: 10.2196/jmir.1948] [Medline: 22394997]
9. Figueroa C, Johnson C, Ford N, Sands A, Dalal S, Meurant R, et al. Reliability of HIV rapid diagnostic tests for self-testing compared with testing by health-care workers: a systematic review and meta-analysis. Lancet HIV 2018 Jun;5(6):e277-e290. [doi: 10.1016/s2352-3018(18)30044-4]
10. Lunny C, Taylor D, Hoang L, Wong T, Gilbert M, Lester R, et al. Self-Collected versus Clinician-Collected Sampling for Chlamydia and Gonorrhea Screening: A Systemic Review and Meta-Analysis. PLoS One 2015;10(7):e0132776 [FREE Full text] [doi: 10.1371/journal.pone.0132776] [Medline: 26168051]
11. Siegler AJ, Mayer KH, Liu AY, Patel RR, Ahlschlager LM, Kraft CS, et al. Developing and Assessing the Feasibility of a Home-based Preexposure Prophylaxis Monitoring and Support Program. Clin Infect Dis 2019 Jan 18;68(3):501-504 [FREE Full text] [doi: 10.1093/cid/ciy529] [Medline: 29982304]
12. Sullivan GM, Artino AR. Analyzing and Interpreting Data From Likert-Type Scales. J Grad Med Educ 2013 Dec;5(4):541-542. [doi: 10.4300/jgme-5-4-18]
13. Carifio J, Perla R. Resolving the 50-year debate around using and misusing Likert scales. Med Educ 2008 Dec;42(12):1150-1152. [doi: 10.1111/j.1365-2923.2008.03172.x] [Medline: 19120943]
14. Cohen J. Statistical Power Analysis for the Behavioral Sciences 2nd Edition. New York, NY: Routledge; 2013.
15. Pearson WS, Kreisel K, Peterman TA, Zlotorzynska M, Dittus PJ, Habel MA, et al. Improving STD service delivery: Would American patients and providers use self-tests for gonorrhea and chlamydia? Prev Med 2018 Oct;115:26-30. [doi: 10.1016/j.ypmed.2018.08.007] [Medline: 30096329]
16. Odesanmi TY, Wasti SP, Odesanmi OS, Adegbola O, Oguntuase OO, Mahmood S. Comparative effectiveness and acceptability of home-based and clinic-based sampling methods for sexually transmissible infections screening in females aged 14–50 years: a systematic review and meta-analysis. Sex Health 2013;10(6):559. [doi: 10.1071/sh13029]
17. Coronavirus Disease (COVID-19): Social Distancing, Quarantine, and Isolation. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html [accessed 2020-04-12]
18. Kamerow D. Covid-19: the crisis of personal protective equipment in the US. BMJ 2020 Apr 03;369:m1367. [doi: 10.1136/bmj.m1367] [Medline: 32245847]
19. Colfax GN, Lehman JS, Bindman AB, Vittinghoff E, Vranizan K, Fleming PL, Multisite Evaluation of Surveillance of HIV (MESH) Study Group. What happened to home HIV test collection kits? Intent to use kits, actual use, and barriers to use among persons at risk for HIV infection. AIDS Care 2002 Oct;14(5):675-682. [doi: 10.1080/0954012021000005533a] [Medline: 12419117]

Abbreviations

STI: sexually transmitted infection

Edited by G Eysenbach; submitted 19.04.20; peer-reviewed by S Badawy, C Figueroa, R McGowan; comments to author 20.06.20; revised version received 01.07.20; accepted 11.08.20; published 03.09.20.

Please cite as:
Hall EW, Luisi N, Zlotorzynska M, Wilde G, Sullivan P, Sanchez T, Bradley H, Siegler AJ
Willingness to Use Home Collection Methods to Provide Specimens for SARS-CoV-2/COVID-19 Research: Survey Study
J Med Internet Res 2020;22(9):e19471
URL: https://www.jmir.org/2020/9/e19471
doi:10.2196/19471
PMID:32790639

Original Paper

Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study

Changwon Son[1], BS, MS; Sudeep Hegde[1], BEng, MS, PhD; Alec Smith[1], BS; Xiaomei Wang[1], BS, PhD; Farzan Sasangohar[1,2], BA, BCS, MASc, SM, PhD

[1]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[2]Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, United States

Corresponding Author:
Farzan Sasangohar, BA, BCS, MASc, SM, PhD
Department of Industrial and Systems Engineering
Texas A&M University
3131 TAMU
College Station, TX, 77843
United States
Phone: 1 979 458 2337
Email: sasangohar@tamu.edu

Abstract

Background: Student mental health in higher education has been an increasing concern. The COVID-19 pandemic situation has brought this vulnerable population into renewed focus.

Objective: Our study aims to conduct a timely assessment of the effects of the COVID-19 pandemic on the mental health of college students.

Methods: We conducted interview surveys with 195 students at a large public university in the United States to understand the effects of the pandemic on their mental health and well-being. The data were analyzed through quantitative and qualitative methods.

Results: Of the 195 students, 138 (71%) indicated increased stress and anxiety due to the COVID-19 outbreak. Multiple stressors were identified that contributed to the increased levels of stress, anxiety, and depressive thoughts among students. These included fear and worry about their own health and of their loved ones (177/195, 91% reported negative impacts of the pandemic), difficulty in concentrating (173/195, 89%), disruptions to sleeping patterns (168/195, 86%), decreased social interactions due to physical distancing (167/195, 86%), and increased concerns on academic performance (159/195, 82%). To cope with stress and anxiety, participants have sought support from others and helped themselves by adopting either negative or positive coping mechanisms.

Conclusions: Due to the long-lasting pandemic situation and onerous measures such as lockdown and stay-at-home orders, the COVID-19 pandemic brings negative impacts on higher education. The findings of our study highlight the urgent need to develop interventions and preventive strategies to address the mental health of college students.

(J Med Internet Res 2020;22(9):e21279) doi:10.2196/21279

KEYWORDS

COVID-19; pandemic; college student; mental health; stress; anxiety; self-management

Introduction

Mental health issues are the leading impediment to academic success. Mental illness can affect students' motivation, concentration, and social interactions—crucial factors for students to succeed in higher education [1]. The 2019 Annual Report of the Center for Collegiate Mental Health [2] reported that anxiety continues to be the most common problem (62.7% of 82,685 respondents) among students who completed the Counseling Center Assessment of Psychological Symptoms, with clinicians also reporting that anxiety continues to be the most common diagnosis of the students that seek services at university counseling centers. Consistent with the national trend, Texas A&M University has seen a rise in the number of students seeking services for anxiety disorders over the past 8 years. In 2018, slightly over 50% of students reported anxiety as the main reason for seeking services. Despite the increasing need for mental health care services at postsecondary institutions,

alarmingly, only a small portion of students committing suicide contact their institution counseling centers [3], perhaps due to the stigma associated with mental health. Such negative stigma surrounding mental health diagnosis and care has been found to correlate with a reduction in adherence to treatment and even early termination of treatment [4].

The COVID-19 pandemic has brought into focus the mental health of various affected populations. It is known that the prevalence of epidemics accentuates or creates new stressors including fear and worry for oneself or loved ones, constraints on physical movement and social activities due to quarantine, and sudden and radical lifestyle changes. A recent review of virus outbreaks and pandemics documented stressors such as infection fears, frustration, boredom, inadequate supplies, inadequate information, financial loss, and stigma [5]. Much of the current literature on psychological impacts of COVID-19 has emerged from the earliest hot spots in China. Although several studies have assessed mental health issues during epidemics, most have focused on health workers, patients, children, and the general population [6,7]. For example, a recent poll by The Kaiser Family Foundation showed that 47% of those sheltering in place reported negative mental health effects resulting from worry or stress related to COVID-19 [8]. Nelson et al [9] have found elevated levels of anxiety and depressive symptoms among general population samples in North America and Europe. However, with the exception of a few studies, notably from China [10-12], there is sparse evidence of the psychological or mental health effects of the current pandemic on college students, who are known to be a vulnerable population [13]. Although the findings from these studies thus far converge on the uptick of mental health issues among college students, the contributing factors may not necessarily be generalizable to populations in other countries. As highlighted in multiple recent correspondences, there is an urgent need to assess effects of the current pandemic on the mental health and well-being of college students [14-17].

The aim of this study is to identify major stressors associated with the COVID-19 pandemic and to understand their effects on college students' mental health. This paper documents the findings from online interview surveys conducted in a large university system in Texas.

Methods

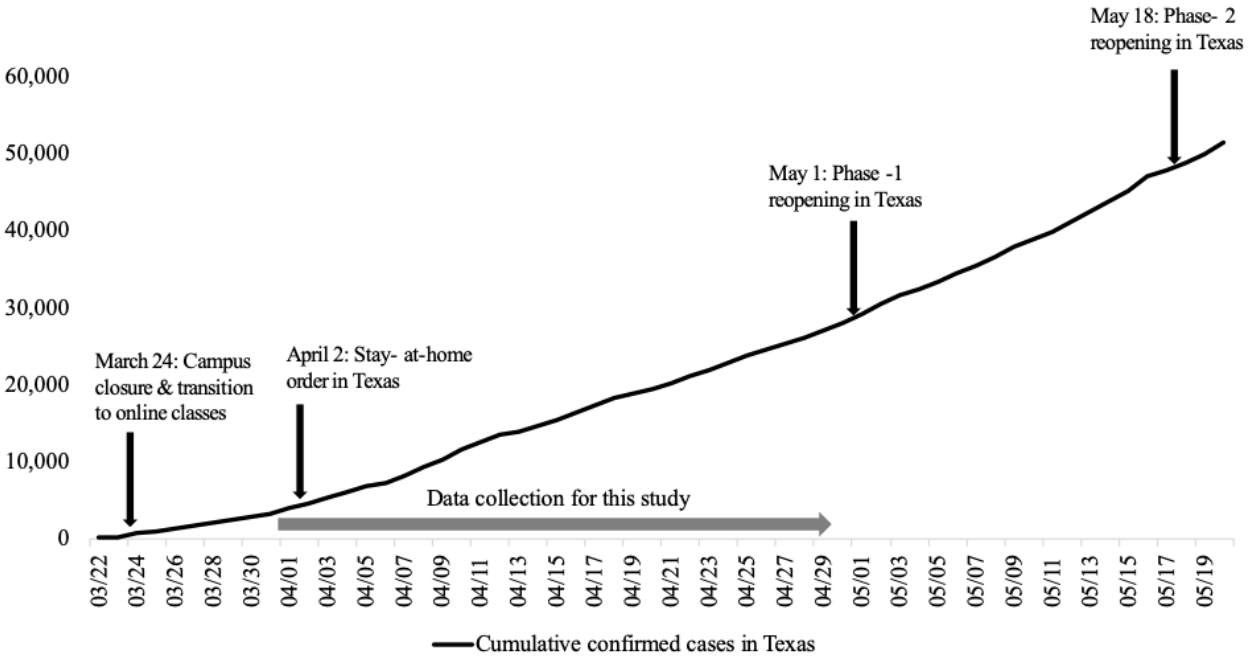
Study Design

A semistructured interview survey guide was designed with the purpose of assessing the mental health status of college students both quantitatively and qualitatively. In addition, the interview aimed to capture the ways that students have been coping with the stress associated with the pandemic situation. First, our study assesses participants' general stress levels using the Perceived Stress Scale-10 (PSS) [18]. PSS is a widely used instrument to measure overall stress in the past month [19]. Second, participants were asked if their own and peers' (two separate questions) stress and anxiety increased, decreased, or remained the same because of the COVID-19 pandemic. For those who indicated increased stress and anxiety during the pandemic, we questioned their stress coping strategies and use of available mental health counseling services. We then elicited pandemic-specific stressors and their manifestations across 12 academic-, health-, and lifestyle-related categories of outcomes such as effects on own or loved ones' health, sleeping habits, eating habits, financial situation, changes to their living environment, academic workload, and social relations. Students were also asked about the impact of COVID-19 on depressive and suicidal thoughts. These constructs were derived from existing literature identifying prominent factors affecting college students' mental health [20,21]. Feedback on the severity of COVID-19's impact on these aspects were elicited using a 4-point scale: 0 (none), 1 (mild), 2 (moderate), and 3 (severe). Participants were asked to elaborate on each response. Third, participants were guided to describe stressors, coping strategies, and barriers to mental health treatment during a typical semester without associating with the COVID-19 pandemic. Although multiple analyses of the collected data are currently under progress, PSS results and the COVID-19–related findings are presented in this paper.

Participants

Participants were recruited from the student population of a large university system in Texas, United States. This particular university closed all their campuses on March 23, 2020, and held all its classes virtually in response to the COVID-19 pandemic. In addition, the state of Texas issued a stay-at-home order on April 2, 2020. Most interviews were conducted about 1 month after the stay-at-home order in April 2020. Figure 1 illustrates the trend of cumulative confirmed cases and a timeline of major events that took place in the university and the state of Texas. Participants were recruited by undergraduate student researchers through email, text messaging, and snowball sampling. The only inclusion criteria for participation was that participants should have been enrolled as undergraduate students in the university at the time of the interviews.

Figure 1. A timeline of major events related to COVID-19 in the university and the state of Texas (source: Texas Department of State Health Services).



Procedures

The interviews were conducted by 20 undergraduate researchers trained in qualitative methods and the use of the interview survey guide described above. None of the authors conducted the interviews. All interviews were conducted via Zoom [22] and were audio recorded. The recordings were later transcribed using Otter.ai [23], an artificial intelligence–based transcription service, and verified for accuracy manually. Prior to the interview, participants were provided an information document about the study approved by the university's Institutional Review Board (No 2019-1341D). Upon verbal consent, participants were asked to respond to a questionnaire about their demographic information such as age, gender, year of college, and program of study before completing the interview. Participation was voluntary and participants were not compensated.

Data Analysis

First, descriptive statistics were compiled to describe participants' demographics (eg, age, gender, academic year, and major) and the distribution of the ratings on PSS-10 survey items. A total PSS score per participant was calculated by first reversing the scores of the positive items (4-7, 9, and 10) and then adding all the ten scores. A mean (SD) PSS score was computed to evaluate the overall level of stress and anxiety among the participants during the COVID-19 pandemic. Second, participants' answers to 12 academic-, health-, and lifestyle-related questions were analyzed to understand relative impacts of the pandemic on various aspects of college students' mental health. Percentages of participants who indicated negative ratings (ie, mild, moderate, or severe influence) on these questions were calculated and ranked in a descending order. Qualitative answers to the 12 stressors and coping strategies were analyzed using thematic analysis [24,25] similar to the deductive coding step in the grounded theory method [26]. A single coder (CS), trained in qualitative analysis methods, analyzed the transcripts and identified themes using an open coding process, which does not use a priori codes or codes created prior to the analysis and places an emphasis on information that can be extracted directly from the data. Following the identification of themes, the coder discussed the codes with two other coders (XW and AS) trained in qualitative analysis and mental health research to resolve discrepancies among related themes and discuss saturation. The coders consisted of two Ph.D. students and one postdoctoral fellow at the same university. MAXQDA (VERBI GmbH) [27] was used as a computer software program to carry out the qualitative analysis.

Results

Participants

Of the 266 university students initially recruited by the undergraduate researchers, 17 retreated and 249 participated in this study. There were 3 graduate students and 51 participants who had missing data points and were excluded, and data from 195 participants were used in the analysis. The average age was 20.7 (SD 1.7) years, and there were more female students (111/195, 57%) than male students (84/195, 43%). Approximately 70% of the participants were junior and senior students. About 60% of the participants were majoring in the college of engineering, which was the largest college in the university population (Table 1). The mean PSS score for the 195 participants was 18.8 (SD 4.9), indicating moderate perceived stress in the month prior to the interview (Table 2).

Table 1. Participants' demographic characteristics.

Variables	Participants (N=195)
Age (years), mean (SD)	20.7 (1.7)
Gender, n (%)	
Male	84 (43.1)
Female	111 (56.9)
Academic year, n (%)	
Freshmen	24 (12.3)
Sophomore	33 (16.9)
Junior	70 (35.9)
Senior	68 (34.9)
Major (college), n (%)	
Agriculture & life science	10 (5.1)
Engineering	117 (60.0)
Liberal arts	20 (10.3)
Architecture	1 (0.5)
Business management	11 (5.6)
Education and human development	12 (6.1)
School of public health	5 (2.5)
Science	5 (2.5)
Veterinary medicine and biomedical sciences	10 (5.1)
Not specified	4 (2.1)

Table 2. Mean score for each of PSS items.

PSS[a] items	Score, mean (SD)
1. In the past month, how often have you felt upset because of something that happened unexpectedly?	2.2 (0.9)
2. In the past month, how often have you felt that you were unable to control the important things in your life?	2.2 (1.0)
3. In the past month, how often have you felt nervous and "stressed"?	2.8 (0.9)
4. In the past month, how often have you dealt successfully with irritating life hassles?	1.5 (0.9)
5. In the past month, how often have you felt that you were effectively coping with important changes that were occurring in your life?	1.5 (0.9)
6. In the past month, how often have you felt confident about your ability to handle your personal problems?	1.3 (0.9)
7. In the past month, how often have you felt that things were going your way?	1.9 (0.8)
8. In the past month, how often have you found that you could not cope with all the things that you needed to do?	1.8 (1.0)
9. In the past month, how often have you been able to control irritations in your life?	1.5 (0.9)
10. In the past month, how often have you felt that you were on top of things?	1.9 (1.0)
Overall PSS scores	18.8 (4.9)

[a]PSS: Perceived Stress Scale-10.

Challenges to College Students' Mental Health During COVID-19

Out of 195 participants, 138 (71%) indicated that their stress and anxiety had increased due to the COVID-19 pandemic, whereas 39 (20%) indicated it remained the same and 18 (9%) mentioned that the stress and anxiety had actually decreased. Among those who perceived increased stress and anxiety, only 10 (5%) used mental health counseling services. A vast majority of the participants (n=189, 97%) presumed that other students were experiencing similar stress and anxiety because of COVID-19. As shown in Figure 2, at least 54% (up to 91% for some categories) of participants indicated negative impacts (either mild, moderate, or severe) of COVID-19 on academic-, health-, and lifestyle-related outcomes. The qualitative analysis yielded two to five themes for each category of outcomes. The

chronic health conditions category was excluded from the qualitative analysis due to insufficient qualitative response. Table 3 presents the description and frequency of the themes and select participant quotes.

Figure 2. Participants' ratings on mental health aspects in an order of negative impacts (mild, moderate, and severe).

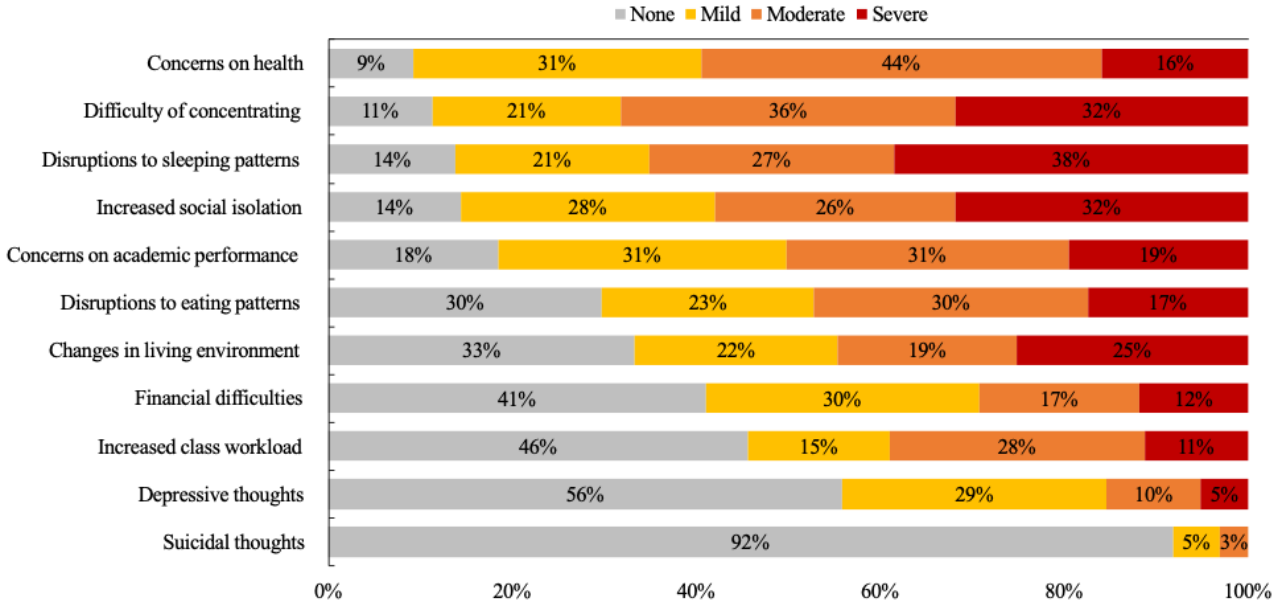


Table 3. Categories and themes of college students' mental health issues and selected participant quotes.

Theme	Participants[a], n (%)	Example quotes[b]
Own health and the health of loved ones (n=177)		
Worry about families and relatives with higher vulnerabilities	76 (43)	• JPP17: "I have a grandma who is affected more by [the virus] and so I'm just afraid that something could happen to her." • SAP16: "My sister just had a baby on Friday and so I've just been worried that her baby or she wouldn't get anything."
Worry about families with more interpersonal contact	26 (15)	• JJP06: "My brother just graduated from a med school and he is doing residency. So, every single patient he sees right now is most likely related to COVID-19." • DDP01: "My mom is actually an essential worker. She works at [a company] warehouse in [a city in Texas]. So, she's coming into contact with people every day."
Worry about themselves being infected	19 (11)	• ACP05: "I always end up having runny nose or just asthma flares up. With this pandemic, the symptoms are very similar to corona so I feel like I would be one of those people who would be highly affected by it."
Difficulty in concentration (n=173)		
Home as a source of distraction	79 (46)	• EGP03: "I'm having difficulty concentrating since I'm home. As I'm around all of my family, it's really hard to focus on what I need to do."
Lack of accountability and motivation	21 (12)	• SAP16: "I just want to lay in my bed. Now no one is keeping me accountable. If I'm on my phone, I'm not paying attention to any of these lectures."
Distracted by social media, internet, and video games	19 (11)	• DDP01: "My desk is right next to my bed so I could just go take a nap or go watch Netflix. Or I could just be on Twitter the whole time and read all the news stories about how people are dying or how bad this is going to get."
Lack of interactive learning environment	18 (10)	• SAP17: "I cannot focus on class when it's online. Through the classes, I don't think there's a lot of interactiveness to make people engaged."
Monotony of life	5 (3)	• DDP07: "Now I'm stuck only doing everything on a computer. So, I'm pretty much on the computer all day."
Sleeping habits (n=168)		
Stay up later or waking up later	84 (50)	• DDP03: "I'll be up until probably four or five in the morning, and sleep through the day usually. Now that most of my classes are online and the lecture isn't mandatory, I sleep through it and I'll watch the lectures later."
Irregular sleep patterns	28 (17)	• SAP03: "I had a really weird sleep schedule now. I stay up really late. And then I wake up very early or sometimes I go to sleep early. I wake up really late. It is just weird."
Increased hours of sleep	12 (7)	• ACP06: "I'm sleeping a lot more now. I'm living at home. I don't have to do anything. I just have more time to sleep."
Difficulty of going/staying asleep	10 (6)	• DDP07: "Now I wake up constantly. I wake up and go to sleep constantly. I have a hard time staying asleep and going asleep."
Social relation/social isolation (n=167)		
Reduced interactions with people	91 (54)	• MBP01: "We're in quarantine so there is significant social isolation from people and from those that I want to hang out with."
Lack of in-person interactions	52 (31)	• JJP02: "I don't see my friends that much and no face to face interaction but only through text."
Restricted outdoor activities	9 (5)	• SNK10: "I also like meeting new people so sometimes I go out climbing or hiking. [COVID-19] has impacted me a lot. I'm not able to do that anymore."
Academic performance (n=159)		
Challenges of online classes	61 (38)	• RMP10: "It's so hard to focus on the lecture because everything is online. And I have to make appointments with a professor or a TA[c]. Then they help me through the Zoom which is online. I think it's hard to have some understanding compared to the face to face meeting."

Theme	Participants[a], n (%)	Example quotes[b]
Impacts on academic progress and future career	36 (23)	• ACP07: "The class I wanted to take over the summer has been canceled, which could potentially push me back a semester." • RMP17: "I think my internship is going to be shortened or cancelled. I need to get more work experience before graduation."
Worry about grades	23 (14)	• ACP12: "Shortly after COVID-19 was declared a pandemic, everything went online. We missed a week of class. So, I had four exams back to back but I didn't transition to online very well. I failed three out of four exams pretty badly. That also got me questioning my entire life and my major."
Reduced motivation or procrastination	12 (8)	• RMP12: "I feel like I started slacking. I was trying to avoid this situation by just not doing some of the work. So, it is stressful academically."
Eating patterns (n=137)		
Increased eating/snacking	35 (26)	• SNK08: "I've been munching a lot on snacks recently since I'm at home."
Inconsistent eating	27 (20)	• SAP02: "I'm home all the time. Sometimes I eat twice a day. Sometimes I don't eat at all. Sometimes it's once a day. It's not something I haven't done before."
Decreased appetite	16 (12)	• SAP15: "I'm having trouble eating. I just don't eat when I'm anxious. So, I've had no appetite."
Emotional eating	7 (5)	• SAP04: "I eat so much now just out of boredom because there's nothing to do really."
Changes in living environment (n=130)		
Changes while staying back home	89 (68)	• YJP05: "I moved back home. So, things are different here. I am having to study now in my bedroom rather than in the library or on campus." • JJP07: "By living with family, you don't have any privacy. You don't feel very focused because you are distracted."
Reduced personal interactions	18 (14)	• ACP02: "I live in the dorm and everybody is moving out so there's basically nobody around me anymore."
Staying longer indoor	9 (7)	• RMP19: "Now I'm at home. I'm literally sitting in the same desk for five or six hours a day."
Financial difficulties (n=115)		
Impacts on current or future employment	44 (38)	• SAP13: "I have rent to pay in [a local town] and I am not sure about my internship this summer. So, I'm going to be basically even in more debt and not unable to pay my bills and my rent."
Impacts on financial situations of families	21 (18)	• ERP03: "My mom has so much that she needs to pay on her own. And she got deduction on her payment, but she still need to pay the same thing. She needs to pay for housing from both mine and my brothers, which is a lot."
Class workload (n=106)		
Catching up with online courses and class projects	51 (48)	• ERP04: "[Professors] still want me to go to a Zoom class. Some of them still record those Zoom meetings and then you can watch it on your own time. It basically doubles the time I have to dedicate each week for that class."
Increased or more difficult assignments	33 (31)	• ERP02: "Four or five out of my six professors have given more work than I would have had if I was there in person. Some of them have to do with participation, just proving that you actually watch the lecture or take notes for the class."
Difficulty of covering the same coursework in shorter time	6 (6)	• NEP04: "A two-week break because of the pandemic made us compress that lost time into our last time we had scheduled."
Depressive thoughts (n=86)		

Theme	Participants[a], n (%)	Example quotes[b]
Loneliness	28 (33)	• MBP02: "I actually suffer from chronic depression. [COVID-19] has definitely made it a lot worse, just being in isolation and being home 24/7. It feels like I need to get out but there's nowhere to go."
Insecurity or uncertainty	10 (12)	• RMP18: "The first couple of days, it was very scary and I think everybody just felt like the world is ending."
Powerlessness or hopelessness	9 (10)	• SNK01: "Maybe [COVID-19] made me really down. Sometimes I feel like I'm incompetent." • SAP20: "It's very easy to fall into a routine of nothingness. And you're seeing no end to this. It's just hopelessness about going back to normal."
Concerns about academic performance	7 (8)	• ACP07: "A lot of hackathons I wanted to go to and a lot of research conferences I wanted to go to have all been shut down. And now it feels like all the work I have been doing for the last few months has been thrown away into the garbage."
Overthinking	4 (5)	• SAP08: "There's just a lot and also you start going crazy in your apartment."
Suicidal thoughts (n=16)		
Linking to depressive thoughts	6 (38)	• JJP03: "[Suicidal thoughts] go hand in hand with depressive thoughts. I am just tired of existing because I am just too hard on myself." • ECP02: "It just has to do with the depressive thoughts and just overthinking. You have a lot of time to think about things that happened in the past like high school. But there's no fixing it. Now, I'm stuck."
Academic issues	1 (6)	• ACP12: "I hate to say it but it comes up on a daily basis. Sometimes as a joke, I want to die. But it's something that I know I have no intention to ever act on and never would like. It's just become incorporated in my life purposely or unconsciously when I do something especially related to academics."
Problems with parents	1 (6)	• SNK09: "I have some problems with my family. And now I'm stuck at home with them. I guess it's more often than normal."
Fear from insecurity	1 (6)	• JPP18: "The biggest thing has been fear of what's next. I think the worst part is more fear of what is to come and what will be the outcome."

[a]Not every participant provided sufficient elaboration to allow for identification of themes, so the frequency of individual themes does not add up to the total number of participants who indicated negative impacts of the COVID-19 outbreak.

[b]The five-digit alphanumeric value indicates the participant ID.

[c]TA: teaching assistant.

Concerns for One's Own Health and the Health of Loved Ones

A vast majority of the participants (177/195, 91%) indicated that COVID-19 increased the level of fear and worry about their own health and the health of their loved ones. Over one-third of those who showed concern (76/177, 43%) were worried about their families and relatives who were more vulnerable, such as older adults, those with existing health problems, and those who are pregnant or gave birth to a child recently. Some of the participants (26/177, 15%) expressed their worry about their family members whose occupation increased their risk of exposure to COVID-19 such as essential and health care workers. Some participants (19/177, 11%) specifically mentioned that they were worried about contracting the virus.

Difficulty With Concentration

A vast majority of participants (173/195, 89%) indicated difficulty in concentrating on academic work due to various sources of distraction. Nearly half of them (79/173, 46%) mentioned that their home is a distractive environment and a more suitable place to relax rather than to study. Participants mentioned that they were more prone to be interrupted by their family members and household chores at home. Other factors affecting students' concentration were lack of accountability (21/173, 12%) and social media, internet, and video games (19/173, 11%). Some (18/173, 10%) stated that online classes were subject to distraction due to lack of interactions and prolonged attention to a computer screen. Additionally, monotonous life patterns were mentioned by some to negatively affect concentration on academic work (5/173, 3%).

Disruption to Sleep Patterns

A majority of participants (168/195, 86%) reported disruptions to their sleep patterns caused by the COVID-19 pandemic, with over one-third (38%) reporting such disruptions as severe. Half of students who reported some disruption (84/168, 50%) stated that they tended to stay up later or wake up later than they did before the COVID-19 outbreak. Another disruptive impact brought by the pandemic was irregular sleep patterns such as inconsistent time to go to bed and to wake up from day to day

(28/168, 17%). Some (12/168, 7%) reported increased hours of sleep, while others (10/168, 6%) had poor sleep quality.

Increased Social Isolation

A majority of participants answered that the pandemic has increased the level of social isolation (167/195, 86%). Over half of these students (91/167, 54%) indicated that their overall interactions with other people such as friends had decreased significantly. In particular, about one-third (52/167, 31%) shared their worries about a lack of in-person interactions such as face-to-face meetings. Others (9/167, 5%) stated that disruptions to their outdoor activities (eg, jogging, hiking) have affected their mental health.

Concerns About Academic Performance

A majority of participants (159/195, 82%) showed concerns about their academic performance being impacted by the pandemic. The biggest perceived challenge was the transition to online classes (61/159, 38%). In particular, participants stated their concerns about sudden changes in the syllabus, the quality of the classes, technical issues with online applications, and the difficulty of learning online. Many participants (36/159, 23%) were worried about progress in research and class projects because of restrictions put in place to keep social distancing and the lack of physical interactions with other students. Some participants (23/159, 14%) mentioned the uncertainty about their grades under the online learning environment to be a major stressor. Others (12/159, 8%) indicated their reduced motivation to learn and tendency to procrastinate.

Disruptions to Eating Patterns

COVID-19 has also negatively impacted a large portion of participants' dietary patterns (137/195, 70%). Many (35/137, 26%) stated that the amount of eating has increased, including having more snacks since healthy dietary options were reduced, and others (27/137, 20%) addressed that their eating patterns have become inconsistent because of COVID-19, for example, irregular times of eating and skipping meals. Some students (16/137, 12%) reported decreased appetite, whereas others (7/137, 5%) were experiencing emotional eating or a tendency to eat when bored. On the other hand, some students (28/195, 14%) reported that they were having healthier diets, as they were cooking at home and not eating out as much as they used to.

Changes in the Living Environment

A large portion of the participants (130/195, 67%) described that the pandemic has resulted in significant changes in their living conditions. A majority of these students (89/130, 68%) referred to living with family members as being less independent and the environment to be more distractive. For those who stayed in their residence either on- or off-campus (18/130, 14%), a main change in their living environment was reduced personal interactions with roommates. Some (9/130, 7%) mentioned that staying inside longer due to self-quarantine or shelter-in-place orders was a primary change in their living circumstances.

Financial Difficulties

More than half of the participants (115/195, 59%) expressed their concerns about their financial situations being impacted by COVID-19. Many (44/115, 38%) noted that COVID-19 has impacted or is likely to impact their own current and future employment opportunities such as part-time jobs and internships. Some (21/115, 18%) revealed the financial difficulties of their family members, mostly parents, getting laid off or receiving pay cuts in the wake of COVID-19.

Increased Class Workload

The effect of COVID-19 on class workload among the college students was not conclusive. Although slightly over half of participants (106/195, 54%) indicated their academic workload has increased due to COVID-19, the rest stated the workload has remained the same (70/195, 36%) or rather decreased (19/195, 10%). For those who were experiencing increased workloads, nearly half (51/106, 48%) thought they needed to increase their own efforts to catch up with online classes and class projects given the lack of in-person support from instructors or teaching assistants. About one-third of the participants (33/106, 31%) perceived that assignments had increased or became harder to do. Some (6/106, 6%) found that covering the remainder of coursework as the classes resumed after the 2-week break to be challenging.

Depressive Thoughts

When asked about the impact of the COVID-19 pandemic on depressive thoughts, 44% (86/195) mentioned that they were experiencing some depressive thoughts during the COVID-19 pandemic. Major contributors to such depressive thoughts were loneliness (28/86, 33%), insecurity or uncertainty (10/86, 12%), powerlessness or hopelessness (9/86, 10%), concerns about academic performance (7/86, 8%), and overthinking (4/86, 5%).

Suicidal Thoughts

Out of 195 participants, 16 (8%) stated that the pandemic has led to some suicidal thoughts with 5% (10/16) reporting these thoughts as mild and 3% (6/16) as moderate. There were 6 participants (38%) that attributed their suicidal thoughts to the presence of depressive thoughts. Other reasons were related to academic performance (1/16, 6%), problems with family as they returned home (1/16, 6%), and fear from insecurity and uncertainty (1/16, 6%).

Coping Mechanism During COVID-19

To cope with stress and anxiety imposed by COVID-19, college students reported seeking support from others but were mainly using various self-management methods.

Self-Management

The majority of the participants (105/138, 76%) with increased stress due to the outbreak of COVID-19 explained that they were using various means to help themselves cope with stress and anxiety during the pandemic. Some (24/105, 23%) relied on negative coping methods such as ignoring the news about COVID-19 (10/105), sleeping longer (7/105), distracting themselves by doing other tasks (5/105), and drinking or smoking (2/105). Approximately one-third (30/105, 29%) used positive coping methods such as meditation and breathing exercises (18/105), spiritual measures (7/105), keeping routines (4/105), and positive reframing (2/105). A majority of the participants (73/105, 70%) who used self-management

mentioned doing relaxing hobbies including physical exercise (31/105), enjoying streaming services and social media (22/105), playing with pets (7/105), journaling (5/105), listening to music (4/105), reading (2/105), and drawing (2/105). Finally, some participants (15/105, 14%) stated that they were planning activities (eg, drafting to-do lists) for academic work and personal matters as a self-distraction method.

Seeking Support From Others

Approximately one-third of the participants (47/138, 34%) mentioned that communicating with their families and friends was a primary way to deal with stress and anxiety during COVID-19. Some explicitly stated that they were using a virtual meeting application such as Zoom frequently to connect to friends and family. Only 1 participant claimed to be receiving support from a professional therapist, and another participant was using Sanvello, a mobile mental health service app provided by the university.

Barriers to Seeking Professional Support During COVID-19

Despite the availability of tele-counseling and widespread promotion of such services by the university, a vast majority of participants who indicated an increase in stress and anxiety (128/138, 93%) claimed that they had not used school counseling services during the pandemic. Reasons for such low use included the condition not being perceived as severe enough to seek the services (4/128, 3%), not comfortable interacting with unfamiliar people (1/128, 0.8%), not comfortable talking about mental health issues over the phone (1/128, 0.8%), and lack of trust in the counseling services (1/128, 0.8%).

Discussion

Principal Findings

College students comprise a population that is considered particularly vulnerable to mental health concerns. The findings of this study bring into focus the effects of pandemic-related transitions on the mental health and well-being of this specific population. Our findings suggest a considerable negative impact of the COVID-19 pandemic on a variety of academic-, health-, and lifestyle-related outcomes. By conducting online survey interviews in the midst of the pandemic, we found that a majority of the participants were experiencing increased stress and anxiety due to COVID-19. In addition, results of the PSS showed moderate levels of stress among our participants. This is in line with a recent pre–COVID-19 survey conducted in the United Kingdom (mean PSS score 19.79, SD 6.37) [28]; however, the administration of PSS as interview questions (compared to allowing participants to read and respond to the 10 questions) might have introduced bias and resulted in underreporting.

Among the effects of the pandemic identified, the most prominent was worries about one's own health and the health of loved ones, followed by difficulty concentrating. These findings are in line with recent studies in China that also found concerns relating to health of oneself and of family members being highly prevalent among the general population during the pandemic. Difficulty in concentrating, frequently expressed by our participants, has previously been shown to adversely affect students' confidence in themselves [29], which has known correlations to increased stress and mental health [30]. In comparison with stress and anxiety in college students' general life, it appears that countermeasures put in place against COVID-19, such as shelter-in-place orders and social distancing practices, may have underpinned significant changes in students' lives. For example, a vast majority of the participants noted changes in social relationships, largely due to limited physical interactions with their families and friends. This is similar to recent findings of deteriorated mental health status among Chinese students [10] and increased internet search queries on negative thoughts in the United States [31]. The findings on the impact of the pandemic on sleeping and eating habits are also a cause for concern, as these variables have known correlations with depressive symptoms and anxiety [20].

Although a majority of participants expressed concerns regarding academic performance, interestingly, almost half of the participants reported lower stress levels related to academic pressure and class workload since the pandemic began. This may be due, in part, to decisions taken by professors and the university to ease the students' sudden transition to distance learning. For instance, this university allowed students to choose a pass/fail option for each course instead of a regular letter grade. Additionally, actions taken by professors, such as reduced course loads, open book examinations, and other allowances on grading requirements, could also have contributed to alleviating or reducing stress. Although participants who returned to their parental home reported concerns about distractions and independence, students might have benefited from family support and reduced social responsibilities. Therefore, the increased stress due to the pandemic may have been offset, at least to some extent.

Alarmingly, 44% (86/195) of the participants reported experiencing an increased level of depressive thoughts, and 8% (16/195) reported having suicidal thoughts associated with the COVID-19 pandemic. Previous research [32] reported about 3%-7% of the college student population to have suicidal thoughts outside of the pandemic situation. Furthermore, with the exception of high-burnout categories, depression levels among students, reported in several recent studies [33-35], have varied between 29% and 38%, which may suggest an uptick in pandemic-related depressive symptoms among college students similar to recent studies in China [10,11]. Although our participants specifically mentioned several factors such as feelings of loneliness, powerlessness, as well as financial and academic uncertainties, other outcomes that were perceived to be impacted by the COVID-19 pandemic may also act as contributors to depressive thoughts and suicidal ideation. In particular, both difficulty concentrating and changes in sleeping habits are associated with depression [20,29,36].

Our study also identifies several coping mechanisms varying between adaptive and maladaptive behaviors. The maladaptive coping behaviors such as denial and disengagement have been shown to be significant predictors of depression among young adults [37]. In contrast, adaptive coping such as acceptance and proactive behaviors are known to positively impact mental health. Our findings suggest that the majority of our participants

exhibited maladaptive coping behaviors. Identifying students' coping behavior is important to inform the planning and design of support systems. In this regard, participatory models of intervention development can be used, in which researchers' and psychologists' engagement with the target population to adapt interventional programs to their specific context has shown promise [37,38]. For instance, Nastasi et al [37] used a participatory model to develop culture-specific mental health services for high school students in Sri Lanka. Similar approaches can be adopted to engage college students as well to develop a mental health program that leverages their natural positive coping behaviors and addresses their specific challenges.

Participants described several barriers to seeking help, such as lack of trust in counseling services and low comfort levels in sharing mental health issues with others, which may be indicative of stigma. Perceiving social stigma as a barrier to seeking help and availing counseling services and other support is common among students [29]. One study showed that only a minor fraction of students who screened positive for a mental health problem actually sought help [39]. Although overcoming the stigma associated with mental health has been discussed at length, practical ways of mitigating this societal challenge remains a gap [40,41]. Our findings suggest that self-management is preferred by students and should be supported in future work. Digital technologies and telehealth applications have shown some promise to enable self-management of mental health issues [42]. For instance, Youn et al [43] successfully used social media networks as a means to reach out to college students and screen for depression by administering a standardized scale, the Patient Health Questionnaire-9. Digital web-based platforms have also been proposed to enhance awareness and communication with care providers to reduce stigma related to mental health among children in underserved communities [44]. For instance, one of the online modules suggested by the authors involves providing information on community-identified barriers to communicating with care providers. Technologies such as mobile apps and smart wearable sensors can also be leveraged to enable self-management and communication with caregivers.

In light of the aforementioned projections of continued COVID-19 cases at the time of this writing [45] and our findings, there is a need for immediate attention to and support for students and other vulnerable groups who have mental health issues [17]. As suggested by a recent study [46] based on the Italian experience of this pandemic, it is essential to assess the population's stress levels and psychosocial adjustment to plan for necessary support mechanisms, especially during the recovery phase, as well as for similar events in the future. Although the COVID-19 pandemic seems to have resulted in a widespread forced adoption of telehealth services to deliver psychiatric and mental health support, more research is needed to investigate use beyond COVID-19 as well as to improve preparedness for rapid virtualization of psychiatric counseling or tele-psychiatry [47-49].

Limitations and Future Work

To our knowledge, this is the first effort in documenting the psychological impacts of the COVID-19 pandemic on a representative sample of college students in the United States via a virtual interview survey method in the middle of the pandemic. However, several limitations should be noted. First, the sample size for our interview survey was relatively small compared to typical survey-only studies; however, the survey interview approach affords the capture of elaboration and additional clarifying details, and therefore complements the survey-based approaches of prior studies focusing on student mental health during this pandemic [10,11,50]. Second, the sample used is from one large university, and findings may not generalize to all college students. However, given the nationwide similarities in universities transitioning to virtual classes and similar stay-at-home orders, we expect reasonable generalizability of these findings. Additionally, a majority of our participants were from engineering majors. Therefore, future work is needed to use a stratified nationwide sample across wider disciplines to verify and amend these findings. Third, although a vast majority of participants answered that they have not used the university counseling service during the pandemic, only a few of them provided reasons. Since finding specific reasons behind the low use is a key to increasing college students' uptake of available counseling support, future research is warranted to unveil underlying factors that hinder college students' access to mental health support. Finally, we did not analyze how student mental health problems differ by demographic characteristics (eg, age, gender, academic year, major) or other personal and social contexts (eg, income, religion, use of substances).

Future work could focus on more deeply probing the relationships between various coping mechanisms and stressors. Additionally, further study is needed to determine the effects of the pandemic on students' mental health and well-being in its later phases beyond the peak period. As seen in the case of health care workers in the aftermath of the severe acute respiratory syndrome outbreak, there is a possibility that the effects of the pandemic on students may linger for a period beyond the peak of the COVID-19 pandemic itself [51].

Acknowledgments

This research was partly funded by a Texas A&M University President's Excellence (X-Grant) award.

Conflicts of Interest

None declared.

References

1. Unger K. Handbook on Supported Education: Providing Services for Students With Psychiatric Disabilities. Charleston, SC: BookSurge Publishing; 2007.
2. 2019 annual report. Center for Collegiate Mental Health. University Park, PA: Penn State University; 2020. URL: https://ccmh.memberclicks.net/assets/docs/2019-CCMH-Annual-Report_3.17.20.pdf [accessed 2020-05-11]
3. Shuchman M. Falling through the cracks — Virginia Tech and the restructuring of college mental health services. N Engl J Med 2007 Jul 12;357(2):105-110. [doi: 10.1056/nejmp078096]
4. Eisenberg D, Downs MF, Golberstein E, Zivin K. Stigma and help seeking for mental health among college students. Med Care Res Rev 2009 Oct;66(5):522-541. [doi: 10.1177/1077558709335173] [Medline: 19454625]
5. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
6. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
7. Xie X, Xue Q, Zhou Y, Zhu K, Liu Q, Zhang J, et al. Mental health status among children in home confinement during the coronavirus disease 2019 outbreak in Hubei Province, China. JAMA Pediatr 2020 Apr 24:e201619 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.1619] [Medline: 32329784]
8. Kirzinger A, Kearney A, Hamel L, Brodie M. KFF health tracking poll - early April 2020: the impact Of coronavirus on life In America. Kaiser Family Foundation. 2020 Apr 02. URL: https://www.kff.org/health-reform/report/kff-health-tracking-poll-early-april-2020/ [accessed 2020-05-10]
9. Nelson B, Pettitt A, Flannery J, Allen N. Rapid assessment of psychological and epidemiological correlates of COVID-19 concern, financial strain, and health-related behavior change in a large online sample. Int J Methods in Psychiatr Res 2020 Apr 13;21(3):169-184. [doi: 10.31234/osf.io/jftze]
10. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
11. Liu X, Liu J, Zhong X. Psychological state of college students during COVID-19 epidemic. SSRN J 2020 Mar 24:A. [doi: 10.2139/ssrn.3552814]
12. Wang C, Zhao H. The impact of COVID-19 on anxiety in Chinese university students. Front Psychol 2020;11:1168. [doi: 10.3389/fpsyg.2020.01168] [Medline: 32574244]
13. Bruffaerts R, Mortier P, Kiekens G, Auerbach RP, Cuijpers P, Demyttenaere K, et al. Mental health problems in college freshmen: prevalence and academic functioning. J Affect Disord 2018 Jan 01;225:97-103 [FREE Full text] [doi: 10.1016/j.jad.2017.07.044] [Medline: 28802728]
14. Zhai Y, Du X. Addressing collegiate mental health amid COVID-19 pandemic. Psychiatry Res 2020 Jun;288:113003 [FREE Full text] [doi: 10.1016/j.psychres.2020.113003] [Medline: 32315885]
15. Zhai Y, Du X. Mental health care for international Chinese students affected by the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e22 [FREE Full text] [doi: 10.1016/S2215-0366(20)30089-4] [Medline: 32199511]
16. de Oliveira Araújo FJ, de Lima LSA, Cidade PIM, Nobre CB, Neto MLR. Impact of Sars-Cov-2 and its reverberation in global higher education and mental health. Psychiatry Res 2020 Jun;288:112977 [FREE Full text] [doi: 10.1016/j.psychres.2020.112977] [Medline: 32302818]
17. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
18. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
19. Cohen S, Kessler R, Gordon L. Measuring stress: A Guide for Health and Social Scientists. New York, NY: Oxford University Press; 1997.
20. Acharya L, Jin L, Collins W. College life is stressful today - emerging stressors and depressive symptoms in college students. J Am Coll Health 2018 Oct;66(7):655-664. [doi: 10.1080/07448481.2018.1451869] [Medline: 29565759]
21. Baghurst T, Kelley BC. An examination of stress in college students over the course of a semester. Health Promot Pract 2014 May;15(3):438-447. [doi: 10.1177/1524839913510316] [Medline: 24231633]
22. Zoom. Zoom Video Communications. URL: https://zoom.us/
23. Otter.ai. URL: https://otter.ai/login
24. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. Guest G, MacQueen K, Namey E. Applied Thematic Analysis. Thousand Oaks, CA: Sage Publications; 2011.
26. Corbin J, Strauss A. Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications; 2014.

27. MAXQDA. VERBI Software. URL: https://www.maxqda.com/
28. Denovan A, Dagnall N, Dhingra K, Grogan S. Evaluating the perceived stress scale among UK university students: implications for stress measurement and management. Stud Higher Education 2017 Jun 22;44(1):120-133. [doi: 10.1080/03075079.2017.1340445]
29. Martin JM. Stigma and student mental health in higher education. Higher Education Res Dev 2010 Jun;29(3):259-274. [doi: 10.1080/07294360903470969]
30. Zuckerman DM. Stress, self-esteem, and mental health: how does gender make a difference? Sex Roles 1989 Apr;20(7-8):429-444. [doi: 10.1007/BF00288001]
31. Jacobson NC, Lekkas D, Price G, Heinz MV, Song M, O'Malley AJ, et al. Flattening the mental health curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. JMIR Ment Health 2020 Jun 01;7(6):e19347 [FREE Full text] [doi: 10.2196/19347] [Medline: 32459186]
32. Zivin K, Eisenberg D, Gollust SE, Golberstein E. Persistence of mental health problems and needs in a college student population. J Affect Disord 2009 Oct;117(3):180-185. [doi: 10.1016/j.jad.2009.01.001] [Medline: 19178949]
33. Zeng Y, Wang G, Xie C, Hu X, Reinhardt JD. Prevalence and correlates of depression, anxiety and symptoms of stress in vocational college nursing students from Sichuan, China: a cross-sectional study. Psychol Health Med 2019 Aug;24(7):798-811. [doi: 10.1080/13548506.2019.1574358] [Medline: 30714819]
34. Nahas ARMF, Elkalmi R, Al-Shami A, Elsayed T. Prevalence of depression among health sciences students: findings from a public university in Malaysia. J Pharm Bioallied Sci 2019;11(2):170-175 [FREE Full text] [doi: 10.4103/jpbs.JPBS_263_18] [Medline: 31148894]
35. Fitzpatrick O, Biesma R, Conroy RM, McGarvey A. Prevalence and relationship between burnout and depression in our future doctors: a cross-sectional study in a cohort of preclinical and clinical medical students in Ireland. BMJ Open 2019 May 01;9(4):e023297 [FREE Full text] [doi: 10.1136/bmjopen-2018-023297] [Medline: 31048421]
36. Heiligenstein E, Guenther G, Hsu K, Herman K. Depression and academic impairment in college students. J Am Coll Health 1996 Sep;45(2):59-64. [doi: 10.1080/07448481.1996.9936863] [Medline: 8908879]
37. Nastasi BK, Sarkar S, Varjas K, Jayasena A. Participatory model of mental health programming: lessons learned from work in a developing country. Sch Psychol Rev 1998;27(2):260-276.
38. Cappella E, Jackson DR, Bilal C, Hamre BK, Soulé C. Bridging mental health and education in urban elementary schools: participatory research to inform intervention development. Sch Psychol Rev 2011;40(4):486-508.
39. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
40. Husain W. Barriers in seeking psychological help: public perception in Pakistan. Community Ment Health J 2020 Jan;56(1):75-78. [doi: 10.1007/s10597-019-00464-y] [Medline: 31542848]
41. Fischer EP, McSweeney JC, Wright P, Cheney A, Curran GM, Henderson K, et al. Overcoming barriers to sustained engagement in mental health care: perspectives of rural veterans and providers. J Rural Health 2016 Sep;32(4):429-438. [doi: 10.1111/jrh.12203] [Medline: 27557334]
42. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
43. Youn SJ, Trinh N, Shyu I, Chang T, Fava M, Kvedar J, et al. Using online social media, Facebook, in screening for major depressive disorder among college students. Int J Clin Health Psychol 2013 Jan;13(1):74-80. [doi: 10.1016/s1697-2600(13)70010-3]
44. Ginossar T, Nelson S. Reducing the health and digital divides: a model for using community-based participatory research approach to e-health interventions in low-income Hispanic communities. J Computer-Mediated Commun 2010;15(4):530-551. [doi: 10.1111/j.1083-6101.2009.01513.x]
45. COVID-19 projections. Institute for Health Metrics and Evaluation. Seattle, WA: University of Washington URL: https://covid19.healthdata.org/united-states-of-america [accessed 2020-05-20]
46. de Girolamo G, Cerveri G, Clerici M, Monzani E, Spinogatti F, Starace F, et al. Mental health in the coronavirus disease 2019 emergency-the Italian response. JAMA Psychiatry 2020 Apr 30:A. [doi: 10.1001/jamapsychiatry.2020.1276] [Medline: 32352480]
47. Zhou J, Liu L, Xue P, Yang X, Tang X. Mental health response to the COVID-19 outbreak in China. Am J Psychiatry 2020 Jul 01;177(7):574-575. [doi: 10.1176/appi.ajp.2020.20030304] [Medline: 32375540]
48. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
49. Shore JH, Waugh M, Calderone J, Donahue A, Rodriguez J, Peters D, et al. Evaluation of telepsychiatry-enabled perinatal integrated care. Psychiatr Serv 2020 May 01;71(5):427-432. [doi: 10.1176/appi.ps.201900143] [Medline: 32019433]
50. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]

51. McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. Can J Psychiatry 2007 Apr;52(4):241-247. [doi: 10.1177/070674370705200406] [Medline: 17500305]

Abbreviations

PSS: Perceived Stress Scale-10

Edited by G Eysenbach, G Fagherazzi, J Torous; submitted 10.06.20; peer-reviewed by T Liu, V Hagger; comments to author 28.07.20; revised version received 01.08.20; accepted 15.08.20; published 03.09.20.

Please cite as:
Son C, Hegde S, Smith A, Wang X, Sasangohar F
Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study
J Med Internet Res 2020;22(9):e21279
URL: https://www.jmir.org/2020/9/e21279
doi:10.2196/21279
PMID:32805704

Original Paper

Rapid Implementation and Innovative Applications of a Virtual Intensive Care Unit During the COVID-19 Pandemic: Case Study

Atiya Dhala[1], MD, FACP; Farzan Sasangohar[2,3], BA, BCS, MASc, SM, PhD; Bita Kash[3,4], BS, MBA, PhD; Nima Ahmadi[3], BS, MSEM, PhD; Faisal Masud[5], MD, FCCM, FCCP

[1]Department of Surgery, Houston Methodist Hospital, Houston, TX, United States

[2]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[3]Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, United States

[4]School of Public Health, Texas A&M University, College Station, TX, United States

[5]Departments of Anesthesiology and Critical Care, Houston Methodist DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston, TX, United States

Corresponding Author:
Farzan Sasangohar, BA, BCS, MASc, SM, PhD
Department of Industrial and Systems Engineering
Texas A&M University
3131 TAMU
College Station, TX, 77843
United States
Phone: 1 9794582337
Email: sasangohar@tamu.edu

Abstract

Background: The COVID-19 pandemic has necessitated a rapid increase of space in highly infectious disease intensive care units (ICUs). At Houston Methodist Hospital (HMH), a virtual intensive care unit (vICU) was used amid the COVID-19 outbreak.

Objective: The aim of this paper was to detail the novel adaptations and rapid expansion of the vICU that were applied to achieve patient-centric solutions while protecting staff and patients' families during the pandemic.

Methods: The planned vICU implementation was redirected to meet the emerging needs of conversion of COVID-19 ICUs, including alterations to staged rollout timing, virtual and in-person staffing, and scope of application. With the majority of the hospital critical care physician workforce redirected to rapidly expanded COVID-19 ICUs, the non–COVID-19 ICUs were managed by cardiovascular surgeons, cardiologists, neurosurgeons, and acute care surgeons. HMH expanded the vICU program to fill the newly depleted critical care expertise in the non–COVID-19 units to provide urgent, emergent, and code blue support to all ICUs.

Results: Virtual family visitation via the Consultant Bridge application, palliative care delivery, and specialist consultation for patients with COVID-19 exemplify the successful adaptation of the vICU implementation. Patients with COVID-19, who were isolated and separated from their families to prevent the spread of infection, were able to virtually see and hear their loved ones, which bolstered the mental and emotional status of those patients. Many families expressed gratitude for the ability to see and speak with their loved ones. The vICU also protected medical staff and specialists assigned to COVID-19 units, reducing exposure and conserving personal protective equipment.

Conclusions: Telecritical care has been established as an advantageous mechanism for the delivery of critical care expertise during the expedited rollout of the vICU at Houston Methodist Hospital. Overall responses from patients, families, and physicians are in favor of continued vICU care; however, further research is required to examine the impact of innovative applications of telecritical care in the treatment of critically ill patients.

(J Med Internet Res 2020;22(9):e20143) doi:10.2196/20143

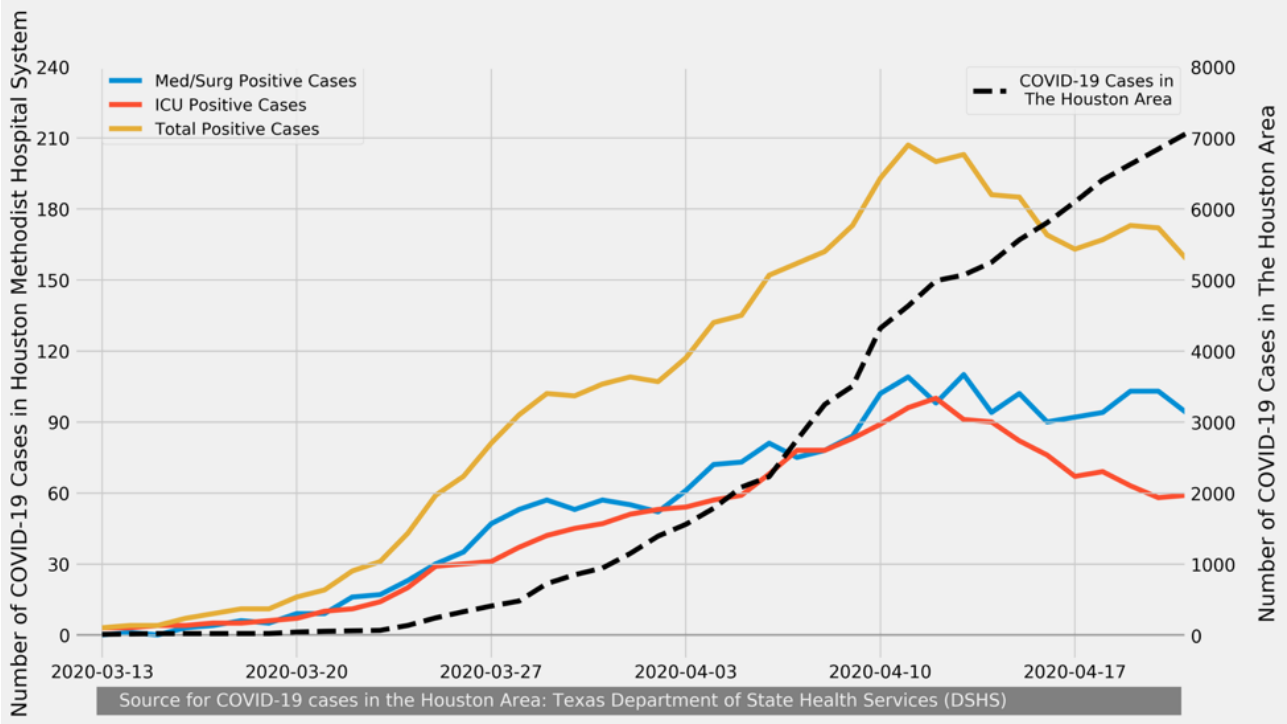
KEYWORDS

intensive care units; critical care; pandemics; SARS-CoV-2; telemedicine; infection control; COVID-19

Introduction

The outbreak of the SARS-CoV-2 pandemic has been widely reported in news articles and journals [1]. Although the initial flow of patients with COVID-19 in Houston, Texas, began slowly compared to that in other large metropolises such as New York City, Houston Methodist Hospital (HMH) began to marshal all its resources to plan for the exponential growth of patients who tested positive for or were suspected of having COVID-19. HMH coordinated responses with other hospitals in the Texas Medical Center (TMC), the largest medical district in the world. As the number of COVID-19 cases increased across the region, the intensive care unit (ICU) beds dedicated to COVID-19 patients in HMH also increased (Figure 1). At the peak of the pandemic, HMH had dedicated 150 ICU beds to patients with COVID-19 (with potential for a two- to four-fold increase) and dedicated bedside critical care physicians to serve these patients.

Figure 1. Number of Houston area COVID-19 cases (dashed line) from March 13-April 21, 2020 compared to the numbers of Houston Methodist cases (solid lines). ICU: intensive care unit; Med/Surg: Medical/Surgical.



In this article, we review how HMH implemented its telecritical care program by focusing on the broader innovative application of this virtual technology to find patient-centric solutions while protecting staff and patient families during an extraordinary pandemic situation. The documented work has been identified as exempt by the institutional review board at HMH.

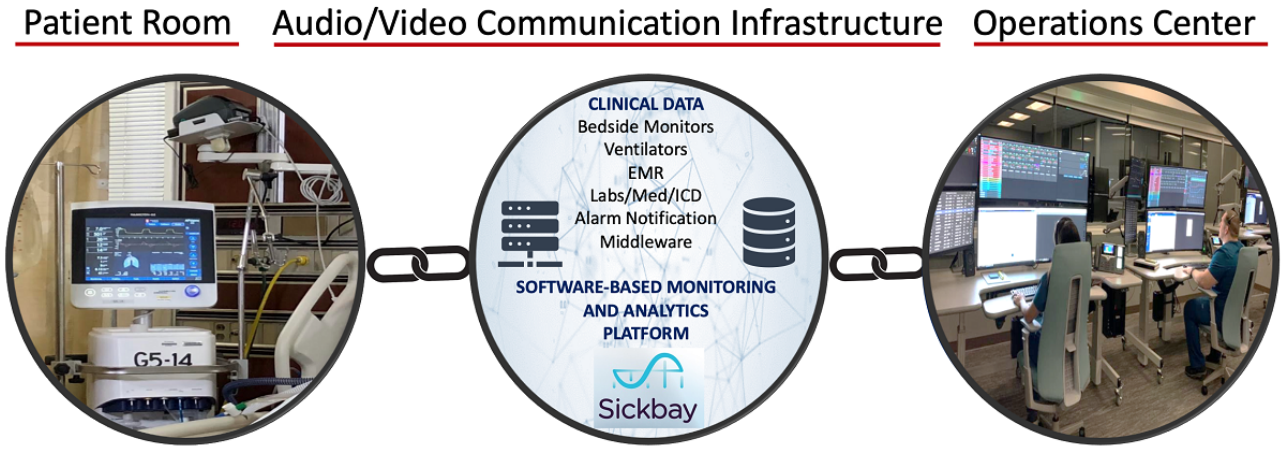
Background

Telecritical care has revolutionized the delivery of care by enabling remote monitoring and treatment of ICU patients. Since their first pilot implementation in 1997 [2], telecritical care platforms, initially known and trademarked as electronic intensive care units (eICUs), have extended access to resources (eg, critical care physicians, specialized consults, and nurses), provided a wide range of decision-support tools, and enabled monitoring and analysis of a large amount of physiological data [3]. Recent studies show large variability in the mortality impacts of telecritical care platforms; for instance, in a 2016 national effectiveness study by Kahn et al [4], the positive impacts on the length of stay, cost, and quality of care contributed to increased popularity and adoption rates of telecritical care (see also [5,6]). A 2017 survey of 722 hospitals worldwide (672 in the United States) showed that 35% had formal telecritical care programs [6].

The HMH vICU

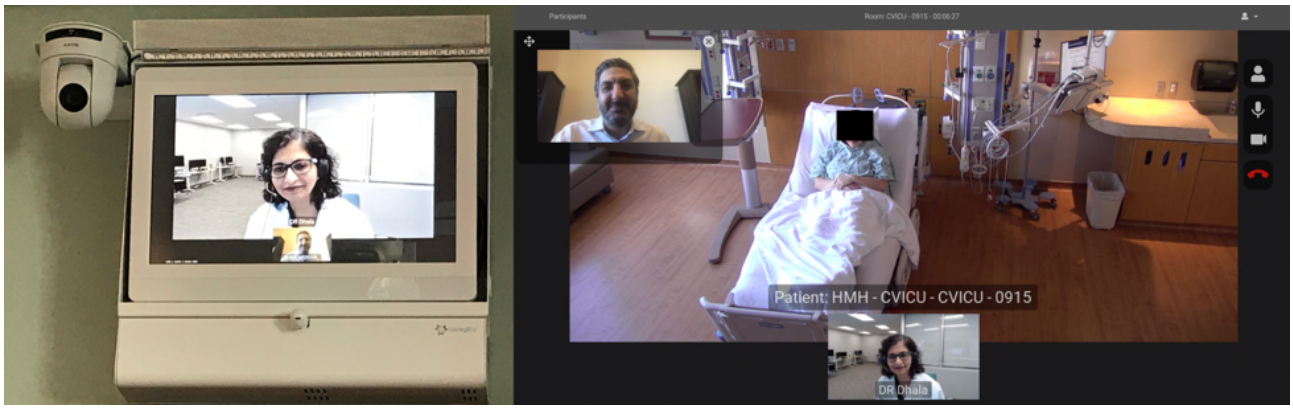
Before the onset of the COVID-19 pandemic, HMH launched its innovative telecritical care program, branded as the Virtual Intensive Care Unit (vICU), to augment the critical care services being provided in its ICUs. The HMH vICU has three main components: the operations center, the patient room, and the audiovisual (AV) communication infrastructure linking the first two components (Figure 2). The operations center, located at HMH's main campus at the TMC, provides a central command capability where medical doctors and registered nurses, referred to as Virtual MDs (vMDs) and Virtual RNs (vRNs), connect and monitor the status of patients using the AV equipment installed in the patients' rooms. The AV communication infrastructure includes a single camera in the patient rooms with 360-degree pan, tilt, and zoom capability that enables the vMDs and vRNs to examine the patient; they can also focus on intravenous fluids, drip rates, monitors, and ventilator settings, and, of course, they can communicate with the patients, their visiting families, and the bedside providers.

Figure 2. The three main components of the virtual intensive care unit system. AV: audiovisual; EMR: electronic medical record; ICD: Implantable Cardioverter-Defibrilators.



Our operations center uses a US Food and Drug Administration–cleared software-based monitoring and analytics platform called Sickbay (Medical Informatics Corp). The physiological data from several bedside monitors and devices, including ventilators and hemodynamic monitors, and the static data from the electronic medical record (EMR) interface with this platform. Novel algorithms transform these big data into actionable information in the form of risk scores, which feed into clinical decision support systems. The operations center is also where virtual AV connections are established with the patients' rooms. For emergent or urgent calls, the bedside teams can access the vICU by pressing a virtual alert button in each room to engage a vMD or vRN from the operations center. The vMDs and the vRNs can connect by camera within seconds to respond. In addition, the operations center can connect a consultant or an expert from any remote location through an application called Consultant Bridge. The operations center staff can send an SMS text message or email a link to the consultant to "bridge" them into the patient's room. Consultant Bridge enables the vMD, the consultant, and the bedside team to have a three-way video call using the cameras and monitors already installed in the ICU patient rooms (Figure 3).

Figure 3. Virtual intensive care unit audio and video equipment installed in an intensive care unit room (left) and caller view (right). For the purposes of illustration, a member of the hospital staff played the role of the patient and one of the authors played the role of the caller.



Initially, the vICU staff, consisting of one vMD and two vRNs, provided nocturnal coverage for HMH's 36-bed neurology ICU (Neuro-ICU). To ensure a smooth transition and to minimize any disruptions to existing processes, the implementation plan called for a stepwise rollout of vICU to the other three units: the medical ICU (MICU), cardiovascular ICU (CVICU), and cardiac ICU (CICU). The vICU process required that all the staff's expectations be clearly defined and workflows integrated appropriately while nurturing a robust and productive relationship with the bedside teams. The vICU and the ICU teams initiated a collaborative effort to design the essential workflows almost four months before the actual launch of the vICU.

Methods

ICU Expansions in Response to the COVID-19 Surge

The expected surge of the COVID-19 pandemic necessitated a sudden change in the rollout plan for the vICU. To fully prepare for the surge, HMH forecasted the number of beds that might be needed during the peak of the outbreak. At the time, the vICU had already started providing coverage to the Neuro-ICU during the night shift. The 24-bed MICU was the next unit scheduled to go live with vICU coverage. Instead, on March 12, 2020, the MICU became the first dedicated COVID-19 ICU at HMH. The novelty of SARS-CoV-2, combined with the complexity of treating COVID-19 patients, forced HMH to redirect all its critical care intensivists to the COVID-19 units, providing 24-hour bedside coverage.

Staffing Changes in Response to the COVID-19 Surge

By March 26, 2020, the first COVID-19 ICU had almost reached capacity; this required the establishment of a second dedicated COVID-19 ICU, which was created by converting the 36-bed Neuro-ICU. By April 8, 2020, the Neuro-ICU had also reached full capacity, and a third COVID-19 ICU was set up by converting a 19-bed step down unit. For all the COVID-19 ICUs, the staffing ratio of intensivists to patients was 1:8 to 1:12 during the day and 1:12 to 1:18 during the night. Because the hospital had redirected the entire critical care physician workforce to the COVID-19 ICUs, the non–COVID-19 ICUs were managed by cardiovascular surgeons, cardiologists, neurosurgeons, and acute care surgeons.

To fill out the newly depleted critical care expertise in the non–COVID-19 ICUs, HMH expanded the night vICU program to provide urgent, emergent, and code blue support for all ICUs, including the COVID-19 ICUs. The bedside teams in the non–COVID-19 ICUs included advanced practice providers (APPs) who reached out to the vMDs for critical care support and admitted medicine overflow patients with the vMDs at night. HMH's next step in the surge planning was to open the vICU in the daytime, providing critical care support from vMDs and vRNs during the day for the non–COVID-19 ICUs.

In the first week of April 2020, HMH began surveillance testing of its medical staff, and many frontline employees tested positive for COVID-19. All the COVID-19–positive workers, although asymptomatic, were removed from their clinical duties in the ICUs, and the resulting staff shortages expedited the integration of the vICU with the rest of the ICUs. Before the staff shortages, the vICU staff were only providing support for emergent, urgent, and code blue situations. However, the expanded vICU staff began to cover admissions and triage during the night for more comprehensive integration.

Results

ICU Expansions in Response to the COVID-19 Surge

It is well documented that family engagement has an enormous positive impact on ICU patients by decreasing anxiety, confusion, agitation, and delirium [7-9]. Evidence suggests that separating families from the patients can adversely impact the patient's feelings of security and of ultimate outcome [10].

During the COVID-19 pandemic, nearly all hospitals disallowed visitors for adult inpatients, including all COVID-19 and non–COVID-19 ICU patients. While video chat technology (eg, FaceTime and Skype) is commonly available among patients and their families, in an ICU setting with a highly infectious, critically ill cohort of patients who may be sedated and frequently intubated, the use of this common technology was not feasible because it would require staff to bring in and position the equipment (eg, smartphone or tablet) while using personal protective equipment (PPE).

However, the vICU infrastructure provided a readily available and much more accessible means of connecting the patients with COVID-19 with their families. HMH began to offer this technology to the families for emotional support and improved patient care. Two vRNs (in the operation center) were tasked with reaching out to the bedside teams and collaborating with the bedside nurses, physicians, and unit managers to gain access to the patients' families. Family members received links on their smartphones that instantly connected them with their loved one's ICU room using the Consultant Bridge feature. The vICU patients logged approximately 20 to 40 calls per day using the Consultant Bridge during this period. Because restrictions on visitors also applied to non–COVID-19 patients and their families, the Consultant Bridge was used for all ICU patients. The results of a short postcall quality assessment survey showed overwhelming satisfaction with the access to the patient using vICU technology.

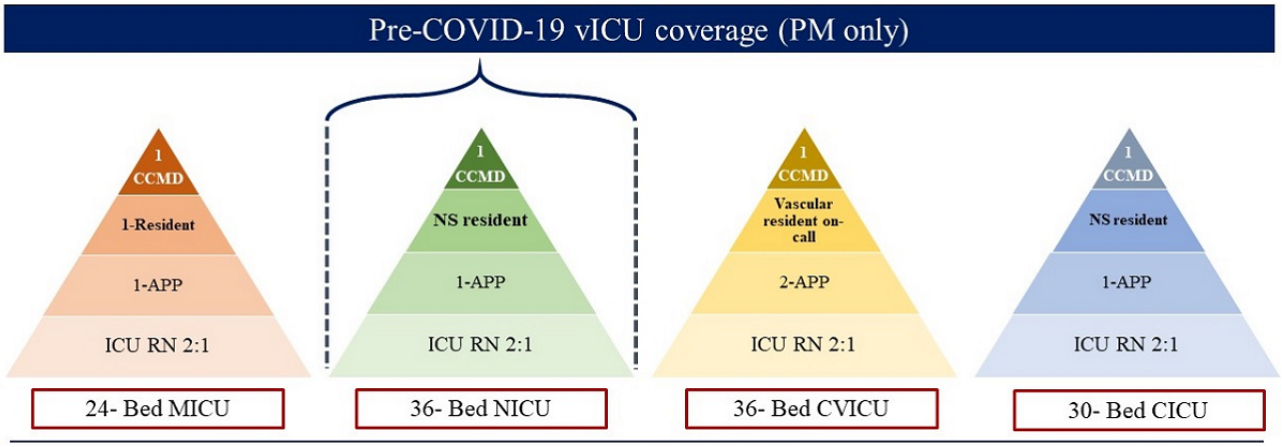
Palliative Care for ICU Patients

In standard palliative care situations, family members are involved in the decision-making process at the bedside. However, in a pandemic situation, when the family members cannot be near their terminally ill loved ones, vICU technology may provide a vehicle to perform the critical steps with the full participation of the palliative care team and the family members. Similar to regular visitations, we used the Consultant Bridge feature of vICU to enable palliative care.

Staffing Changes in Response to the COVID-19 Surge

The demands of the COVID-19 pandemic resulted in the establishment of a tiered staffing model, with vICU providing essential support to all the ICUs. This model bears a strong resemblance to the tiered staffing model described by the Society of Critical Care Medicine for surge planning [11]. Prior to the COVID-19 pandemic, HMH had commenced its vICU program, providing nocturnal coverage for the Neuro-ICU combined with full workflow implementation. During this period, HMH used four in-house intensivists and the following staffing ratios: 2:1 for ICU RNs; 1 APP for 24 beds; 1 resident on call in the night; and 1 critical care bedside physician per unit (4 in total) (Figure 4).

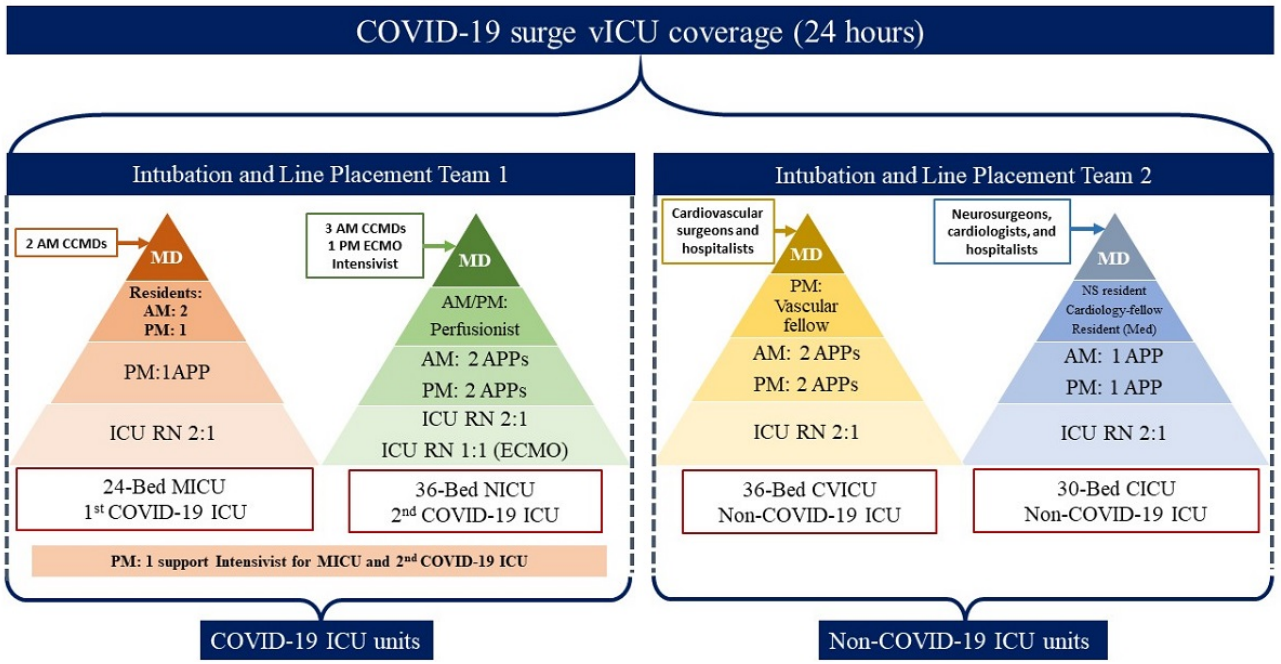
Figure 4. Houston Methodist Hospital tiered staffing model for critical care services supported by the virtual intensive care unit prior to the COVID-19 pandemic. APP: advanced practice provider; CCMD: critical care intensivist; CICU; cardiac intensive care unit; CVICU: cardiovascular intensive care unit; ICU: intensive care unit; MICU: medical intensive care unit; NS: neurosurgery; RN: registered nurse; vICU: virtual intensive care unit.



There are many consultants whose presence at the bedside is not absolutely necessary during this pandemic despite their useful contributions to patient care plans. For these cases, vICU was able to bridge these consultants and their expertise to patient rooms without exposing the consultants to a highly infectious disease or requiring the expenditure of scarce PPE. The superior quality of the real-time video feed coupled with advanced zoom capability greatly enhanced the quality of these remote consults. For example, when a patient who tested positive for COVID-19 developed a maculopapular rash, the bedside team was able to perform a skin biopsy with the guidance of a dermatologist connected through the Consultant Bridge. The high quality of video transmission enabled the dermatologist to clearly see the skin lesion and direct the procedure.

To adapt to the needs of patients with COVID-19, HMH expanded its vICU coverage from one unit to all units. HMH assigned all the intensivists to serve patients with COVID-19, with one support intensivist covering both COVID-19 ICUs at night. For the non–COVID-19 ICUs, coverage was provided by non–critical care attending physicians, mostly specialty surgeons, supported by vMDs and vRNs (Figure 5).

Figure 5. Houston Methodist Hospital tiered staffing model for critical care services supported by a virtual intensive care unit during the COVID-19 pandemic. APP: advanced practice provider; CCMD: critical care intensivist; CICU: cardiac intensive care unit; ECMO: extracorporeal membrane oxygenation; ICU: intensive care unit: MD: medical doctor; Med: Medical; MICU: medical intensive care unit; NICU: neurology intensive care unit; RN: registered nurse: vICU: virtual intensive care unit.



Discussion

Principal Findings

The experiences of HMH clinicians show that telecritical care platforms can be an essential part of the critical care services toolkit. While the benefits of having bedside intensivists have long been well known, telecritical care experts can support and augment the work of bedside teams treating critically ill patients, thereby expanding the scope and boundaries of traditional ICUs. Under normal conditions, telecritical care services may help reduce staff burnout in ICUs, lower ICU mortality rates, reduce length of stay in the hospital or ICU, and promote greater adherence to best practices.

During the COVID-19 pandemic, the challenges of treating patients with a highly infectious disease brought several novel applications of HMH's vICU into sharp relief with positive patient outcomes. While HMH's response to the COVID-19 pandemic quickly paved the way for the full integration of the vICU with all its ICUs, more work is needed to improve preparedness for and resilience in future pandemics. In particular, frequent drills [12] and simulation training [13] have shown promise in preparing ICU staff for rare but high-consequence pandemic events. These efforts can focus on identifying necessary adjustments to staffing, layout, processes, and new use cases for vICU technology. Participatory ergonomics methods and models such as technology acceptance have shown promise in eliciting ICU staff acceptance and intention to use these technological innovations [14].

One of the most important serendipitous benefits came from connecting patients to their family members through the use of the vICU platform via the Consultant Bridge application. Patients with COVID-19, who were isolated and separated from their families due to their highly infectious disease, were able to virtually see and hear their loved ones, bolstering the mental and emotional status of those patients. Many families expressed gratitude for this ability to see and speak with their loved ones. We observed tremendous potential in the use of this technology, and we believe that even non-ICU units may benefit from some form of telepresence technology to facilitate family engagement. While the Health Insurance Portability and Accountability Act (HIPAA) and consent regulations have been relaxed in light of the current pandemic, our vICU platform, including the Consultant Bridge, has been tailored to be HIPAA-compliant. In addition, the bedside staff ensure that the patient is "camera-ready" before any camera in the room is allowed to go live. Despite challenges related to privacy and security, a recent attitude shift toward viewing family members as partners in shared decision-making [15] has set the stage for a potential "open" vICU.

One of the significant contributors to burnout among HMH ICU providers is the anxiety associated with PPE shortages and exposure [16]. The use of the vICU provided a unique level of protection to the medical staff assigned to COVID-19 units. The ICU staff were given local access to patient rooms by vICU-enabled laptops that allowed the bedside physicians and nurses to connect to the patient's room by camera without having to don and doff protective gear each time when checking on a patient. Such remote access reduced the usage of PPEs, which were already in short supply. The vICU's Consultant Bridge also allowed specialty consultants, including extracorporeal membrane oxygenation specialists, cardiologists, and endocrinologists, to examine patients with COVID-19 virtually.

Installation

A financial cost-benefit analysis of the vICU installation is beyond the scope of this paper. However, it should be noted that physical installation of the cameras, the alert buttons, and the rest of the communication infrastructure requires careful planning, including moving current patients while the rooms are being retrofitted with the equipment. Proper planning and advanced notification of the clinical teams are essential to remove any risk of disruption to patient care during the installation process.

Scalability

With the ongoing resurgence of COVID-19 cases, HMH has been able to scale its ICU bed capacity by leveraging its vICU resources. The hospital converted previously mothballed ICU beds or intermediate care units into functioning ICU COVID-19–specific units by rolling out a mobile vICU "cart" that contains the camera and other communication hardware. These carts can be moved into any room or near any bed, providing instantaneous vICU connectivity to the vICU operation center and staff. While they are not a perfect substitute for a fully fitted vICU patient room, these mobile vICU carts provide great flexibility in deploying the vICU resources to newly converted ICU units or emergency rooms where critically ill patients may be waiting for an ICU bed.

More importantly, the structure of the vICU platform enables a high degree of scalability, as each vMD can provide coverage for up to 200 ICU patients. The patient-to-vMD ratio may vary among hospitals and health care systems. Finally, the vICU platform also enables our hospital system to use critical care MDs or APPs who may be high-risk individuals or may have tested positive for COVID-19 without requiring them to enter an ICU or even a hospital.

Effectiveness of Reducing Infection Rate

The effectiveness of a vICU program must be measured in a broader strategic framework that can quantify how the vICU platform has expanded the availability of ICU beds and leveraged the staffing of critical care experts. At HMH, the vICU has allowed the hospital to significantly expand its number of beds without compromising the quality of patient care.

While we did not include empirical data regarding infection control in this study, the inherent structure of delivering critical care remotely protects the vICU providers as well as the consultants and family members, who do not come in contact with the patient with an infectious disease.

Limitations

During the surge period, all the critical care physicians were conscripted to work in the COVID-19 ICUs, leaving the non–COVID-19 patients to be managed by non–critical care physicians and surgeons with support and oversight by critical

care specialists in the vICU. While this adaptive behavior was a good indicator of resilience, the development of protocols informed by simulated and proactive efforts may provide more explicit guidelines for future staffing adjustments. In addition, while the tiered staffing model documented here shows promise for future pandemic surge planning, more work is warranted to improve this model and to develop generalizable guidelines for optimal and flexible allocation of resources during a pandemic.

Conclusions

Telecritical care has been established as an advantageous mechanism for the delivery of critical care expertise. The current COVID-19 pandemic has brought multiple new useful applications to light that could be transformative in how telecritical care is perceived and deployed in the future, especially during highly infectious disease outbreaks. However, further research is required to examine the impact of innovative applications of telecritical care in the treatment of critically ill patients.

Acknowledgments

The authors do not have any sources of funding to report.

Conflicts of Interest

None declared.

References

1. Santangelo G. New NIH resource to analyze COVID-19 literature: The COVID-19 portfolio tool. National Institutes of Health Extramural Nexus. 2020 Apr 15. URL: https://nexus.od.nih.gov/all/2020/04/15/new-nih-resource-to-analyze-covid-19-literature-the-covid-19-portfolio-tool/ [accessed 2020-04-21]
2. Rosenfeld BA, Dorman T, Breslow MJ, Pronovost P, Jenckes M, Zhang N, et al. Intensive care unit telemedicine: alternate paradigm for providing continuous intensivist care. Crit Care Med 2000 Dec;28(12):3925-3931. [doi: 10.1097/00003246-200012000-00034] [Medline: 11153637]
3. Celi LA, Hassan E, Marquardt C, Breslow M, Rosenfeld B. The eICU: it's not just telemedicine. Crit Care Med 2001 Aug;29(8 Suppl):N183-N189. [doi: 10.1097/00003246-200108001-00007] [Medline: 11496041]
4. Kahn JM, Le TQ, Barnato AE, Hravnak M, Kuza CC, Pike F, et al. ICU Telemedicine and Critical Care Mortality: A National Effectiveness Study. Med Care 2016 Mar;54(3):319-325 [FREE Full text] [doi: 10.1097/MLR.0000000000000485] [Medline: 26765148]
5. Udeh C, Udeh B, Rahman N, Canfield C, Campbell J, Hata JS. Telemedicine/Virtual ICU: Where Are We and Where Are We Going? Methodist Debakey Cardiovasc J 2018;14(2):126-133 [FREE Full text] [doi: 10.14797/mdcj-14-2-126] [Medline: 29977469]
6. Subramanian S, Pamplin JC, Hravnak M, Hielsberg C, Riker R, Rincon F, et al. Tele-Critical Care: An Update From the Society of Critical Care Medicine Tele-ICU Committee. Crit Care Med 2020 Apr;48(4):553-561. [doi: 10.1097/CCM.0000000000004190] [Medline: 32205602]
7. Volland J, Fisher A, Drexler D. Delirium and Dementia in the Intensive Care Unit: Increasing Awareness for Decreasing Risk, Improving Outcomes, and Family Engagement. Dimens Crit Care Nurs 2015;34(5):259-264. [doi: 10.1097/DCC.0000000000000133] [Medline: 26244239]
8. Ely EW. The ABCDEF Bundle: Science and Philosophy of How ICU Liberation Serves Patients and Families. Crit Care Med 2017 Feb;45(2):321-330 [FREE Full text] [doi: 10.1097/CCM.0000000000002175] [Medline: 28098628]
9. Marra A, Ely EW, Pandharipande PP, Patel MB. The ABCDEF Bundle in Critical Care. Crit Care Clin 2017 Apr;33(2):225-243 [FREE Full text] [doi: 10.1016/j.ccc.2016.12.005] [Medline: 28284292]
10. Giannini A. The "open" ICU: not just a question of time. Minerva Anestesiol 2010 Feb;76(2):89-90 [FREE Full text] [Medline: 20150847]
11. Halpern NA, Tan KS, Society of Critical Care Medicine Ventilator Taskforce. US ICU resource availability for COVID-19. Version 2. Society of Critical Care Medicine. 2020 Mar 25. URL: https://sccm.org/getattachment/Blog/March-2020/United-States-Resource-Availability-for-COVID-19/United-States-Resource-Availability-for-COVID-19.pdf [accessed 2020-04-21]
12. King MA, Niven AS, Beninati W, Fang R, Einav S, Rubinson L, Task Force for Mass Critical Care. Evacuation of the ICU: care of the critically ill and injured during pandemics and disasters: CHEST consensus statement. Chest 2014 Oct;146(4 Suppl):e44S-e60S [FREE Full text] [doi: 10.1378/chest.14-0735] [Medline: 25144509]
13. Brazzi L, Lissoni A, Panigada M, Bottino N, Patroniti N, Pappalardo F, et al. Simulation-based training of extracorporeal membrane oxygenation during H1N1 influenza pandemic: the Italian experience. Simul Healthc 2012 Feb;7(1):32-34. [doi: 10.1097/SIH.0b013e31823ebccb] [Medline: 22293665]
14. Kowitlawakul Y. The technology acceptance model: predicting nurses' intention to use telemedicine technology (eICU). Comput Inform Nurs 2011 Jul;29(7):411-418. [doi: 10.1097/NCN.0b013e3181f9dd4a] [Medline: 20975536]

15. Azoulay E, Chaize M, Kentish-Barnes N. Involvement of ICU families in decisions: fine-tuning the partnership. Ann Intensive Care 2014 Nov 30;4(1):37 [FREE Full text] [doi: 10.1186/s13613-014-0037-5] [Medline: 25593753]
16. Sasangohar F, Jones SL, Masud FN, Vahidy FS, Kash BA. Provider Burnout and Fatigue During the COVID-19 Pandemic: Lessons Learned From a High-Volume Intensive Care Unit. Anesth Analg 2020 Jul;131(1):106-111 [FREE Full text] [doi: 10.1213/ANE.0000000000004866] [Medline: 32282389]

Abbreviations

APP: advanced practice provider
AV: audiovisual
CICU: cardiac intensive care unit
CVICU: cardiovascular intensive care unit
eICU: electronic intensive care unit
EMR: electronic medical record
HIPAA: Health Insurance Portability and Accountability Act
HMH: Houston Methodist Hospital
ICU: intensive care unit
MICU: medical intensive care unit
Neuro-ICU: neurology intensive care unit
PPE: personal protective equipment
TMC: Texas Medical Center
vICU: virtual intensive care unit
vMD: virtual medical doctor
vRN: virtual registered nurse

Edited by G Eysenbach; submitted 12.05.20; peer-reviewed by SC Pan, K Chen, D Peña; comments to author 13.07.20; revised version received 19.07.20; accepted 10.08.20; published 03.09.20.

Please cite as:
Dhala A, Sasangohar F, Kash B, Ahmadi N, Masud F
Rapid Implementation and Innovative Applications of a Virtual Intensive Care Unit During the COVID-19 Pandemic: Case Study
J Med Internet Res 2020;22(9):e20143
URL: http://www.jmir.org/2020/9/e20143/
doi:10.2196/20143
PMID:32795997

Original Paper

Depression and Psychological-Behavioral Responses Among the General Public in China During the Early Stages of the COVID-19 Pandemic: Survey Study

Weiyu Zhang[1], BM; Xiaoting Yang[2], AD; Jinfeng Zhao[3], PhD; Fengzhi Yang[1], BM; Yajing Jia[1], BA; Can Cui[1], BA; Xiaoshi Yang[1], PhD

[1]Department of Social Medicine, School of Public Health, China Medical University, Shenyang, China

[2]Department of Gastroenterology, Benxi General Hospital, Liaoning Health Industry Group, Benxi, China

[3]The University of Auckland, Auckland, New Zealand

Corresponding Author:
Xiaoshi Yang, PhD
Department of Social Medicine, School of Public Health
China Medical University
No 77 Puhe Road
Shenyang North New Area
Shenyang, 110122
China
Phone: 86 18900910796
Email: xsyang@cmu.edu.cn

Abstract

Background: The COVID-19 pandemic has recently spread dramatically worldwide, raising considerable concerns and resulting in detrimental effects on the psychological health of people who are vulnerable to the disease. Therefore, assessment of depression in members of the general public and their psychological and behavioral responses is essential for the maintenance of health.

Objective: This study aimed to assess the prevalence of depression and the associated factors among the general public during the early stages of the COVID-19 pandemic in China.

Methods: A cross-sectional survey with convenience sampling was conducted from February 11 to 16, 2020, in the early stages of the COVID-19 outbreak in China. A self-administrated smartphone questionnaire based on the Patient Health Questionnaire-9 (PHQ-9) and psychological and behavioral responses was distributed to the general public. Hierarchical multiple regression analysis and multivariate logistic regression analysis were conducted to explore the associated factors of depression.aA cross-sectional survey with convenience sampling was conducted from February 11 to 16, 2020, in the early stages of the COVID-19 outbreak in China. A self-administrated smartphone questionnaire based on the Patient Health Questionnaire-9 (PHQ-9) and psychological and behavioral responses was distributed to the general public. Hierarchical multiple regression analysis and multivariate logistic regression analysis were conducted to explore the associated factors of depression.

Results: The prevalence of depression (PHQ-9 score ≥10) among the general public during the COVID-19 pandemic was 182/1342 (13.6%). Regression analysis indicated that feeling stressed, feeling helpless, persistently being worried even with support, never feeling clean after disinfecting, scrubbing hands and items repeatedly, hoarding food, medicine, or daily supplies, and being distracted from work or study were positively associated with depression, while social support and being calm were negatively associated with depression.

Conclusions: The general public suffered from high levels of depression during the early stages of the COVID-19 pandemic. Thus, COVID-19–related mood management and social support should be provided to attenuate depression in the general public.

(J Med Internet Res 2020;22(9):e22227) doi:10.2196/22227

KEYWORDS

depression; COVID-19; social support; the general public

Introduction

COVID-19 is a highly infectious disease that was identified at the end of 2019 and has been listed as Public Health Emergency of International Concern by the World Health Organization [1,2]. COVID-19 rapidly spread to 31 provinces, municipalities, and autonomous regions in China and then spread worldwide, developing into a serious global pandemic and resulting in harm to public health, the global economy, and social development. As of 4 PM CST, July 2, 2020, 10,726,907 confirmed cases and 514,458 deaths have been reported from 26 countries around the world; 85,264 confirmed cases and 4648 deaths have been reported in China [3,4].

The unexpected outbreak of COVID-19 resulted in adverse impacts on the general public, restricting normal activities and disrupting festive plans. A series of immediate emergency responses were initiated by Chinese government to prevent the spread of COVID-19. The responses included sealing off cities and provinces, restrictions on travel, control of civilian air traffic, self-quarantine, delaying the resumption of work and school, and setting up targeted hospitals to receive and treat patients. The rapid spread of COVID-19 and uncertainty related to the disease have created a great deal of stress and difficult predicaments for the public; this has resulted in the emergence of mental disorders, such as depression and anxiety [5]. Scarcities of medicine and personal protective equipment (eg, antibiotics, masks, alcohol sanitizer) and basic food supplies (eg, cereals and vegetables) during the earliest weeks of the pandemic seriously affected the mental health of the public [6,7]. Recent studies revealed that extensive exposure to COVID-19 pandemic stress can render the public vulnerable to depression [8].

The psychological and behavioral responses to the COVID-19 pandemic can be vulnerability factors of mental disorders during pandemics [9,10]. Previous research has indicated that adverse psychological and behavioral responses to public health emergencies may exacerbate the spread of disease and induce the development of mental disorders such as acute stress disorder, panic attacks, posttraumatic stress disorder, anxiety and depression, and even suicidal tendencies [11-17]. However, the patterns of psychological responses to pandemics are various and unpredictable [18], and they can have considerable influences on the psychological health of the public. A great number of previous studies indicate that negative emotionality, maladaptive cognitive style, and low levels of optimism are associated with the risks of negative psychological impacts [9,19,20]. People who overestimate threats or lack tolerance of uncertainty in pandemics can become highly worried and anxious during pandemics [21]. Perception of stress and helplessness arising from the stressful situations of the epidemic could have direct effects on the mental health of individuals and lead to prevalence of depression and anxiety [22,23]. Moreover, psychoneuroimmunology studies indicate that people with predisposition to emotional disorders may be particularly vulnerable as a result of immune responses [24,25]. However, positive emotions such as optimism play a moderating role in the relationship between stress and depression, and lower optimism is associated with affective disorders [26,27]. Additionally, it has been observed that during pandemics, people can share negative emotions, including disappointment, anger, stress, fear, and anxiety, in social life; this is also known as emotional contagion [28].

According to the cognitive-behavioral model of health anxiety, people who experience excessively high health anxiety are likely to become highly anxious during a pandemic and engage in maladaptive behaviors, such as excessive hand washing [9,18]. Studies of the concept of the behavioral immune system indicate that people can respond to pandemics excessively (ie, to medically unnecessary levels) and may avoid infection-related contaminants [9,29]. The information-motivation-behavior skill model and the modified behavioral framework mentioned in recent studies indicate that motivation, behavior skills and positive perception of risk have direct effects on health behaviors [30,31]. Conversely, to maintain health and avoid infection, the general public may also behave irrationally (eg, superstitious behaviors, overdose of vitamins, and use of herbal supplements or even folk remedies) when people are excessively worried or anxious during pandemics [32,33]. In a recent study, according to the conceptual model of Stimulus-Organism-Response, governance such as lockdown measures taken to cut off channels of infection decreased psychological distance, leading to a buffer effect on perceived risks and anxiety [34]. Adversity in pandemics, such as economic depression and shortage of supplements, can lead to antisocial behaviors such as violence, rioting, looting, and even civil unrest and mass panic [35]. Scientific prediction and appropriate guidance of public behaviors have important implications for people's mental health [36,37].

Social support, which is perceived as care or help from others, has a protective effect on depression by regulating stress, as presented in previous studies [38,39], and it is closely associated with the generation, control, and prevention of mental disorders. Adequate social support and appropriate sources of support are beneficial to the public health because they can release stress, maintain individuals' emotional responses, and prevent mental symptoms [40].

The COVID-19 pandemic has resulted in adverse effects on public mental health, which has recently raised considerable concerns [41,42]. Studies have been conducted to assess and prevent psychological health crises during the outbreak of COVID-19 [43]. As a result of the pandemic, the public has reported discomforts such as heavy mental health burdens, poor sleep quality, and psychological distress [44-48]. Therefore, in our study, we aimed to appraise depression and explore its associated factors, including psychological responses, behavioral responses, and social support, among the general public during the early stages of the COVID-19 pandemic. Our findings can be used to provide evidence-based advice on early psychological and behavioral interventions to reduce depression in the general public.

Methods

Participants, Procedure, and Ethics Statement

A cross-sectional survey with convenience sampling was conducted in mainland China from February 11 to 16, 2020. A web link and quick response code were distributed via WeChat, a widely used social network platform in China, to collect self-administrated questionnaires with support from the Environmental Health Institute at China Medical University. Participants anonymously completed a questionnaire in 15 to 20 minutes comprising questions based on the Patient Health Questionnaire-9 (PHQ-9) and concerning psychological and behavioral responses to the COVID-19 pandemic. Participants included in this study met the following inclusion criteria: aged ≥18 years; able to read and write Chinese; able to use WeChat to complete the questionnaire independently; willing to participate and provide signed web-based informed consent. Any participant meeting one or more of the following criteria was excluded: receiving treatment for any psychological illness; having significant visual impairment; having a history of drug dependence; having been diagnosed with a disease that would prevent them from completing the questionnaire independently.

This survey recruited a total of 1675 adults from the general public in 3 municipalities and 22 provinces or autonomous regions in China. Each participant was well informed of the aims, funding, and contents of the questionnaire and the commitment to the privacy of the participants. Questionnaires that were answered completely and logically were regarded as valid. A total of 1342/1675 participants provided valid answers to the questionnaire, resulting in a valid response rate of 80.12%. This study conformed to ethical standards and was conducted in accordance with the Helsinki Declaration as revised in 1989. The Ethics Committee of China Medical University approved the protocols of this study.

Demographic Characteristics of the Participants

The demographic characteristics of the participants that were collected included gender (male, female), age (≤35 years, >35 years), marital status (married, other), occupation, education, and monthly income. Age was categorized according to proportionality, population distribution, and previous studies on the relationship between age and use of WeChat [49,50]. Occupation was categorized as government worker/civil servant/village committee worker/enterprise employee, health care worker, student, teacher/lawyer/journalist, and other. Education was categorized as below junior college, junior college, and bachelor's degree and above. Monthly income (RMB) was classified as ≤¥5000 (≤US $725.19), ¥5001 to ¥10,000 (US $725.34 to $1,450.39), and >¥10,000 (>US $1,450.39).

Measurement of Depression

Depression was assessed using the 9-item PHQ-9, which is one of the most widely used tools to assess depression [51]. Each item asked about the situation in the past two weeks using a 4-point Likert-type scale with choices of "not at all," "a few days," "more than half of the two weeks," and "nearly every day," giving a total score between 0 and 21. A score above 10 is regarded to be indicative of depression [52-55]. The Cronbach α coefficient of the PHQ-9 in this study was .91.

Measurement of Psychological Responses

Psychological responses were measured by self-developed questions. Psychological responses included "feeling stressed," "feeling helpless," "persistently being worried even with support," "being calm," and "being optimistic." Answers were grouped into three categories of "disagree," "not sure," and "agree" to assess the participants' psychological responses during the past two weeks.

Measurement of Behavioral Responses

Behavioral responses were measured by self-developed questions. Options for reflecting behavioral responses for this study included "never feeling clean after disinfecting; scrubbing hands and items repeatedly" (answers were "no," "sometimes," and "always"), "hoarding food, medicine, or daily supplies," "social avoidance to avoid infection," and "being distracted from work or study." Answers were grouped into 3 categories of "disagree," "not sure," and "agree" to evaluate the participants' behavioral responses during the past two weeks.

Measurement of Social Support

Social support was assessed by a "yes or no" question that asked whether people had received social support in the past two weeks.

Statistical Analyses

All analyses were performed using SPSS version 23.0 for Windows (IBM Corporation). A two-tailed probability value <.05 was considered statistically significant. We used *t* tests and one-way analysis of variance to compare differences in depression among categorical variables. Hierarchical multiple regression (HMR) analysis was conducted to test incremental variance using the following independent variables: Step 1: demographic characteristics of the general public; Step 2: psychological responses to the COVID-19 pandemic; Step 3: behavioral responses to the COVID-19 pandemic; and Step 4: social support. The depression scores were continuous in HMR and used as the dependent variables. Standardized parameter estimates (standardized β) were used to compare the magnitude of associations of the independent variables. The depression scores in the multivariate logistic regression analysis were binary (depression or no depression) with a cutoff score of 10. Multivariate logistic regression analysis was performed to explore risk factors associated with depression using odds ratios (ORs). Among all categorically independent variables, items for which more than 95% of individuals had the same response were not included in the data analysis. A two-tailed *P* value <.05 was considered to be statistically significant.

Results

Demographic Characteristics and Depression Distribution of the Participants

A total of 1342 subjects participated in this study, giving a valid response rate of 1342/1675 (80.12%). The distribution of the demographic characteristics of the participants, prevalence of

depression, and univariate analysis of depression are listed in Table 1. The prevalence of depression among the general public was 182/1342 (13.60%).

The 1342 participants included 500 men (37.26%) and 842 women (62.74%). Approximately one-third of the participants (467/1342, 35.77%) had a bachelor's degree or higher degree. Depression scores for the participants who were aged <35 years were significantly higher than those of the other participants (P=.001). Participants who were married tended to have lower depression scores than participants who were not (P<.001). Government workers, civil servants, village committee workers, enterprise employees, and health care workers had lower depression scores than those engaging in other occupations (P<.001).

Table 1. Characteristics of the general public and distribution of depression in China during the early stages of the COVID-19 pandemic (N=1342).

Variable	Total, n (%)	Depression, n (%)	No depression, n (%)	Depression score, mean (SD)	*P* value
Demographic characteristics					
Gender					.30
Male	500 (37.26)	66 (13.20)	434 (86.80)	4.04 (5.68)	
Female	842 (62.74)	116 (13.78)	726 (86.22)	4.35 (5.13)	
Age (years)					.001
≤35	597 (44.49)	98 (16.42)	499 (83.58)	4.80 (5.57)[a]	
>35	745 (55.51)	84 (11.28)	661 (88.72)	3.78 (5.11)	
Marital status					<.001
Married	898 (66.92)	107 (11.92)	791 (88.08)	3.85 (5.07)	
Other	444 (33.09)	75 (16.89)	369 (83.11)	5.03 (5.77)[a]	
Occupation					<.001
Government worker, civil servant, village committee worker, or enterprise employee	377 (28.09)	34 (9.02)	343 (90.98)	3.56 (4.99)	
Health care worker	279 (20.79)	30 (10.75)	249 (89.25)	3.59 (4.71)	
Student	203 (15.13)	37 (18.23)	166 (81.77)	5.00 (5.29)[a]	
Teacher, lawyer, or journalist	244 (18.18)	37 (15.16)	207 (84.84)	4.48 (5.36)[b]	
Other	239 (17.81)	44 (18.41)	195 (81.59)	5.19 (6.30)[a]	
Education					.08
Below junior college	166 (12.37)	165 (15.15)	141 (84.85)	4.41 (5.74)	
Junior college	709 (52.83)	107 (15.09)	602 (84.91)	4.49 (5.67)[b]	
Bachelor’s degree and above	467 (34.80)	50 (10.71)	417 (89.29)	3.80 (4.62)	
Monthly income (¥)[c]					.006
≤5000	445 (33.16)	77 (17.30)	368 (82.70)	3.72 (4.98)	
5001-10,000	535 (39.87)	53 (9.91)	482 (90.09)	3.21 (4.17)	
>10,000	362 (26.97)	52 (14.36)	310 (85.64)	3.54 (4.44)	
Psychological Responses					
Feeling stressed					<.001
Disagree	370 (27.57)	16 (4.32)	354 (95.68)	1.72 (3.48)	
Not sure	634 (47.24)	63 (9.94)	571 (90.06)	3.85 (4.48)[a]	
Agree	338 (25.19)	103 (30.47)	235 (69.53)	7.74 (6.60)[a]	
Feeling helpless					<.001
Disagree	554 (41.28)	103 (18.59)	451 (81.41)	2.44 (4.18)	
Not sure	443 (33.01)	49 (11.06)	394 (88.94)	4.09 (4.61)[a]	
Agree	345 (25.71)	30 (8.70)	315 (91.30)	7.32 (6.41)[a]	
Persistently being worried even with support					<.001
Disagree	648 (48.29)	24 (3.70)	624 (96.30)	2.05 (3.46)	
Not sure	313 (23.32)	37 (11.82)	276 (88.18)	4.30 (4.67)[a]	
Agree	381 (28.39)	121 (31.76)	260 (68.24)	7.91 (6.39)[a]	

Variable	Total, n (%)	Depression, n (%)	No depression, n (%)	Depression score, mean (SD)	*P* value
Being calm					<.001
Agree	521 (38.82)	33 (6.33)	488 (93.67)	2.74 (4.44)	
Not sure	548 (40.83)	73 (13.32)	475 (86.68)	4.36 (4.94)[a]	
Disagree	273 (20.35)	76 (27.84)	197 (72.16)	6.86 (6.53)[a]	
Being optimistic					<.001
Agree	410 (30.55)	39 (9.51)	371 (90.49)	3.32 (5.10)	
Not sure	425 (31.67)	42 (9.88)	383 (90.12)	3.63 (4.50)	
Disagree	507 (37.78)	101 (19.92)	406 (80.08)	5.50 (5.99)[a]	
Behavioral responses					
Never feeling clean after disinfecting; scrubbing hands and items repeatedly					<.001
No	543 (40.46)	35 (6.45)	508 (93.55)	2.73 (4.41)	
Sometimes	402 (29.96)	60 (14.93)	342 (85.07)	4.83 (4.96)[a]	
Always	397 (29.58)	87 (21.91)	310 (78.09)	5.71 (6.28)[a]	
Hoarding food, medicine, or daily supplies					<.001
Disagree	595 (44.34)	55 (9.24)	540 (90.76)	3.07 (4.60)	
Not sure	492 (36.66)	57 (11.59)	435 (88.41)	4.27 (4.66)[a]	
Agree	255 (19.00)	70 (27.45)	185 (72.55)	6.89 (7.00)[a]	
Social avoidance to avoid infection					.053
Disagree	275 (20.49)	36 (13.09）	239 (86.91)	3.89 (5.24)	
Not sure	547 (40.76)	63 (11.52)	484 (88.48)	3.99 (4.76)	
Agree	520 (38.75)	83 (15.96)	437 (84.04)	4.68 (5.93)	
Being distracted from work or study					<.001
Disagree	443 (33.01)	3 (0.68)	440 (99.32)	1.23 (2.33)	
Not sure	379 (28.24)	32 (8.44)	347 (91.56)	3.27 (3.90)[a]	
Agree	520 (38.75)	147 (28.27)	373 (71.73)	7.51 (6.24)[a]	
Social support					**<**.001
Yes	1198 (89.27)	146 (12.19)	1052 (87.81)	4.00 (5.08)	
No	144 (10.73)	36 (25.00)	108 (75.00)	6.21 (6.89)[a]	

[a]Significant at the 0.01 level (two-tailed).

[b]Significant at the 0.05 level (two-tailed).

[c]1 ¥ = US $0.15.

Psychological and Behavioral Responses to the COVID-19 Pandemic

The distribution of depression scores according to psychological and behavioral responses to the COVID-19 pandemic are shown in Table 1. The participants who disagreed that they felt stressed or helpless tended to have lower depression scores than those who agreed or were not sure (*P*<.001). Depression scores were lower among participants who were persistently worried even with support (*P*<.001). The depression scores for the individuals who agreed that they were calm or optimistic about the COVID-19 pandemic were significantly lower (*P*<.001).

The participants who never felt clean after disinfecting and who scrubbed their hands and items repeatedly had significantly higher depression scores (*P*<.001). Participants who agreed that they hoarded food, medicine, or daily supplies tended to have significantly higher depression scores (*P*<.001). Scores were higher for the participants who agreed that they were distracted from work or study (*P*<.001).

The results also showed that participants who reported that they did not receive any support from others when they had difficulties during the COVID-19 pandemic had a significantly higher score of depression than those who received social support (*P*<.001).

Risk Factors of Depression

Table 2 shows the final results of the HMR models of depression among the general public. A total of 42.1% of the variance was explained by the final model. The R^2 changes indicated that the incremental variances explained by each block of variables were 3.1%, 28.6%, 10.1%, and 0.3% for demographic characteristics, psychological responses, behavioral responses, and social support, respectively. In the final model of the HMR and the forest plot (Figure 1), the risk factors of depression included feeling stressed; feeling helpless; persistently being worried, even with support; never feeling clean after disinfecting, and scrubbing hands and items repeatedly; hoarding food, medicine, or daily supplies; and being distracted from work or study (all $P<.01$). However, being calm and social support were negatively associated with depression.

Table 2. Hierarchical linear regression analysis of depression among the general population in China during the early stages of the COVID-19 pandemic.

Variable	β	Standardized β	95% CI	*P* value	Adjusted R^2	ΔR^2
Demographic characteristics					0.023	0.031
Gender (male vs female)	–0.429	–0.039	–0.926 to 0.068	.09		
Age (>35 years vs ≤35 years)	–0.082	–0.008	–0.675 to 0.511	.79		
Marital status (married vs other)	–0.669	–0.059		.06		
Occupation						
Health care worker	0.407	0.031	–0.259 to 1.073	.23		
Student	0.277	0.019	–0.599 to 1.154	.54		
Teacher, lawyer, or journalist	0.745	0.054	0.066 to 1.425	.03[a]		
Other	0.502	0.036	–0.188 to 1.192	.15		
Education						
Below junior college vs bachelor's degree and above	0.437	0.027	–0.299 to 1.174	.25		
Junior college vs bachelor's degree and above	0.433	0.040	–0.053 to 0.919	.08		
Monthly income (¥)[b]						
≤5000 vs >10,000	0.382	0.034	–0.226 to 0.991	.22		
5001-10,000 vs >10,000	–0.231	–0.021	–0.791 to 0.330	.42		
Psychological responses					0.306	0.286
Feeling stressed						
Not sure vs disagree	–0.161	–0.015	–0.805 to 0.482	.62		
Agree vs disagree	1.631	0.133	0.813 to 2.448	<.001[c]		
Feeling helpless						
Not sure vs disagree	0.665	0.059	0.097 to 1.234	.02[a]		
Agree vs disagree	1.717	0.141	1.076 to 2.358	<.001[c]		
Persistently being worried even with support						
Not sure vs disagree	0.822	0.065	0.156 to 1.488	.016[a]		
Agree vs disagree	2.016	0.170	1.299 to 2.733	<.001[c]		
Being calm						
Agree vs disagree	–1.365	–0.125	–2.081 to –0.649	<.001[c]		
Not sure vs disagree	–0.477	–0.044	–1.115 to 0.160	.14		
Being optimistic						
Agree vs disagree	0.318	0.027	–0.296 to 0.933	.31		
Not sure vs disagree	–0.407	–0.035	–0.968 to 0.153	.15		
Behavioral responses					0.404	0.101
Never feeling clean after disinfecting; scrubbing hands and items repeatedly						
Sometimes vs no	0.175	0.015	–0.398 to 0.748	.55		
Always vs no	0.803	0.069	0.231 to 1.374	.006[c]		
Hoarding food, medicine, or daily supplies						
Not sure vs disagree	–0.157	–0.014	–0.683 to 0.369	.56		
Agree vs disagree	1.393	0.102	0.686 to 2.099	<.001[c]		
Social avoidance to avoid infection						

Variable	β	Standardized β	95% CI	*P* value	Adjusted R^2	ΔR^2
Not sure vs disagree	–0.421	–0.039	–1.060 to 0.218	.20		
Agreed vs disagree	–0.121	–0.011	–0.793 to 0.550	.72		
Being distracted from work or study						
Not sure vs disagree	0.951	0.080	0.311 to 1.590	.004[c]		
Agree vs disagree	3.842	0.351	3.219 to 4.464	<.001[c]		
Social support (yes vs no)	–0.943	–0.055	–1.682 to –0.205	.012[a]	0.406	0.003

[a]Significant at the .05 level (two-tailed).

[b]1 ¥ = US $0.15.

[c]Significant at the .01 level (two-tailed).

Figure 1. Forest plot of the risk factors of depression (hierarchical multiple regression).

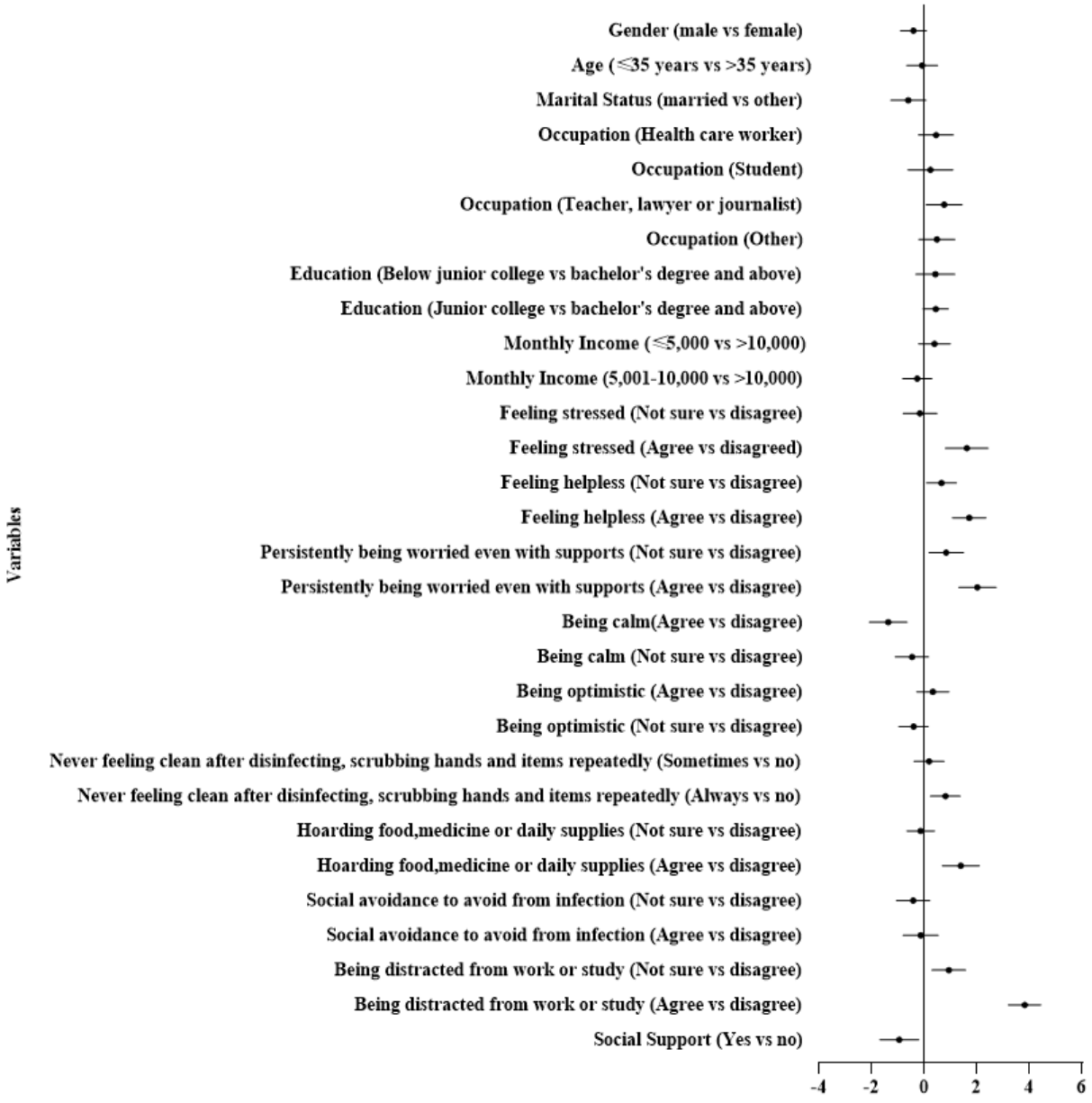


Table 3 presents the results of the multivariate logistic regression analysis. Feeling helpless (OR 2.341, 95% CI 1.367-4.009), persistently being worried even with support (OR 3.315, 95% CI 1.696-6.479), never feeling clean after disinfecting and

scrubbing hands and items repeatedly (OR 1.941, 95% CI 1.153-3.267), hoarding food, medicine, or daily supplies (OR 1.822, 95% CI 1.012-3.279), and being distracted from work or study (OR 27.225, 95% CI 8.243-89.918) increased the risk of depression. However, being calm (OR 0.344, 95% CI 0.186-0.635) and social support (OR 0.529, 95% CI 0.308-0.908) decreased the chance of suffering from depression. Additionally, compared with other occupations, teachers, lawyers, and journalists (OR 2.053, 95% CI 1.117-3.776) had higher chances of depression. In the final model of the multivariate logistic regression and the forest plot (Figure 2), the risk factors of depression included occupation of teacher, lawyer, or journalist; junior college education; feeling helpless; persistently being worried, even with support; never feeling clean after disinfecting and scrubbing hands and items repeatedly; hoarding food, medicine, or daily supplies; and being distracted from work or study (all $P<.01$). However, being calm and social support were protective factors of depression.

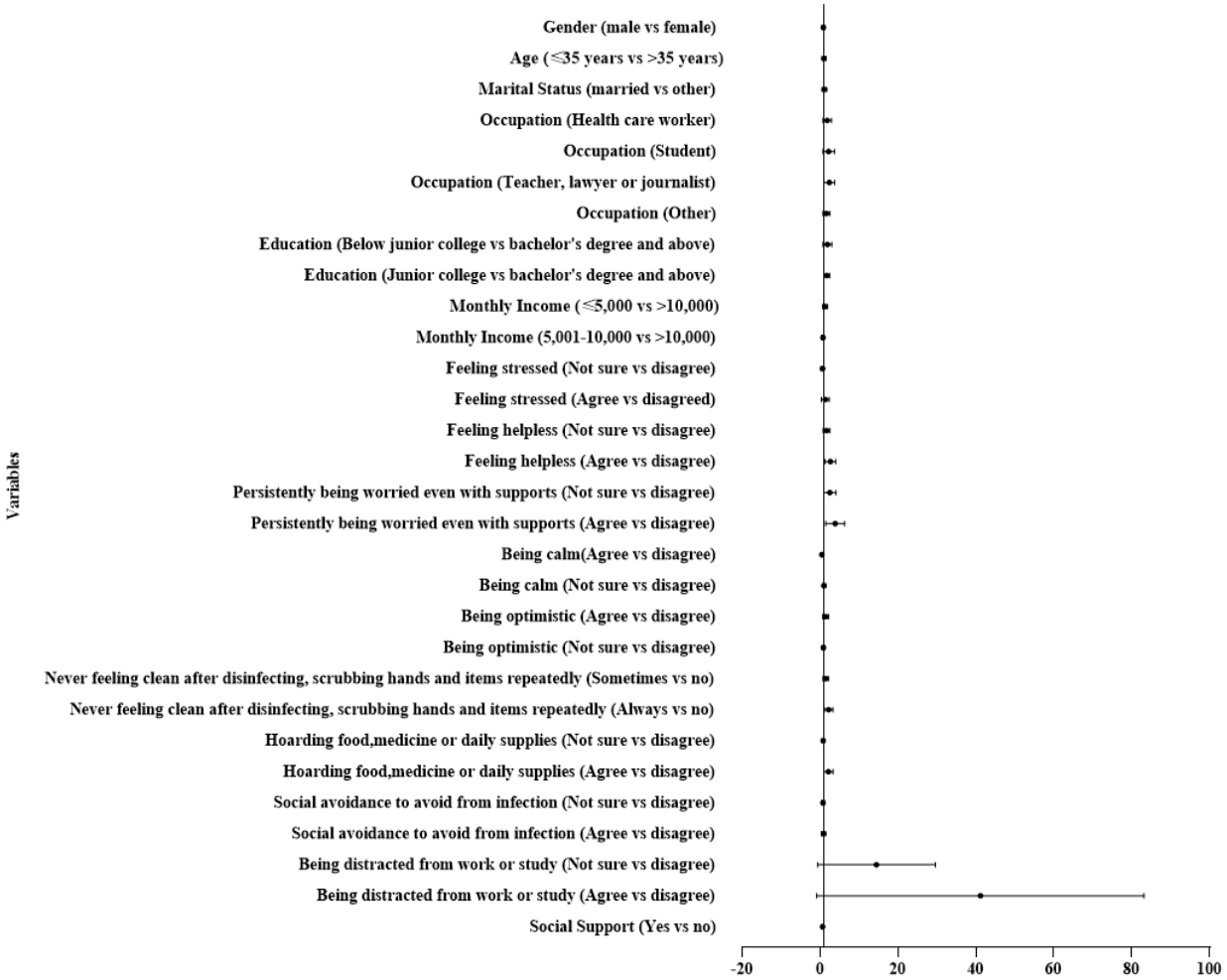
Table 3. Multivariate logistic regression analysis exploring factors of depression among the general population in China during the early stages of the COVID-19 pandemic.

Variable	Odds ratio	95% CI
Demographic characteristics		
Gender (male vs female)	0.742	0.478-1.150
Age (>35 years vs ≤35 years)	0.885	0.532-1.474
Marital status (married vs other)	0.936	0.531-1.651
Occupation		
Health care worker	1.530	0.804-2.914
Student	1.810	0.872-3.754
Teacher, lawyer, or journalist	2.049	1.115-3.768
Other	1.403	0.765-2.573
Education		
Below junior college vs bachelor's degree and above	1.596	0.839-3.035
Junior college vs bachelor's degree and above	1.572	1.018-2.428
Monthly income (¥)[a]		
≤5000 vs >10,000	1.160	0.697-1.931
5001-10,000 vs >10,000	0.617	0.374-1.018
Psychological responses		
Feeling stressed		
Not sure vs disagree	0.465	0.216-1.003
Agree vs disagree	1.018	0.436-2.376
Feeling helpless		
Not sure vs disagree	1.455	0.826-2.562
Agree vs disagree	2.348	1.373-4.016
Persistently being worried even with support		
Not sure vs disagree	2.120	1.081-4.160
Agree vs disagree	3.303	1.696-6.433
Being calm		
Agree vs disagree	0.345	0.188-0.636
Not sure vs disagree	0.876	0.553-1.386
Being optimistic		
Agree vs disagree	1.262	0.729-2.186
Not sure vs disagree	0.713	0.443-1.149
Behavioral responses		
Never feeling clean after disinfecting; scrubbing hands and items repeatedly		
Sometimes vs no	1.265	0.746-2.145
Always vs no	1.942	1.158-3.257
Hoarding food, medicine, or daily supplies		
Not sure vs disagree	0.655	0.397-1.082
Agree vs disagree	1.862	1.039-3.338
Social avoidance to avoid infection		
Not sure vs disagree	0.619	0.347-1.104
Agree vs disagree	0.746	0.404-1.379

Variable	Odds ratio	95% CI
Being distracted from work or study		
Not sure vs disagree	9.131	2.648-31.495
Agree vs disagree	26.810	8.128-88.430
Social support (yes vs no)	0.524	0.305-0.901

[a]1 ¥ = US $0.15.

Figure 2. Forest plot of the risk factors of depression (multivariate logistic regression).



Discussion

Principal Findings

This study revealed that 13.6% of the general public suffered from depression during the early stages of the pandemic of COVID-19 in China. The prevalence of depression was slightly lower than the prevalence before the outbreak of COVID-19 from previous studies [56-58] (13.8% vs 18.8%). The results also indicated that the prevalence of depression during the COVID-19 pandemic is almost identical to that during the SARS epidemic (11.9% vs 18.0%) among the general public in China [59,60]. Most of the research on mental disorders during public health emergencies in China focuses on the psychological status of patients or health care workers; depression in the general public is neglected and untreated [61-64]. The general public are extremely vulnerable to symptoms of depression during the COVID-19 pandemic. People with mental symptoms of depression should be identified and managed in a timely manner to improve their psychological health status [65].

Depression is closely associated with the psychological and behavioral responses of the general public during the COVID-19 pandemic. This study found that psychological responses contributed the most (28.6%) to the variance of depression. Feeling stressed, feeling helpless, and feeling persistently worried even with support were positively associated with depression, which is in agreement with previous studies indicating that the increase of perceived susceptibility to the epidemic is associated with the prevalence of mental disorders [66]. Perception of the risks of infection is a possible reason for negative psychological responses to pandemics and is highly

correlated with distress in individuals [67-69]. Additionally, poor self-rated health status can be causative and can result in psychological burdens and excessive worries about the pandemic [70,71]. In addition, individuals who are worried even when they have support may be more prone to behavior problems, which can exacerbate the symptoms of depression [72,73].

Notably, being calm and optimistic are negatively associated with depression. Calmness and optimism can help people cope with problems rationally and may mediate the negative effects of perception of epidemic-related stress and indirectly attenuate the symptoms of depression [74]. On the other hand, positive cognition of public health emergencies has close correlations with stress and social functions [75]. Additionally, positive psychological capabilities such as calmness and optimism can ameliorate depression, in particular for people who experience higher levels of helplessness [76]. Lessened psychological capabilities can result in high susceptibility to anxiety and depression during negative emotional experiences or periods of adversity such as the COVID-19 pandemic. Thus, psychological management and regulation of the responses to a public health emergency may benefit from consciously controlling stress, helplessness, and worry and from maintaining calm and optimism.

This study showed that depression was significantly linked to behavioral responses. More than half of the participants (799/1342, 59.54%) reported that they never felt clean after they disinfected and scrubbed hands and items repeatedly, and 1067/1342 (79.51%) of these members of the general public practiced social avoidance to avoid the infection of COVID-19. Preventive behaviors are key to reducing the spread and impact of a pandemic. However, excessively preventive and avoidant behavioral responses, which possibly resulted from the perceived stress induced by the COVID-19 pandemic, were positively associated with depression. Recent studies have reported that psychological impacts, including depression, anxiety, and stress, are linked with the adoption of precautionary measures to prevent the spread of COVID-19 [16]. For example, using medical preventive equipment and washing after touching contaminated surfaces can provide potential psychological benefits by offering a sense of security and comport for the general public and can prevent the spread of COVID-19. However, these precautionary measures can also have negative effects on individuals' mental health when they become excessive [44,77]. People may feel more susceptible to becoming infected during pandemics. Negative psychological impacts revealed by these behaviors can be easily popularized among the general public. This finding is consistent with other studies showing that protective behaviors such as avoiding sharing of utensils during meals and washing hands frequently increase the likelihood of stress, anxiety, and depression [78]. Health professionals should pay more attention to these behaviors, which can have adverse impacts on mental health, with a view to realizing early warning signs and conducting psychological interventions to protect vulnerable groups [79].

Specifically, social support appeared to be a protective factor for depression. Social support can enhance self-esteem and positive adaptation to stress and combat the detrimental effects of stress or worry on the development of depression [44]. Social support such as positive communications and entertainment may help individuals to recover quickly the from the COVID-19 outbreak [80]. In contrast, lack of social support can result in hostility and uncertainty in social life, thus increasing susceptibility to serious mental disorders, particularly depression [16]. Therefore, adequate and appropriate social support should be provided to prevent the development of mental disorders in the general public.

Limitations

Several limitations should be acknowledged when presenting our findings. First, this survey was conducted with convenience sampling, which limits our ability to generalize the results of this study to the overall population. Second, this study was based on a self-administrated questionnaire completed via smartphone, which limited the diversity of the sample and the authenticity of the answers. Depression and the associated factors revealed in this study were limited to the early stages of the COVID-19 pandemic.

Conclusion

The general public in China suffered from high levels of depression during the early stages of the COVID-19 pandemic. Being stressed, feeling helpless or worried, and being distracted from work or study increased the risks of depression. People who experience these adverse psychological responses should be provided with psychological interventions to reduce the symptoms of depression. Moreover, positive emotions and psychological responses such as calmness and optimism decreased the risks of depression and should be encouraged by health professionals and educators. Behavioral responses, including repeated cleaning, hoarding supplies, and distraction from work or study, were not only influencing factors of depression but also expressions of depression. Therefore, early detection and psychological and behavioral interventions should be developed to help people cope rationally with pandemic-related stress and improve the mental health of the general public.

Acknowledgments

This work was supported by the Research on Behaviors Rules and Emotions and of Public during the Epidemic of COVID-19, Public Behavior Research Project of Prevention and Control of COVID-19, China Medical University (grant number: 121-1210120025). The authors are very grateful to all the study participants and research assistants for data provision and collection.

Authors' Contributions

WZ contributed to the acquisition and analysis of data and the drafting and revision of the manuscript. JZ contributed to the revision of the manuscript. Xiaoting Y contributed to the acquisition of data. FY, YJ, and CC contributed to the acquisition and interpretation of data. Xiaoshi Y was responsible for the conception and design and contributed to the revision of the manuscript.

Conflicts of Interest

None declared.

References

1. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
2. 2020 WHO Timeline-COVID-19. World Health Organization. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-08-26]
3. 2020 Coronavirus disease (COVID-19) Situation Report-148. World Health Organization. 2020 Jun 16. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200616-covid-19-sitrep-148-draft.pdf?sfvrsn=9b2015e9_2 [accessed 2020-08-26]
4. Real-time big data of the pandemic of COVID-19. Webpage in Chinese. National Health Commission of the People's Republic of China. URL: https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_3#tab0 [accessed 2020-08-26]
5. Van Bortel T, Basnayake A, Wurie F, Jambai M, Koroma AS, Muana AT, et al. Psychosocial effects of an Ebola outbreak at individual, community and international levels. Bull World Health Organ 2016 Mar 01;94(3):210-214 [FREE Full text] [doi: 10.2471/BLT.15.158543] [Medline: 26966332]
6. Klompas M, Morris CA, Sinclair J, Pearson M, Shenoy ES. Universal Masking in Hospitals in the Covid-19 Era. N Engl J Med 2020 May 21;382(21):e63 [FREE Full text] [doi: 10.1056/NEJMp2006372] [Medline: 32237672]
7. Devi S. Travel restrictions hampering COVID-19 response. Lancet 2020 Apr 25;395(10233):1331-1332 [FREE Full text] [doi: 10.1016/S0140-6736(20)30967-3] [Medline: 32334692]
8. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]
9. Taylor S. The Psychology of Pandemics: Preparing for the Next Global Outbreak of Infectious Disease. Newcastle upon Tyne, UK: Cambridge Scholars; 2019.
10. Allden K, Jones L, Weissbecker I, Wessells M, Bolton P, Betancourt T, et al. Mental health and psychosocial support in crisis and conflict: report of the Mental Health Working Group. Prehosp Disaster Med 2009;24 Suppl 2:s217-s227 [FREE Full text] [doi: 10.1017/s1049023x00021622] [Medline: 19806544]
11. Shultz JM, Baingana F, Neria Y. The 2014 Ebola outbreak and mental health: current status and recommended response. JAMA 2015 Feb 10;313(6):567-568 [FREE Full text] [doi: 10.1001/jama.2014.17934] [Medline: 25532102]
12. Zhang W, Lee L, Connor KM, Chang C, Lai T, Davidson JR. Symptoms of neurasthenia following earthquake trauma: re-examination of a discarded syndrome. Psychiatry Res 2007 Oct 31;153(2):171-177 [FREE Full text] [doi: 10.1016/j.psychres.2006.04.021] [Medline: 17451809]
13. Wu KK, Chan SK, Ma TM. Posttraumatic stress, anxiety, and depression in survivors of severe acute respiratory syndrome (SARS). J Trauma Stress 2005 Feb;18(1):39-42 [FREE Full text] [doi: 10.1002/jts.20004] [Medline: 16281194]
14. Liu T, Chen X, Miao G, Qian M, He Y, Yu X, et al. Recommendations on diagnostic criteria and prevention of SARS-related mental disorders. Article in Chinese. Journal of Clinic Psychology Medicine 2003;13(3):188-191. [doi: 10.3969/j.issn.1005-3220.2003.03.043]
15. Hawkey PM, Bhagani S, Gillespie SH. Severe acute respiratory syndrome (SARS): breath-taking progress. J Med Microbiol 2003 Aug;52(Pt 8):609-613 [FREE Full text] [doi: 10.1099/jmm.0.05321-0] [Medline: 12867552]
16. Ko CH, Yen CF, Yen JY, Yang MJ. Psychosocial impact among the public of the severe acute respiratory syndrome epidemic in Taiwan. Psychiatry Clin Neurosci 2006 Aug;60(4):397-403 [FREE Full text] [doi: 10.1111/j.1440-1819.2006.01522.x] [Medline: 16884438]
17. Vijaya K, Low YY, Chan SP, Foo LL, Lee M, Deurenberg-Yap M. Behaviour of Singaporeans during the SARS outbreak: The impact of anxiety and public satisfaction with media information. Int J Health Promot Educ 2005 Jan;43(1):17-22 [FREE Full text] [doi: 10.1080/14635240.2005.10708030]
18. Goodwin R, Gaines SO, Myers L, Neto F. Initial psychological responses to swine flu. Int J Behav Med 2011 Jun 2;18(2):88-92 [FREE Full text] [doi: 10.1007/s12529-010-9083-z] [Medline: 20195809]
19. Timpano KR, Buckner JD, Richey JA, Murphy DL, Schmidt NB. Exploration of anxiety sensitivity and distress tolerance as vulnerability factors for hoarding behaviors. Depress Anxiety 2009 Apr;26(4):343-353 [FREE Full text] [doi: 10.1002/da.20469] [Medline: 19123454]

20. Ferguson E. Hypochondriacal concerns and the five factor model of personality. J Pers 2000 Aug;68(4):705-724 [FREE Full text] [doi: 10.1111/1467-6494.00113] [Medline: 10934687]
21. Green JS, Teachman BA. Predictive Validity of Explicit and Implicit Threat Overestimation in Contamination Fear. J Obsessive Compuls Relat Disord 2013 Jan 01;2(1):1-8 [FREE Full text] [doi: 10.1016/j.jocrd.2012.09.002] [Medline: 24073390]
22. Wang J, Li Y, Chen J, Liang W, Yang T, Lee Y, et al. Investigating the Relationships among Stressors, Stress Level, and Mental Symptoms for Infertile Patients: A Structural Equation Modeling Approach. PLoS One 2015 Oct 20;10(10):e0140581 [FREE Full text] [doi: 10.1371/journal.pone.0140581] [Medline: 26484531]
23. Maier SF. Exposure to the stressor environment prevents the temporal dissipation of behavioral depression/learned helplessness. Biol Psychiatry 2001 May 01;49(9):763-773 [FREE Full text] [doi: 10.1016/s0006-3223(00)01095-7] [Medline: 11331084]
24. Cacioppo J, Tassinary L, Berntson G. The Handbook of Psychophysiology, 3rd Edition. Cambridge, UK: Cambridge University Press; 2007.
25. Kiecolt-Glaser JK. Psychoneuroimmunology: Psychology's Gateway to the Biomedical Future. Perspect Psychol Sci 2009 Jul;4(4):367-369 [FREE Full text] [doi: 10.1111/j.1745-6924.2009.01139.x] [Medline: 19750133]
26. Romswinkel E, König HH, Hajek A. The role of optimism in the relationship between job stress and depressive symptoms. Longitudinal findings from the German Ageing Survey. J Affect Disord 2018 Dec 01;241:249-255 [FREE Full text] [doi: 10.1016/j.jad.2018.08.005] [Medline: 30138809]
27. Broekhof R, Rius-Ottenheim N, Spinhoven P, van der Mast RC, Penninx BW, Zitman FG, et al. Long-lasting effects of affective disorders and childhood trauma on dispositional optimism. J Affect Disord 2015 Apr 01;175:351-358 [FREE Full text] [doi: 10.1016/j.jad.2015.01.022] [Medline: 25665495]
28. Scheve C, Salmela M. Collective Emotions. Oxford, UK: Oxford University Press; 2014.
29. Taylor S, Asmundson GJG. Treatment of Health Anxiety. In: The Wiley Handbook of Obsessive Compulsive Disorders, Volume II. Hoboken, NJ: Wiley; 2017:977-989.
30. Luo Y, Yao L, Zhou L, Yuan F, Zhong X. Factors influencing health behaviours during the coronavirus disease 2019 outbreak in China: an extended information-motivation-behaviour skills model. Public Health 2020 Aug;185:298-305 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.057] [Medline: 32717671]
31. Ahmad M, Iram K, Jabeen G. Perception-based influence factors of intention to adopt COVID-19 epidemic prevention in China. Environ Res 2020 Jul 27;190:109995 [FREE Full text] [doi: 10.1016/j.envres.2020.109995] [Medline: 32739626]
32. Ariely D. Predictably Irrational: The Hidden Forces That Shape Our Decisions. New York, NY: Harper Collins; 2008.
33. Lee JD. An Epidemic of Rumors: How Stories Shape Our Perception of Disease. Logan, UT: Utah State University Press; 2014.
34. Zheng L, Miao M, Lim J, Li M, Nie S, Zhang X. Is Lockdown Bad for Social Anxiety in COVID-19 Regions?: A National Study in The SOR Perspective. Int J Environ Res Public Health 2020 Jun 24;17(12):4561 [FREE Full text] [doi: 10.3390/ijerph17124561] [Medline: 32599911]
35. Shultz J, Espinel Z, Flynn B, Hoffmann Y, Cohen R. DEEP PREP All-Hazards Disaster Behavioral Health Training. Coral Gables, FL: DEEP Center, University of Miami; 2008.
36. Dezecache G. Human collective reactions to threat. Wiley Interdiscip Rev Cogn Sci 2015 Feb 20;6(3):209-219 [FREE Full text] [doi: 10.1002/wcs.1344] [Medline: 26263225]
37. Ursano RJ, Norwood AE, Fullerton CS. Bioterrorism: Psychological and Public Health Interventions. Cambridge, UK: Cambridge University Press; Mar 17, 2005:595-596.
38. He F, Guan H, Kong Y, Cao R, Peng J. Some Individual Differences Influencing the Propensity to Happiness: Insights from Behavioral Economics. Soc Indic Res 2013 Dec 8;119(2):897-908 [FREE Full text] [doi: 10.1007/s11205-013-0519-0]
39. Wang X, Cai L, Qian J, Peng J. Social support moderates stress effects on depression. Int J Ment Health Syst 2014;8(1):41 [FREE Full text] [doi: 10.1186/1752-4458-8-41] [Medline: 25422673]
40. Maulik PK, Eaton WW, Bradshaw CP. The role of social network and support in mental health service use: findings from the Baltimore ECA study. Psychiatr Serv 2009 Sep;60(9):1222-1229 [FREE Full text] [doi: 10.1176/ps.2009.60.9.1222] [Medline: 19723737]
41. Dalton L, Rapa E, Stein A. Protecting the psychological health of children through effective communication about COVID-19. Lancet Child Adolesc Health 2020 May;4(5):346-347 [FREE Full text] [doi: 10.1016/s2352-4642(20)30097-3]
42. Yang Y, Li W, Zhang Q, Zhang L, Cheung T, Xiang Y. Mental health services for older adults in China during the COVID-19 outbreak. Lancet Psychiat 2020 Apr;7(4):e19 [FREE Full text] [doi: 10.1016/S2215-0366(20)30079-1] [Medline: 32085843]
43. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addiction 2020 Mar 27:6353-6359. [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]
44. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
45. Choi EPH, Hui BPH, Wan EYF. Depression and Anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740 [FREE Full text] [doi: 10.3390/ijerph17103740] [Medline: 32466251]

46. Gómez-Salgado J, Andrés-Villas M, Domínguez-Salas S, Díaz-Milanés D, Ruiz-Frutos C. Related Health Factors of Psychological Distress During the COVID-19 Pandemic in Spain. Int J Environ Res Public Health 2020 Jun 02;17(11):3947 [FREE Full text] [doi: 10.3390/ijerph17113947] [Medline: 32498401]
47. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
48. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
49. Yang X, Kovarik CL. A systematic review of mobile health interventions in China: Identifying gaps in care. J Telemed Telecare 2019 Jul 18:1357633X19856746. [doi: 10.1177/1357633X19856746] [Medline: 31319759]
50. Cai B, Xu Z. Investigation on reading behaviors of medical science articles in Wechat public platform. Article in Chinese. Journal of Public Health Management 2016 Jun:783-787.
51. Kroenke K, Spitzer RL. The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
52. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. Gen Hosp Psychiatry 2015;37(1):67-75 [FREE Full text] [doi: 10.1016/j.genhosppsych.2014.09.009] [Medline: 25439733]
53. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry 2015;37(6):567-576 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.06.012] [Medline: 26195347]
54. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
55. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 09;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
56. Zhao D, Hu C, Chen J, Dong B, Ren Q, Yu D, et al. Risk factors of geriatric depression in rural China based on a generalized estimating equation. Int Psychogeriatr 2018 Oct;30(10):1489-1497 [FREE Full text] [doi: 10.1017/S1041610218000030] [Medline: 29380720]
57. Lu S, Reavley N, Zhou J, Su J, Pan X, Xiang Q, et al. Depression among the general adult population in Jiangsu Province of China: prevalence, associated factors and impacts. Soc Psychiatry Psychiatr Epidemiol 2018 Oct 30;53(10):1051-1061 [FREE Full text] [doi: 10.1007/s00127-018-1568-0] [Medline: 30062483]
58. Ryder AG, Sun J, Zhu X, Yao S, Chentsova-Dutton YE. Depression in China: integrating developmental psychopathology and cultural-clinical psychology. J Clin Child Adolesc Psychol 2012;41(5):682-694 [FREE Full text] [doi: 10.1080/15374416.2012.710163] [Medline: 22900498]
59. Tao J, Gong F, Lu J, Liang L, Liao X. Meta-analysis on the prevalence of depressive and /or anxiety disorder in general hospitals in China. Article in Chinese. Sichuan Mental Health 2018;31:73-78. [doi: 10.11886/j.issn.1007-3256.2018.01.018]
60. Lee DTS, Sahota D, Leung TN, Yip ASK, Lee FFY, Chung TKH. Psychological responses of pregnant women to an infectious outbreak: a case-control study of the 2003 SARS outbreak in Hong Kong. J Psychosom Res 2006 Nov;61(5):707-713 [FREE Full text] [doi: 10.1016/j.jpsychores.2006.08.005] [Medline: 17084150]
61. Smith GD, Ng F, Ho Cheung Li W. COVID-19: Emerging compassion, courage and resilience in the face of misinformation and adversity. J Clin Nurs 2020 May;29(9-10):1425-1428 [FREE Full text] [doi: 10.1111/jocn.15231] [Medline: 32155302]
62. Zhang W, Wang K, Yin L, Zhao W, Xue Q, Peng M, et al. Mental Health and Psychosocial Problems of Medical Health Workers during the COVID-19 Epidemic in China. Psychother Psychosom 2020;89(4):242-250 [FREE Full text] [doi: 10.1159/000507639] [Medline: 32272480]
63. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiat 2020 Apr;7(4):e15-e16 [FREE Full text] [doi: 10.1016/S2215-0366(20)30078-X] [Medline: 32085839]
64. Liu Y, Cao L, Li X, Jia Y, Xia H. Awareness of mental health problems in patients with coronavirus disease 19 (COVID-19): A lesson from an adult man attempting suicide. Asian J Psychiatr 2020 Jun;51:102106 [FREE Full text] [doi: 10.1016/j.ajp.2020.102106] [Medline: 32334411]
65. Anmella G, Arbelo N, Fico G, Murru A, Llach C, Madero S, et al. COVID-19 inpatients with psychiatric disorders: Real-world clinical recommendations from an expert team in consultation-liaison psychiatry. J Affect Disord 2020 Sep 01;274:1062-1067 [FREE Full text] [doi: 10.1016/j.jad.2020.05.149] [Medline: 32663933]
66. Yao H, Chen J, Xu Y. Patients with mental health disorders in the COVID-19 epidemic. Lancet Psychiat 2020 Apr;7(4):e21 [FREE Full text] [doi: 10.1016/S2215-0366(20)30090-0] [Medline: 32199510]
67. Schibalski JV, Müller M, Ajdacic-Gross V, Vetter S, Rodgers S, Oexle N, et al. Stigma-related stress, shame and avoidant coping reactions among members of the general population with elevated symptom levels. Compr Psychiatry 2017 Apr;74:224-230 [FREE Full text] [doi: 10.1016/j.comppsych.2017.02.001] [Medline: 28236772]

68. Goulia P, Mantas C, Dimitroula D, Mantis D, Hyphantis T. General hospital staff worries, perceived sufficiency of information and associated psychological distress during the A/H1N1 influenza pandemic. BMC Infect Dis 2010 Nov 09;10:322 [FREE Full text] [doi: 10.1186/1471-2334-10-322] [Medline: 21062471]
69. Maguire PA, Reay RE, Looi JC. A sense of dread: affect and risk perception in people with schizophrenia during an influenza pandemic. Australas Psychiatry 2019 Oct;27(5):450-455 [FREE Full text] [doi: 10.1177/1039856219839467] [Medline: 30983379]
70. Alkhamees AA, Alrashed SA, Alzunaydi AA, Almohimeed AS, Aljohani MS. The psychological impact of COVID-19 pandemic on the general population of Saudi Arabia. Compr Psychiatry 2020 Jul 12;102:152192 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152192] [Medline: 32688022]
71. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Feb 07;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
72. Zainal NH, Newman MG. Worry amplifies theory-of-mind reasoning for negatively valenced social stimuli in generalized anxiety disorder. J Affect Disord 2018 Feb;227:824-833 [FREE Full text] [doi: 10.1016/j.jad.2017.11.084] [Medline: 29254067]
73. Fanti KA, Hellfeldt K, Colins OF, Meehan A, Andershed A, Andershed H. Worried, sad, and breaking rules? Understanding the developmental interrelations among symptoms of anxiety, depression, and conduct problems during early childhood. J Crim Justice 2019 May;62:23-28 [FREE Full text] [doi: 10.1016/j.jcrimjus.2018.09.006]
74. Heinen I, Bullinger M, Kocalevent R. Perceived stress in first year medical students - associations with personal resources and emotional distress. BMC Med Educ 2017 Jan 06;17(1):4 [FREE Full text] [doi: 10.1186/s12909-016-0841-8] [Medline: 28056972]
75. Trindade IA, Mendes AL, Ferreira NB. The moderating effect of psychological flexibility on the link between learned helplessness and depression symptomatology: A preliminary study. J Contextual Behav Sci 2020 Jan;15:68-72 [FREE Full text] [doi: 10.1016/j.jcbs.2019.12.001]
76. Leung GM, Lam T, Ho L, Ho S, Chan BHY, Wong IOL, et al. The impact of community psychological responses on outbreak control for severe acute respiratory syndrome in Hong Kong. J Epidemiol Community Health 2003 Nov;57(11):857-863 [FREE Full text] [doi: 10.1136/jech.57.11.857] [Medline: 14600110]
77. Ho RC, Zhang MW, Ho CS, Pan F, Lu Y, Sharma VK. Impact of 2013 south Asian haze crisis: study of physical and psychological symptoms and perceived dangerousness of pollution level. BMC Psychiatry 2014 Mar 19;14:81 [FREE Full text] [doi: 10.1186/1471-244X-14-81] [Medline: 24642046]
78. Li S, Wang Y, Xue J, Zhao N, Zhu T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
79. Roohafza HR, Afshar H, Keshteli AH, Mohammadi N, Feizi A, Taslimi M, et al. What's the role of perceived social support and coping styles in depression and anxiety? J Res Med Sci 2014 Oct;19(10):944-949 [FREE Full text] [Medline: 25538777]
80. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of Prevalence and Associated Factors of Anxiety and Depression Among People Affected by versus People Unaffected by Quarantine During the COVID-19 Epidemic in Southwestern China. Med Sci Monit 2020 Apr 26;26:e924609 [FREE Full text] [doi: 10.12659/MSM.924609] [Medline: 32335579]

Abbreviations

HMR: hierarchical multiple regression
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by EPH Choi; comments to author 17.07.20; revised version received 05.08.20; accepted 10.08.20; published 04.09.20.

Please cite as:
Zhang W, Yang X, Zhao J, Yang F, Jia Y, Cui C, Yang X
Depression and Psychological-Behavioral Responses Among the General Public in China During the Early Stages of the COVID-19 Pandemic: Survey Study
J Med Internet Res 2020;22(9):e22227
URL: https://www.jmir.org/2020/9/e22227
doi:10.2196/22227
PMID:32886066

Original Paper

Perinatal Distress During COVID-19: Thematic Analysis of an Online Parenting Forum

Bonnie R Chivers[1*], MHumNutr; Rhonda M Garad[1*], PhD; Jacqueline A Boyle[1,2], PhD; Helen Skouteris[1], PhD; Helena J Teede[1,3], PhD; Cheryce L Harrison[1], PhD

[1]Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

[2]Department of Obstetrics and Gynaecology, Monash Health, Clayton, Australia

[3]Diabetes and Vascular Medicine Unit, Monash Health, Clayton, Australia

[*]these authors contributed equally

Corresponding Author:
Cheryce L Harrison, PhD
Monash Centre for Health Research and Implementation
School of Public Health and Preventive Medicine
Monash University
Wellington Rd
Clayton
Australia
Phone: 61 3 8572 2662
Email: cheryce.harrison@monash.edu

Abstract

Background: The COVID-19 global pandemic has impacted the whole of society, requiring rapid implementation of individual-, population-, and system-level public health responses to contain and reduce the spread of infection. Women in the perinatal period (pregnant, birthing, and postpartum) have unique and timely needs for directives on health, safety, and risk aversion during periods of isolation and physical distancing for themselves, their child or children, and other family members. In addition, they are a vulnerable group at increased risk of psychological distress that may be exacerbated in the context of social support deprivation and a high-risk external environment.

Objective: The aim of this study is to examine the public discourse of a perinatal cohort to understand unmet health information and support needs, and the impacts on mothering identity and social dynamics in the context of COVID-19.

Methods: A leading Australian online support forum for women pre- through to postbirth was used to interrogate all posts related to COVID-19 from January 27 to May 12, 2020, inclusive. Key search terms included “COVID,” “corona,” and “pandemic.” A three-phase analysis was conducted, including thematic analysis, sentiment analysis, and word frequency calculations.

Results: The search yielded 960 posts, of which 831 were included in our analysis. The qualitative thematic analysis demonstrated reasonable understanding, interpretation, and application of relevant restrictions in place, with five emerging themes identified. These were (1) heightened distress related to a high-risk external environment; (2) despair and anticipatory grief due to deprivation of social and family support, and bonding rituals; (3) altered family and support relationships; (4) guilt-tampered happiness; and (5) family future postponed. Sentiment analysis revealed that the content was predominantly negative (very negative: n=537 and moderately negative: n=443 compared to very positive: n=236 and moderately positive: n=340). Negative words were frequently used in the 831 posts with associated derivatives including “worried” (n=165, 19.9%), “risk” (n=143, 17.2%), “anxiety” (n=98, 11.8%), “concerns” (n=74, 8.8%), and “stress” (n=69, 8.3%).

Conclusions: Women in the perinatal period are uniquely impacted by the current pandemic. General information on COVID-19 safe behaviors did not meet the particular needs of this cohort. The lack of nuanced and timely information may exacerbate the risk of psychological and psychosocial distress in this vulnerable, high-risk group. State and federal public health departments need to provide a central repository of information that is targeted, consistent, accessible, timely, and reassuring. Compensatory social and emotional support should be considered, using alternative measures to mitigate the risk of mental health disorders in this cohort.

(J Med Internet Res 2020;22(9):e22002) doi:10.2196/22002

KEYWORDS

pregnancy; perinatal; maternal; COVID-19; communication; social support; qualitative research; mental health; health information; online support; thematic analysis; sentiment analysis; word frequency

Introduction

COVID-19, a novel strain of coronavirus, is an acute, highly infectious virus that has affected tens of millions of people and has caused close to 1 million deaths as of August 2020 [1]. The disease is spread directly through respiratory droplets from the mouth or nose. It is estimated that up to 40% of transmission is presymptomatic, with an average incubation period of 5-6 days [2]. The vast majority of those who are infected with COVID-19 experience mild to moderate illness arising from a cluster of mild symptoms including fever, dry cough, and lethargy [3]. Currently, there is no effective COVID-19 medical prophylaxis and limited treatment, mandating a rigorous individual-, population-, and system-level public health policy and behavioral change response to minimize transmission [4].

COVID-19 originated in Wuhan, China, with the first cases reported in December 2019 [5], and was officially declared as a public health emergency of international concern on January 30, 2020, by the World Health Organization. The first Australian confirmed case of COVID-19 was on January 25, 2020, in Victoria, originating from Wuhan, China, subsequently resulting in the closing of Australian borders to all nonresidents in March [6]. Physical distancing rules were imposed on March 21, 2020, with the associated closure of all *nonessential* services including retail outlets, cafés, restaurants, schools, recreational facilities, and playgrounds [7]. To assist in physical distancing, additional measures including working, studying, or completing school from home were imposed; social gatherings were banned, and stringent restrictions on individual movement were put in place [7]. These public health policies profoundly impacted individual- and population-level health, disrupted normal social interactions, and contributed to economic insecurity.

As COVID-19 continues to disrupt human interactions, published data on the risk, transmission, and health outcomes of specific populations, including women in the perinatal period and their neonate, are evolving, yet are currently inconclusive [8-10]. Women within the perinatal period are a vulnerable population, both physiologically, with changes during pregnancy that reduce immunity [11], and psychologically, with increased risk of psychological distress including stress, anxiety, and depression [12-15], all of which may increase maternal and neonatal morbidity [16]. Consequently, it is recognized that women during this period require unique and specific health, support, and information needs to avoid stress [17]. Currently, the impact of COVID-19 on such needs and associated levels of distress is poorly understood, yet is critical in ascertaining adverse implications, as well as in identifying strategies to protect and optimize women's psychosocial health during this time. Therefore, this study aims to understand the sentiment and impacts to emotional well-being as well as the unmet information and support needs arising from changes to social dynamics and support in a perinatal cohort during the COVID-19 pandemic.

Methods

Overview

We conducted an observational, qualitative analysis of online discussions within a leading Australian forum for new or expecting parents. The most popular Australian pre- and postbirth forum was identified by searching the term "new mum forum" in Google. The top 10 (first page) results were assessed, and all websites with publicly available forums (n=7) were analyzed using a website analytics tool (Alexa, Amazon.com). This software was used to determine the global page views, global rank, and Australian rank of the 7 websites with publicly available forums. The highest ranked website for Australian users was identified and used as the sampling platform for this study. To confirm this website's suitability for this study, member requirements were assessed to ensure forum users were new or expecting mothers.

Within the selected forum, the search function was used to identify user-generated content relating to COVID-19. No date restrictions were applied in the search with posts collected on May 12, 2020. We searched for website content using key search terms including "COVID," "corona," or "pandemic." The search identified articles, comments, and posts, which was then narrowed to posts. All posts were extracted in a deidentified format into a Word (Microsoft Corporation) document, which included the post title, date, and content. The inclusion criteria were posts related to COVID-19 up until May 12, 2020 (inclusive). The exclusion criteria were posts that included a title only (content had been deleted), that did not relate to COVID-19, and that were duplicate posts (original post was collected once).

Analysis

Data was processed using NVivo Pro 12 (QSR International) software [18]. Analysis comprised three phases, including thematic analysis, sentiment analysis, and word frequency calculations of stemmed words. Thematic analysis was undertaken using a modified grounded theory approach that was informed by Braun and Clarke's [19] six phase approach. A single researcher (BC) became familiar with the data, generated initial codes, and searched for themes. The team then collaborated to discuss themes and a >25% check of themes was conducted by two additional researchers (CH and RG). To support the themes identified, NVivo Pro 12 automatic sentiment analysis and a text frequency search were run to identify emotional indicators. NVivo searched for expressions of sentiment in the source material then used predefined scores for words classified as containing sentiment [20]. Words are considered in isolation, and the program then determines the sentiment of the paragraph as a calculation of each word containing the sentiment. Sentiment results include the number of references (paragraphs with sentiment) that are categorized as very positive (VP), moderately positive (MP), moderately negative (MN), and very negative (VN). A single researcher

(BC) conducted a >10% cross-check of the sentiment results. Word frequency calculations were used to identify all stemmed words (minimum 3 letters) used 50 times or more. These words were screened by a single researcher (BC) to identify negative words. The frequency of key terms used was divided by the number of total posts to derive an overall percentage, and a weighted percentage on the total word count was calculated by NVivo Pro 12.

Ethics

Ethics approval for this study was granted by the Monash Health (RES-19-0000-291A) and Monash University (Project no 20196) Human Research Ethics Committees. Although ethical oversight of publicly available data is not strictly required, the authors sought approval as per Monash University protocol.

Results

Overview

A total of 960 posts were identified using the search terms ("corona" n=589, "COVID" n=257, and "pandemic" n=114). As per the exclusion criteria, 114 posts were excluded, resulting in a final sample of 831 unique posts. The first relevant post identified was dated January 27, 2020.

Thematic Analysis

We identified five themes from the analyzed content: (1) heightened distress related to a high-risk external environment; (2) despair and anticipatory grief due to deprivation of social and family support, and bonding rituals; (3) altered family and support relationships; (4) guilt-tampered happiness; and (5) family future postponed.

Theme 1: Heighted Distress Related to a High-Risk External Environment

Women expressed concerns and unease due to a range of factors such as the lack of access to particular information on risk (eg, risk during pregnancy, risk to baby in utero, risk to a new born baby, risk from the hospital environment, risk to mental health from reduced social supports). They asked questions within the discussion forum, such as should pregnant women cease working, what is the risk of COVID-19 to unborn or newly born babies, should working partners isolate from their pregnant spouse or from new babies, are pregnant women at increased risk, or should I be leaving the house? Many women were unable to locate information to fully answer these concerns and sought confirmation of both their concerns and their risk reduction actions from their peers within the forum.

Women who stated they were close to giving birth demonstrated significant levels of worry in relation to the safety of antenatal appointments and the hospital environment. Some indicated they were considering a home birth, and others stated they were considering not attending antenatal visits due to fear of contracting COVID-19.

> *I am due to get my NT scan done...I am getting nervous about having to travel there from a small rural town to have it done. I am considering whether I should get it done or not worry.*

One woman stated that, although she did not want to be "alarmist" the lack of information on hospital safety made her feel "vulnerable." Others stated that "being pregnant during this time is so scary...it's certainly not how I envisioned it to be going or to end," and another woman said, "I almost don't want to give birth right now."

Women said they found it difficult to disengage with the constant worry and concern about COVID-19 with one woman stating that

> *...all I'm thinking and dreaming about is COVID-19! I'm super paranoid, I feel like even getting groceries, I have anxiety.*

Many women said they had disengaged with mainstream news sources as they found these to be fear inducing. Posts reflected heightened levels of worry and stress in this cohort. One woman stated, "I feel the most anxious, overwhelmed, isolated, out of control as I've ever felt before."

Theme 2: Despair and Anticipatory Grief due to Deprivation of Social and Family Support, and Bonding Rituals

Due to the unforeseen restrictions from the pandemic, usual perinatal social rituals such as baby showers, celebrations, or gender reveals were not possible. In addition, areas where women heavily rely on social support, such as multiple support persons during birth and postbirth, and family or social support providing physical, psychological, and social care, were denied to this group of women. When one woman was told she was not allowed to have her partner with her during a routine scan she said, "I am really sad as it's our first pregnancy and these scans are small ways where our husbands can take part in our pregnancy." Many expressed a range of feelings such as sadness, anger, and a sense of loss. One woman stated "the happiest time in our lives is being over shadowed by this virus." Another said "I've played many times over, my siblings and parents coming to hospital to meet my first child and that was a once in a life time moment for me; something I think very special that I won't get anymore." Upon reading about the COVID-19–related grief, one woman said:

> *totally made me realise that I'm grieving the loss of a normal pregnancy...I checked out the five stages of grief, and I'm definitely going through the emotions they list! Anger and bargaining - hopefully I'll move on to true acceptance soon.*

Theme 3: Altered Family and Support Relationships

Customary ways of strengthening relationships with family and social networks during pregnancy and early motherhood were denied to this group of women due to the demands of isolation. Many of the women discussed conflicts with family members over interpretations of physical distancing rules. This conflict was expressed by one woman as:

> *I don't understand how they whinge that they can't see their grandson yet continue to do all the things THAT ARE THE REASON we don't want our son around them...Has anyone else had to deal with family not taking this virus seriously?*

The impact of COVID-19 isolating new mothers from their support and family networks, and the unique need to protect the health of grandparents reduced social support for new mothers and resulted, in some cases, in unanticipated interfamily conflict. This had the effect of dividing some families rather than the bonding experiences that many may have anticipated.

Theme 4: Guilt-Tampered Happiness

Many women expressed feelings of guilt due to the contrast between the positivity and happiness they were experiencing at the news of their pregnancy or arrival of their child, and the difficult situation of many others in the community.

> *I feel a bit guilty about wanting to share my good news. I...have been looking forward to sharing about our baby especially after I was told I would not be able to have kids at all.*

Other women identified guilt relating to their wishes to experience baby showers or events that their peers have formerly enjoyed:

> *this might sound a little selfish but it's our first and I was looking forward to the gifts and games etc. I hope by August maybe we can have a small gathering.*

Although another woman stated she was feeling guilty for thinking about her pregnancy plans. "I hope everyone is coping with the current covid crisis - I feel a bit guilty thinking about IVF."

Theme 5: Family Future Postponed

Many women discussed their concerns about family planning and of wanting to postpone plans to extend their family. "I've heard so much information (and misinformation) the last few weeks it's really made me fearful about trying to conceive with all the madness around." Another posed the questions:

> *[is anyone] temporarily pausing trying to conceive whilst our world is being rattled by corona? Has it changed your timeline in terms of when you want baby to born? Have you decided to wait or maybe not even have anymore?*

Sentiment and Word Frequency Analysis

There were 1556 references of sentiment. Of these, most were found to be within the negative range, with 980 (63.0%) references classified as negative (VN: n=537 and MN: n=443) compared with 576 positive references (VP: n=236 and MP: n=340). A cross-check of sentiment results found that less than 10% of sentiment references were coded to incorrect categories.

A list of words used 50 times or more (n=189) was examined to screen for negative words. Multiple negative emotive stem words (n=7) were identified (Table 1). Percentages derived from the total number of posts provide an indication of the spread of these words within the data set.

Table 1. Word frequency in total posts (N=831).[a]

Word	Similar word	Frequency, n (%)	Weighted percentage (%)
Worrying	Worried, worries, worry	165 (19.9)	0.32
Risk	Risking, risks	143 (17.0)	0.27
Anxiety	Anxieties, anxious, anxiousness	98 (11.7)	0.19
Concerns	Concern, concerned, concerning	74 (8.8)	0.14
Stress	Stressed, stressful, stressing	69 (8.2)	0.13
Struggling	Struggle, struggled	59 (7.0)	0.11
Scare	Scares, scared, scaring	55 (6.6)	0.11

[a]Percentage calculations were derived based on the word count in context of the total posts (identifying an approximate number of posts including these words; however, this cannot account for words used multiple times in one post) and as a weighted percentage of the total word count.

Discussion

Principal Results

The rapidly evolving COVID-19 pandemic is a global emergency, requiring unprecedented public health policy changes and behavioral adherence to limit viral spread. This vital public health response comes at significant health, societal, and economic cost, and is anticipated to have a profound effect on emotional well-being and mental health. During the perinatal period, women have specific health care and emotional needs, and are vulnerable to mental health challenges. Understanding the unique perinatal information needs and impact of losing support networks is, therefore, critical in informing public health and health care system's interventions for the well-being of women and their families.

Our results demonstrate a significant response to the pandemic within this cohort, with an average of 275 posts per month directly related to COVID-19 across the evaluation period. Thematic analysis captured five themes: heightened distress related to a high-risk external environment; despair and anticipatory grief due to deprivation of social and family support, and bonding rituals; altered family and support relationships; guilt-tampered happiness; and a family future postponed. Sentiment analysis showed heightened negativity with high frequency use of negative terms including "stress," "anxious," "worry," and "risk," with emergent qualitative themes identified related to anxiety, grief, guilt, social support, and disrupted family planning.

To enable rapid evaluation and dissemination of findings, we chose to examine a leading online Australian perinatal forum,

accessed broadly by reproductive-aged women across preconception, pregnancy, postpartum, and into early parenting and childhood. Online support groups have previously been shown to be commonly accessed by women during this period. They provide an accessible, peer-to-peer opportunity for information sharing, individualized information seeking, and social and emotional support from women in a comparable life stage [21,22]. Accessibility and anonymity provided by online forums facilitates a disinhibited expression of feelings without concern of consequences or conflict from close connections. Engagement with and preference for interacting with peers in an online forum may reduce engagement with other forms of media. Indeed, this is supported by our thematic analysis showing many women reporting disengagement with mainstream news sources, as they found these to be fear-provoking.

Anticipatory grief for loss of social and family affirming opportunities during the perinatal period was emphasized as a consequence of physical distancing measures in this cohort. Pregnancy and birth are often a celebratory period for women and are associated with joy experienced by both the parents and close family members. Traditional rituals and milestones including the early pregnancy growth and development scans, baby shower celebrations, and hospital visits by family and grandparents were all effectively denied in accordance with physical distancing and personal safety measures. Such milestones present significant opportunities for social support and likely play a role in strengthening family and partner relationships, in turn increasing well-being, self-efficacy, and coping during this time. Disruption or removal of these milestones is, therefore, likely to impact well-being, potentially increasing feelings of isolation and consequently driving conflict due to anger and frustration, exacerbating pre-existing negativity within the wider external environment.

Conversely, many women also expressed guilt as a result of feeling joy and happiness related to their pregnancy or impending birth. This *paradox of guilt* in the context of COVID-19 is reportedly similar to the phenomenon of survivor's guilt, an experience of immense guilt toward surviving a traumatic event that others in a given population did not [23]. Here, women felt happiness yet felt exclaiming such feelings were inappropriate in the context of widespread illness, grief, and loss at a population level. Although this may not be directly termed as survivor's guilt, a correlation is plausible. Women may also be disproportionately affected by feelings of guilt in this context, with research showing increased concern for the health of others including older family members and children, rather than the health and well-being of themselves [24].

Our results show an interruption in family planning as a consequence of COVID-19. This may be due to several factors including economic and employment insecurity or fear of direct health impacts. The sudden closure of many sectors has profoundly impacted the economy, contributing to the largest rise in unemployment rate since the depression, with women disproportionately impacted [25]. Hesitancy related to health outcomes during pregnancy may also impact family planning. Although women are generally more resistant to viral infections than men, physiological changes that occur during pregnancy increase vulnerability to severe infections due to a reduced immune response. A recent systemic review of the first 108 pregnant women infected with COVID-19 found a majority of case reports occurred in the last trimester with associated severe maternal morbidity [8]. This included a cesarean section rate of 91% across cases, predominantly due to fetal distress, with 1 intrauterine and 1 neonatal death recorded. Despite the majority of women recovering without major complication, the authors concluded that mother-to-baby transmission of COVID-19 remains unclear [8], with similar reports published elsewhere [9,10].

Our sentiment analysis revealed increased negativity during the evaluation period, reflected by ~60% of posts related to distress including anxiety, worry, and risk. These results are similar to recent research evaluating user sentiment on Twitter during the COVID-19 pandemic [26]. Following a brief analysis of daily tweets related to four key emotions, the authors concluded fear was most strongly represented, followed by anger; joy; and, lastly, sadness. Increased negativity on social media is likely to be indicative, in part, of the broader population's overall sentiment. Indeed, this is supported by recent population-based research by the UK Office of National Statistics, reporting a doubling in high levels of anxiety between October 2019 to April 2020, compared with an earlier reference period [27]. Notably, women appear disproportionately affected, with 24% higher frequency of anxiety than men overall. Authors postulated that increased anxiety in women may be related to financial impacts, including a reduced likelihood of employment due to child rearing and gender inequities in salary. Increased anxiety and depressive symptoms in women are also common during pregnancy [15], and recent preliminary research reported a two-fold increase during COVID-19 [28]. The impact of social isolation was reported as a significant predictor of increased depression on a regression analysis [28]. Pregnancy, birth, and the postpartum periods are life stages requiring increased levels of social support, and reduced support at this time has been shown to have detrimental impacts on maternal mental health outcomes [29]. Women in this cohort frequently reported having reduced access to social support networks such as partners (ie, essential workers or those requiring quarantine due to business-related travel), family members, significant others, social networks, mothering groups, and health care professionals. In turn, this may have increased engagement with the forum as a source of like-minded support and comfort.

The women using the discussion forum demonstrated reasonable understanding and knowledge of relevant public health restrictions. This is encouraging with respect to behavior changes in response to the COVID-19 pandemic, as it indicates successful public health communication to the general community. However, women in the perinatal period are highly motivated to seek nuanced information that reduces risk to the mother, unborn child, or new baby. This includes information relating to self-isolation and safety while pregnant, employment implications if unable to work from home, the risks of COVID-19 infection at health facilities during perinatal medical visits, restrictions on birthing supports, and isolation from working partners. Women indicated information was sought from multiple sources but indicated, however, that it did not

meet their current needs and was either general in nature, ambiguous, or inconsistent. It is critical that the information needs of this cohort are understood and met to reduce the risk of mental health disorders in an already high-risk group and to reduce the risk of avoidance of health care monitoring. This is particularly imperative during pregnancy with some women indicating they would postpone or avoid antenatal visits, which may lead to negative health outcomes for both mother and baby. State and federal public health departments need to provide a central repository of information for this cohort that is accessible, timely, and reassuring. It is also important to understand that women are using discussion forums as their primary source of information in lieu of official sources.

Limitations

The following limitations should be considered when interpreting these findings. Anonymized data was interrogated and, therefore, demographical and geographical information about the user could not be obtained that may influence the user's engagement with the forum, as well as their perception and reaction to the COVID-19 pandemic. However, in accessing real-world nonidentified data, information is less likely to be influenced by social desirability or recall bias. We cannot confirm all posts were written by women in the perinatal period, however, forum users must enter an expected due date, child's birth date, or declare that they are trying to conceive to become a member of the forum.

Additionally, the real-world implications of users' online posts are unclear, and detrimental effects may be overemphasized in the absence of sufficient data pertaining to real-world behavior and state of mind. Due to early inconsistency in pandemic-related terminology, users may have used other keywords to describe COVID-19 that were not collected in this study. Finally, due to site management and restrictions relating to the seeking or provision of medical advice via the forum, posts may have been deleted before data was collected. In addition, the limitations of the software used in the sentiment analysis must be acknowledged, such as the inability to recognize sarcasm, double negatives, slang, dialect variations, idioms, and ambiguity. However, to address this limitation, we cross-checked the results and determined less than 10% were incorrectly classified. We believe this to be one of the first evaluations of the social impacts of COVID-19 in mothers or pregnant women in Australia [24,30-33]. This paper provides critical insights into the unanticipated impacts of this pandemic on a high need's perinatal cohort.

Conclusion

The perinatal period involves a major life transition requiring increased levels of social, emotional, and health professional support. Our results demonstrate pregnant women and new mothers are uniquely impacted by the COVID-19 pandemic. General information on COVID-19 safe behaviors does not appear to meet the needs of this population. The lack of nuanced and timely information appears to have exacerbated the risk of psychological and psychosocial distress in this vulnerable group who demonstrate heightened distress, reduced social and emotional support, anticipatory grief, increasing interfamily conflicts, and direct impacts on family planning behaviors. These findings suggest the need for targeted, consistent, accessible, and timely information on risk and risk aversion strategies, and adoption of strategies to de-escalate anxiety and concern. It also suggests that support strategies are needed to compensate for the loss of family, social support, and health professional contact, and lastly, that mental health interventions tailored to the unique needs of this cohort are likely important during the pandemic and for related public health policies.

Conflicts of Interest

None declared.

References

1. How to use Our World in Data. Our World in Data. URL: https://ourworldindata.org/how-to-use-our-world-in-data [accessed 2020-08-17]
2. Information for clinicians: frequently asked questions. Australian Government Department of Health. 2020 Mar 04. URL: https://www.health.gov.au/sites/default/files/documents/2020/03/coronavirus-covid-19-information-for-clinicians.docx [accessed 2020-06-12]
3. Clinical questions about COVID-19: questions and answers. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html [accessed 2020-06-12]
4. High consequence infectious diseases (HCID). United Kingdom Government. 2018 Oct 22. URL: https://www.gov.uk/guidance/high-consequence-infectious-diseases-hcid [accessed 2020-06-12]
5. Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [FREE Full text] [doi: 10.1186/s40779-020-00240-0] [Medline: 32169119]
6. First confirmed case of novel coronavirus in Australia. Australian Government Department of Health. 2020 Jan 25. URL: https://www.health.gov.au/ministers/the-hon-greg-hunt-mp/media/first-confirmed-case-of-novel-coronavirus-in-australia [accessed 2020-06-12]
7. Australia's social distancing rules have been enhanced to slow coronavirus — here's how they work. Australian Broadcasting Corporation. 2020 Mar 20. URL: https://www.abc.net.au/news/2020-03-20/coronavirus-covid-19-scott-morrison-enhanced-social-distancing/12075532 [accessed 2020-06-13]

8. Zaigham M, Andersson O. Maternal and perinatal outcomes with COVID-19: a systematic review of 108 pregnancies. Acta Obstet Gynecol Scand 2020 Jul;99(7):823-829 [FREE Full text] [doi: 10.1111/aogs.13867] [Medline: 32259279]
9. Zhu H, Wang L, Fang C, Peng S, Zhang L, Chang G, et al. Clinical analysis of 10 neonates born to mothers with 2019-nCoV pneumonia. Transl Pediatr 2020 Feb;9(1):51-60. [doi: 10.21037/tp.2020.02.06] [Medline: 32154135]
10. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet 2020 Mar;395(10226):809-815. [doi: 10.1016/s0140-6736(20)30360-3]
11. Robinson DP, Klein SL. Pregnancy and pregnancy-associated hormones alter immune responses and disease pathogenesis. Horm Behav 2012 Aug;62(3):263-271 [FREE Full text] [doi: 10.1016/j.yhbeh.2012.02.023] [Medline: 22406114]
12. Melville JL, Gavin A, Guo Y, Fan M, Katon WJ. Depressive disorders during pregnancy. Obstet Gynecol 2010;116(5):1064-1070. [doi: 10.1097/aog.0b013e3181f60b0a]
13. Rubertsson C, Hellström J, Cross M, Sydsjö G. Anxiety in early pregnancy: prevalence and contributing factors. Arch Womens Ment Health 2014 Jun;17(3):221-228. [doi: 10.1007/s00737-013-0409-0] [Medline: 24442712]
14. Seyfried LS, Marcus SM. Postpartum mood disorders. Int Rev Psychiatry 2003 Aug;15(3):231-242. [doi: 10.1080/0954026031000136857] [Medline: 15276962]
15. Maternal mental health. World Health Organization. URL: https://www.who.int/mental_health/maternal-child/maternal_mental_health/en/ [accessed 2020-07-01]
16. Lobel M, Cannella DL, Graham JE, DeVincent C, Schneider J, Meyer BA. Pregnancy-specific stress, prenatal health behaviors, and birth outcomes. Health Psychol 2008 Sep;27(5):604-615. [doi: 10.1037/a0013242] [Medline: 18823187]
17. Yali AM, Lobel M. Stress-resistance resources and coping in pregnancy. Anxiety Stress Coping 2002 Jan;15(3):289-309. [doi: 10.1080/1061580021000020743]
18. NVivo qualitative data analysis software. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home
19. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. How auto coding sentiment works. QSR International. URL: http://help-nv11.qsrinternational.com/desktop/concepts/How_auto_coding_sentiment_works.htm [accessed 2020-08-17]
21. Ellis L, Roberts L. Exploring the use and quality of Internet discussion forums in pregnancy: a qualitative analysis. Birth 2020 Mar;47(1):153-161. [doi: 10.1111/birt.12459] [Medline: 31583769]
22. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536x.2010.00390.x]
23. Iyer P. The paradox of privilege: guilt during COVID19. Observer Research Foundation. 2020 May 01. URL: https://www.orfonline.org/expert-speak/the-paradox-of-privilege-guilt-during-covid19-65509/ [accessed 2020-07-01]
24. Corbett GA, Milne SJ, Hehir MP, Lindow SW, O'connell MP. Health anxiety and behavioural changes of pregnant women during the COVID-19 pandemic. Eur J Obstet Gynecol Reprod Biol 2020 Jun;249:96-97 [FREE Full text] [doi: 10.1016/j.ejogrb.2020.04.022] [Medline: 32317197]
25. McLaren HJ, Wong KR, Nguyen KN, Mahamadachchi KND. Covid-19 and women's triple burden: vignettes from Sri Lanka, Malaysia, Vietnam and Australia. Soc Sci 2020 May 21;9(5):87. [doi: 10.3390/socsci9050087]
26. Lwin M, Lu J, Sheldenkar A, Schulz P, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
27. Personal and economic well-being in Great Britain: May 2020. Office for National Statistics. 2020 May 04. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/personalandeconomicwellbeingintheuk/may2020#disproportionately-impacted-groups [accessed 2020-06-23]
28. Durankuş F, Aksu E. Effects of the COVID-19 pandemic on anxiety and depressive symptoms in pregnant women: a preliminary study. J Matern Fetal Neonatal Med 2020 May 18:1-7. [doi: 10.1080/14767058.2020.1763946] [Medline: 32419558]
29. Balaji AB, Claussen AH, Smith DC, Visser SN, Morales MJ, Perou R. Social support networks and maternal mental health and well-being. J Womens Health 2007 Dec;16(10):1386-1396. [doi: 10.1089/jwh.2007.cdc10]
30. Wu Y, Zhang C, Liu H, Duan C, Li C, Fan J, et al. Perinatal depressive and anxiety symptoms of pregnant women during the coronavirus disease 2019 outbreak in China. Am J Obstet Gynecol 2020 Aug;223(2):240.e1-240.e9 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.009] [Medline: 32437665]
31. Chandra PS, Shiva L, Nagendrappa S, Ganjekar S, Thippeswamy H. COVID 19 related Psychosis as an interface of fears, socio-cultural issues and vulnerability- case report of two women from India. Psychiatry Res 2020 Aug;290:113136 [FREE Full text] [doi: 10.1016/j.psychres.2020.113136] [Medline: 32502827]
32. Ahorsu DK, Imani V, Lin C, Timpka T, Broström A, Updegraff JA, et al. Associations between fear of COVID-19, mental health, and preventive behaviours across pregnant women and husbands: an actor-partner interdependence modelling. Int J Ment Health Addiction 2020 Jun 11:A. [doi: 10.1007/s11469-020-00340-x]

33. Milne SJ, Corbett GA, Hehir MP, Lindow SW, Mohan S, Reagu S, et al. Effects of isolation on mood and relationships in pregnant women during the covid-19 pandemic. Eur J Obstet Gynecol Reprod Biol 2020 Sep;252:610-611 [FREE Full text] [doi: 10.1016/j.ejogrb.2020.06.009] [Medline: 32616415]

Abbreviations

MN: moderately negative
MP: moderately positive
VN: very negative
VP: very positive

Edited by G Eysenbach; submitted 01.07.20; peer-reviewed by L Guerra-Reyes, A McKinlay; comments to author 10.08.20; revised version received 17.08.20; accepted 17.08.20; published 07.09.20.

Please cite as:
Chivers BR, Garad RM, Boyle JA, Skouteris H, Teede HJ, Harrison CL
Perinatal Distress During COVID-19: Thematic Analysis of an Online Parenting Forum
J Med Internet Res 2020;22(9):e22002
URL: http://www.jmir.org/2020/9/e22002/
doi:10.2196/22002
PMID:32857707

Original Paper

Perceptions, Knowledge, and Behaviors Related to COVID-19 Among Social Media Users: Cross-Sectional Study

Khawla F Ali[1], MD; Simon Whitebridge[1], BSc; Mohammad H Jamal[2], MBChB, MEd, DABS, FRCSC, FACS, FASMBS; Mohammad Alsafy[3], BDS, MSc; Stephen L Atkin[1], MD, PhD

[1]Royal College of Surgeons in Ireland-Medical University of Bahrain, Adliya, Bahrain

[2]Department of Surgery, Kuwait University, Kuwait

[3]Clinica Dental Center, Kuwait

Corresponding Author:
Khawla F Ali, MD
Royal College of Surgeons in Ireland-Medical University of Bahrain
PO Box 15503
Adliya
Bahrain
Phone: 974 973 34611722
Email: khawlafouad@hotmail.com

Abstract

Background: Social media is one of the most rapid and impactful ways of obtaining and delivering information in the modern era.

Objective: The aim of this study was to rapidly obtain information on public perceptions, knowledge, and behaviors related to COVID-19 in order to identify deficiencies in key areas of public education.

Methods: Using a cross-sectional study design, a survey web link was posted on the social media and messaging platforms Instagram, Twitter, and WhatsApp by the study investigators. Participants, aged ≥18 years, filled out the survey on a voluntary basis. The main outcomes measured were knowledge of COVID-19 symptoms, protective measures against COVID-19, and source(s) of information about COVID-19. Subgroup analyses were conducted to determine the effects of age, gender, underlying illness, and working or studying in the health care industry on the perceived likelihood of acquiring COVID-19 and getting vaccinated.

Results: A total of 5677 subjects completed the survey over the course of 1 week. “Fever or chills” (n=4973, 87.6%) and “shortness of breath” (n=4695, 82.7%) were identified as the main symptoms of COVID-19. Washing and sanitizing hands (n=4990, 87.9%) and avoiding public places and crowds (n=4865, 85.7%) were identified as the protective measures most frequently used against COVID-19. Social media was the most utilized source for information on the disease (n=4740, 83.5%), followed by the World Health Organization (n=2844, 50.1%). Subgroup analysis revealed that younger subjects (<35 years), males, and those working or studying in health care reported a higher perceived likelihood of acquiring COVID-19, whereas older subjects, females, and those working or studying in non–health care areas reported a lower perceived likelihood of acquiring COVID-19. Similar trends were observed for vaccination against COVID-19, with older subjects, females, and those working or studying in non–health care sectors reporting a lower likelihood of vaccinating against COVID-19.

Conclusions: Our results are indicative of a relatively well-informed cohort implementing appropriate protective measures. However, key knowledge deficiencies exist with regards to vaccination against COVID-19, which future efforts should aim at correcting.

(J Med Internet Res 2020;22(9):e19913) doi:10.2196/19913

KEYWORDS

COVID-19; social media; public health; perception; knowledge; health information; health education; virus

Introduction

The novel coronavirus SARS-CoV-2 has been at the core of the devastating COVID-19 pandemic. The pandemic, which originated from Hubei Province, China, has spread around the globe, having claimed 512,842 lives as of July 2, 2020 [1,2].

COVID-19 is predominately characterized by systemic symptoms, such as fever and fatigue, and respiratory symptoms, such as cough, expectoration, and a runny nose [3,4]. In a recent analysis examining the relationship between symptoms of COVID-19 and disease severity, fatigue and expectoration were found to be the most critical, positive prognostic symptoms of COVID-19 severity, whereas a runny nose and nausea were favorable prognostic factors [4]. Transmission of SARS-CoV-2 occurs primarily via direct contact with the respiratory droplets of infected individuals, or via indirect contact with virally contaminated objects [5]. Increasing evidence suggests asymptomatic carriers have the ability to transmit the virus, making it ever so critical to identify and isolate these individuals effectively and rapidly [5-7]. Mortality in patients with COVID-19 differs across countries and has been significantly associated with age and comorbidities in subjects, ranging from 1.9% among ambulatory, low-risk patients to 21.7% among hospitalized, higher-risk individuals [8].

To date, despite intensive medical research on the RNA (ribonucleic acid) of the virus, no treatment has been found for COVID-19; only supportive measures are being used for those who need critical care. Public preventative measures remain key for slowing down its spread. One of the most effective methods for slowing or halting the spread of COVID-19 has been social distancing, and in some instances, social isolation [9,10]. Moreover, hygienic practices such as frequent hand washing, hand sanitizing, and wearing face masks, if implemented widely and correctly, can aid in reducing the spread of the virus [10]. Such simple yet effective measures mandate powerful messages, displayed on widely viewed communication platforms, for impactful dissemination. Additionally, with the rapidly evolving situation surrounding COVID-19, the speed of information dissemination is critical.

One of the fastest and most accessible platforms for broadcasting information is social media. Social media represents a conglomerate of electronic platforms utilized for creating and sharing information, ideas, messages, etc. Such platforms include social networking websites such as Twitter, Instagram, and Facebook, and messaging platforms like WhatsApp. These outlets have had a major global impact, with billions of users worldwide. Their use has extended from personal territories to organizational utilization to spread reliable information. Such organizations include the World Health Organization (WHO) and Centers of Disease Control and Prevention (CDC), both posting daily updates about the current pandemic, and each registering hundreds of thousands of followers worldwide.

The COVID-19 pandemic requires not only rapid information spread but also information identification and collation. This is essential for identifying gaps and misconceptions in the public's knowledge ("fake news") and behaviors toward the novel coronavirus. Social media offers an outlet to address both. Although the utilization of such platforms as tools to undertake research is in its infancy, their speed and extensive reach adds a unique digital print to the field of cross-sectional research [11,12]. Therefore, the aim of this cross-sectional study is to examine, through several social media platforms, the public's perceptions, knowledge, and behaviors related to the current COVID-19 pandemic to identify deficiencies in public education.

Methods

Study Population

Users of the social media and messaging platforms Instagram, Twitter, and WhatsApp, who were aged ≥18 years, were recruited to participate in the survey via a web link. The link was posted on the public social media pages of the following authors KFA, MHJ, and MA. Adult users who viewed the authors' public pages were asked to voluntarily fill out the survey. Resharing of the survey link by users was permitted on all three platforms for snowball sampling. The survey link was active from March 28 to April 4, 2020. Users with the highest visibility of the authors' accounts resided predominantly in the Arabian Gulf countries: Bahrain, Kuwait, Saudi Arabia, and United Arab Emirates. The survey was administered in both English and Arabic, the two predominantly spoken languages in these countries.

Survey Administration

Our survey was composed of 13 questions (see Multimedia Appendix 1 for questions). The first 7 questions addressed topics such as source(s) of information on COVID-19, viral preventative behavior(s), and knowledge of symptoms of COVID-19. We also asked whether the participant had acquired the infection; if they answered "no," they were asked to state their perceived likelihood of acquiring it in the next 3 months. Finally, we asked participants whether they would vaccinate against COVID-19. The other 6 questions pertained to subject demographics: age group, gender, educational level, country of current residence, underlying illness(es), and if the participant works or studies in the health care sector. The survey questions were partially adapted from a questionnaire designed and published by the Understanding America Study, which is maintained by the Center for Economic and Social Research at the University of Southern California [13]. The survey was designed on the online platform Zoho Creator (Zoho Corp). Informed consent in the form of agreement to an information leaflet was obtained from all subjects prior to survey initiation. The survey overall took less than 5 minutes to complete. The study was approved by the Research Ethics Committee at the Royal College of Surgeons in Ireland-Medical University of Bahrain.

Data Collection and Analysis

All data on personal demographics (age groups, gender, highest educational level, working or studying in the health care sector, presence of medical comorbidities, country of current residence, and status of COVID-19 infection) were categorically expressed as counts and percentages of total respondents. Data on source(s) of information during COVID-19, viral preventative behavior(s),

and knowledge of symptoms of COVID-19 were also expressed as counts and percentages of total respondents.

Subgroup categorical analyses were conducted to examine the effects of age (≥35 years versus <35 years), gender (female versus male), underlying illness(es) (yes versus no), and status of working or studying in the health care industry (yes versus no) on the perceived likelihood of being infected with COVID-19 in the next 3 months, as well as the likelihood of vaccinating against COVID-19. Results were categorically expressed as counts and percentages of total respondents in each of the subgroups.

Patient and Public Involvement

The development of the research questions and outcomes was primarily executed by the study investigators and guided by current literature. Patients were not involved in the design of this study. The involvement of patients and the public, however, was crucial during the recruitment phase of the study (ie, dissemination of the survey link, via snowballing effect, through users' social media accounts to others in the digital community). After publication, the results will be disseminated to participants on the same platforms that hosted the survey link, through a short, interactive video recorded and posted by the principle investigator of the study.

Results

A total of 5677 subjects completed the survey. Of the survey respondents, 3945 (69.5%) were female, 3737 (65.9%) were <35 years old, and 4257 participants (75.0%) reported a higher educational level. Amongst respondents, 1250 (22%) reported working or studying in the health care industry and 4003 (70.5%) reported no underlying illness. The majority of respondents resided in Bahrain (3179/5677, 56.0%), Kuwait (784/5677, 13.8%), and Saudi Arabia (693/5677, 12.2%). Only 55 respondents (1%) reported that they had been diagnosed with COVID-19. Other baseline demographics are detailed in Table 1.

Table 1. Baseline characteristics of the study population (N=5677).

Characteristic	Participants
Female, n (%)	3945 (69.5)
Age, n (%)	
18-24 years	1712 (30.2)
25-34 years	2025 (35.7)
35-44 years	1104 (19.4)
45-54 years	534 (9.4)
55-64 years	256 (4.5)
≥65 years	46 (0.8)
Educational status, n (%)	
Primary school	11 (0.2)
Intermediate school	68 (1.2)
High school	1341 (23.6)
College/higher education	4257 (75.0)
Work or study in the health care sector, n (%)	
Yes	1250 (22.0)
No	4427 (78.0)
Medical problems, n (%)	
I have no medical problems	4003 (70.5)
I have medical problems	1674 (29.5)
Type of medical problems[a], n (%)	
High blood pressure	387 (23.1)
Diabetes	362 (21.6)
Heart disease	74 (4.4)
Lung disease	198 (11.8)
Cancer	31 (1.9)
Other	946 (56.5)
Country/region of current residence, n (%)	
Bahrain	3179 (56.0)
Kuwait	784 (13.8)
Saudi Arabia	693 (12.2)
United Arab Emirates	324 (5.7)
Oman	232 (4.1)
Qatar	71 (1.3)
Other Arab countries	166 (2.9)
Asian countries (excluding Arab countries)	43 (0.8)
Europe	118 (2.1)
North America	58 (1.0)
South America	2 (0)
Australia and New Zealand	7 (0.1)
COVID-19 diagnosis, n (%)	

Characteristic	Participants
Yes	55 (1.0)
No	5622 (99.0)

[a]n=1674.

The majority of respondents identified “fever or chills” (4973/5677, 87.6%), “shortness of breath” (4695/5677, 82.7%) and “cough” (4150/5677, 73.1%) as the main symptoms of COVID-19 (Figure 1). Only 165 respondents (2.9%) reported “I do not know” when asked about the main symptoms of COVID-19 (Figure 1). The most reported preventative behaviors were frequent washing and sanitizing hands (4990/5677, 87.9%), avoiding public places and crowds (4865/5677, 85.7%), and canceling or postponing social activities (4371/5677, 77.0%) (Figure 2). Only 74 subjects (1.3%) reported not changing their behavior in response to COVID-19 (Figure 2). The most utilized sources for COVID-19 information were social media platforms (4740/5677, 83.5%), the WHO (2844/5677, 50.1%), and TV (2413/5677, 42.5%) (Figure 3).

Figure 1. Knowledge about main symptoms of COVID-19 infection. Results are expressed as % of respondents. n=5,677.

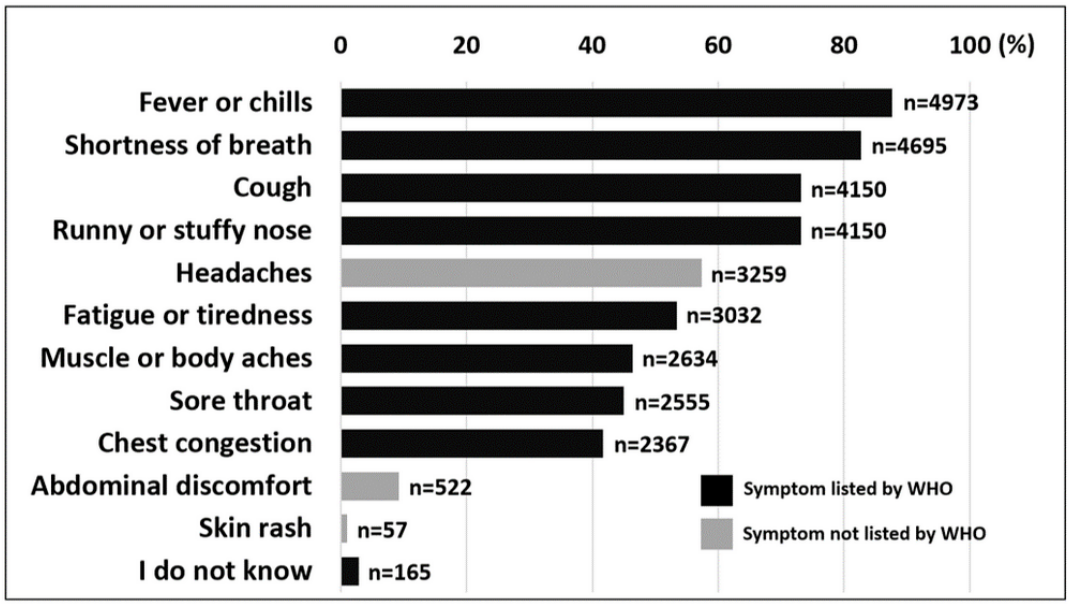


Figure 2. Behaviour(s) done in past week to prevent COVID-19 infection. Results are expressed as % of respondents. n=5,677.

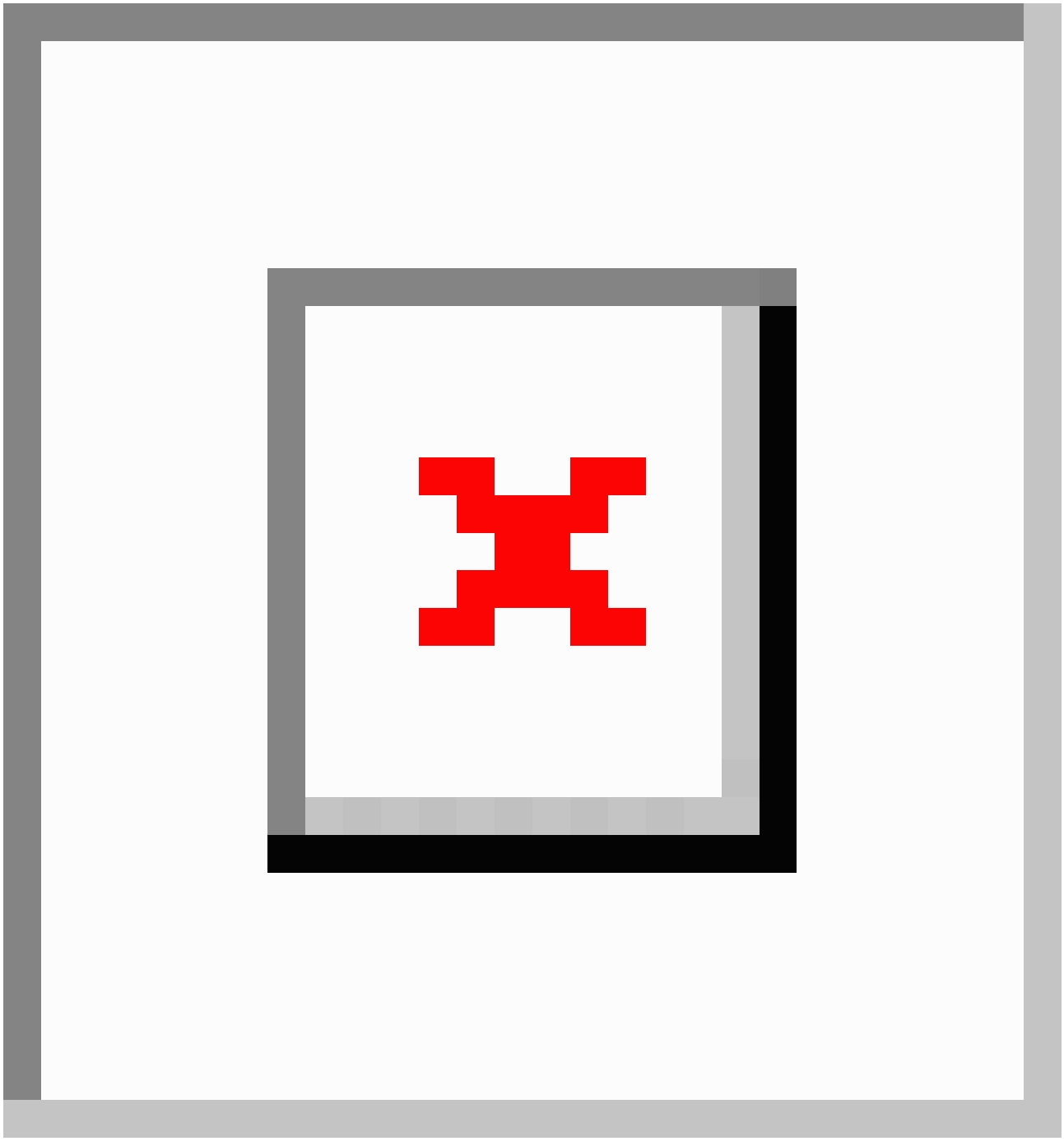
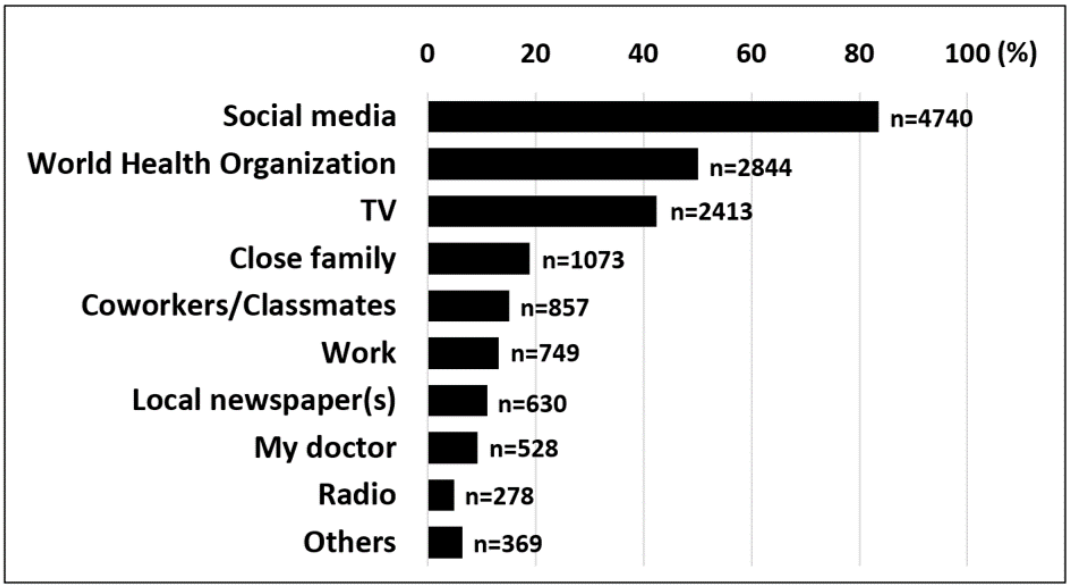


Figure 3. Source(s) of COVID-19 information. Results are expressed as % of respondents (N=5677).



In a subgroup analysis examining the effects of age, gender, underlying illnesses, and work or study in the health care sector on the perceived likelihood of acquiring COVID-19 in the next 3 months, younger subjects (<35 years), male subjects, and health care workers and students reported a higher likelihood of acquiring COVID-19, expressed as "very likely" and "somewhat likely" (Figure 4). In a secondary subgroup analysis investigating the effects of the latter on the likelihood of vaccinating against COVID-19, younger subjects (<35 years), male subjects, and those working or studying in health care also reported a higher likelihood of vaccinating, expressed as "very likely" and "somewhat likely" (Figure 5).

Figure 4. Perceived likelihood of acquiring COVID-19 infection in the next 3 months. Results are categorized by age group, gender, presence of underlying illness(es), and if the subject works or studies in the health care industry. Results are expressed as % of respondents. *n=5602 in all categories (55 subjects reporting COVID-19 infection were eliminated from this analysis; 20 subjects with no response were also eliminated).

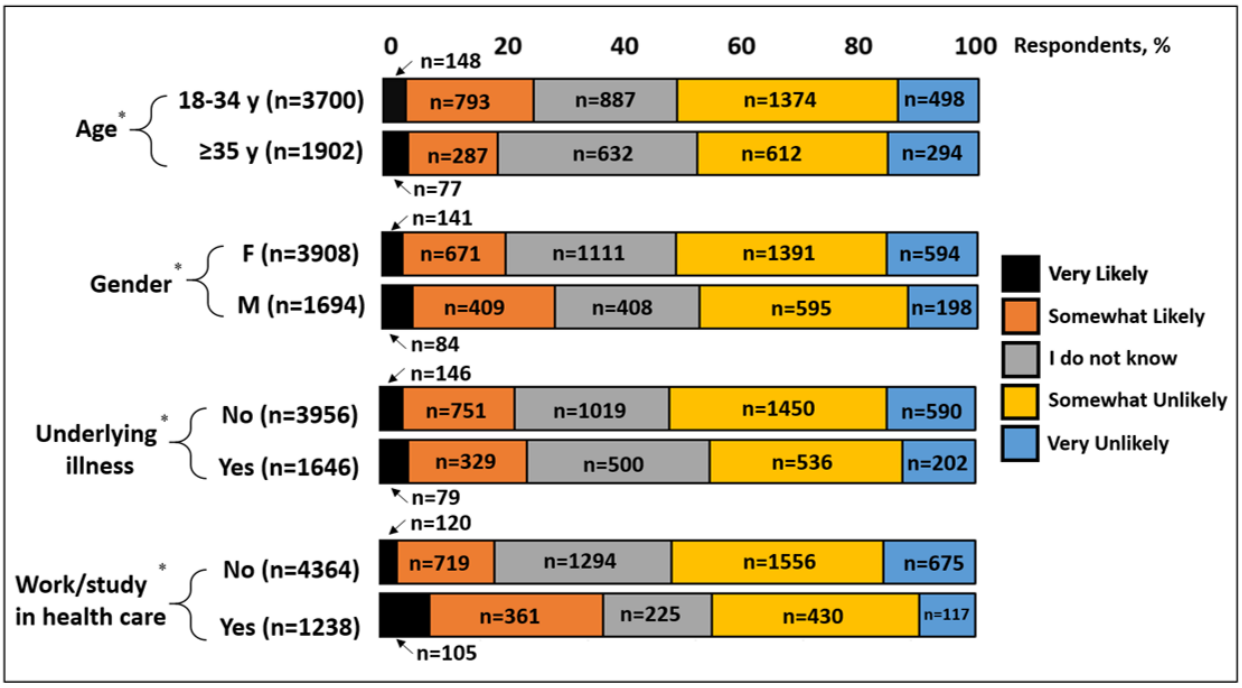
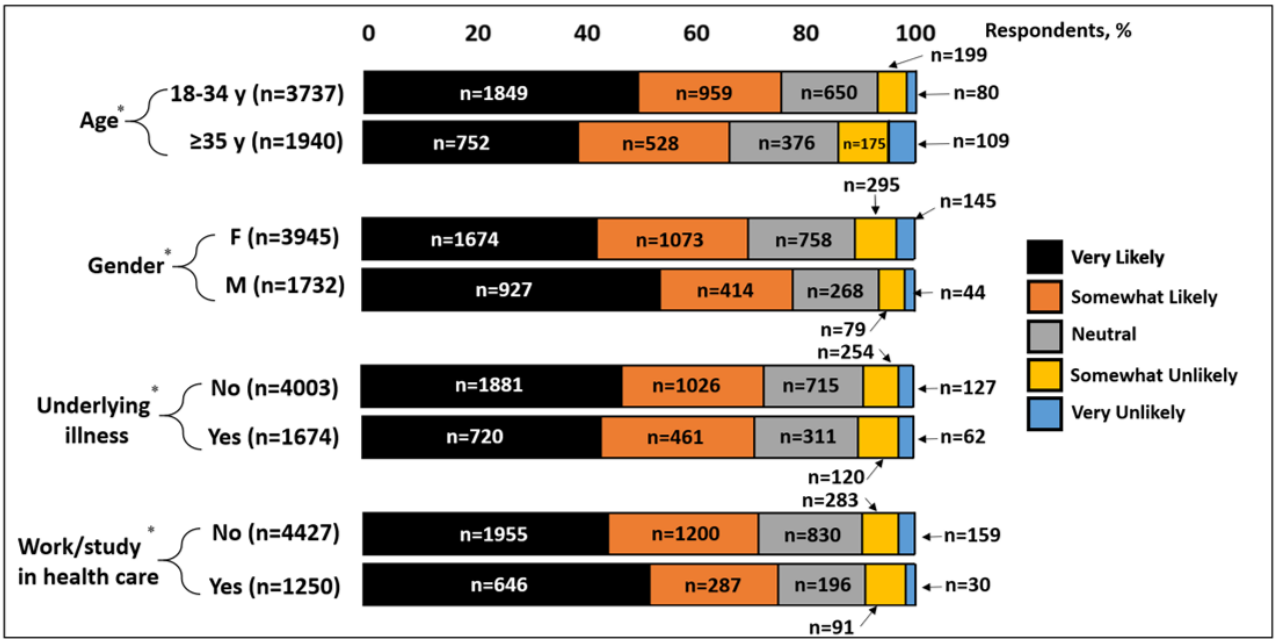


Figure 5. Likelihood of vaccinating against COVID-19. Results are categorized by age group, gender, presence of underlying illness and if subject works/studies in healthcare industry. Results are expressed as % of respondents. *n=5,677 in each category.



Discussion

Principal Findings

In the midst of an unprecedented global health crisis, it is critical to know if the educational messages for the prevention of COVID-19 are being delivered, understood, and implemented by the public. Our survey, conducted through three social media outlets, facilitated a large and rapid collection of 5677 responses within 7 days and showed that the majority of respondents were well aware of the symptoms of COVID-19 and the measures necessary to prevent it, as per the WHO guidelines [14]. Additionally, a large portion of subjects (around 50%) utilized reliable sources of information such as the WHO. Such collective responses are indicative of a well-informed cohort, which may be explained by the younger, more educated (75% had higher education) users of social media. It is of note that a large number of these responses (56.0%) were from people in Bahrain. The spread of COVID-19 in Bahrain was well contained at the time of survey distribution (998 cases as of April 11, 2020) [2]. The number of fatalities (7 as of April 11, 2020) [2] has been very low considering it is the third most densely populated country in the world. Such statistics can be due to the early and extensive communication with the public, particularly through social media outlets [15,16]; this highlights the importance of conveying general knowledge about the virus to residents to control spread.

Our subgroup analysis examining the effects of age (<35 versus ≥35 years old), gender, presence of underlying illnesses, and work or study in health care revealed some intriguing findings. Younger subjects (<35 years old) reported a higher likelihood of vaccinating against COVID-19 compared to older participants. This may be a reflection of their higher perceived likelihood of acquiring COVID-19, as demonstrated in our subgroup analysis. A higher perceived likelihood of infection may be explained by higher chances of coming into contact with perceived high-risk groups as mandated by work, social, or study environments. Similarly, male subjects reported a higher likelihood of acquiring the infection and vaccination than their female counterparts. This may have been triggered by late reports from China, Italy, France, Germany, and South Korea indicating higher rates of mortality, as high as 89%, in males compared to females [17]. Infection rates, however, have not widely differed between the sexes [17]. It has been hypothesized that the differences in mortality rate may be due to higher rates of cigarette smoking, alcohol consumption, and the number of pre-existing comorbidities among men compared to women [18]. Lastly, workers or students in the health care sector reported a higher likelihood of acquiring the infection, as well as a higher likelihood of vaccinating, as one might expect. No major differences were noted amongst those with and without underlying illnesses.

Limitations

One major limitation of this study is its sampling technique. Convenience sampling, typically utilized in cross-sectional studies, is a type of nonprobability sampling that allows for data collection from a group of people who are easy to contact or reach [19]. This may have introduced a sampling bias in our subject cohort. For instance, female and younger subjects are more likely to be represented in social media compared to their male and older counterparts [11], as seen in our study population. Another major limitation is the susceptibility of the study to a nonresponse bias. Our sample also had an unusually high prevalence (22%) of subjects either working or studying in the health care sector. This may be explained by the fact that the survey link was posted on platforms managed by physicians. Such platforms would typically attract users from the same profession. Additionally, the majority of respondents resided in a confined geographical area located within the Arabian Gulf Peninsula. Thus, the data may not be applicable to subjects residing elsewhere. We plan to address such limitations with

future studies targeting a wider and more diverse sample of the population.

Conclusions

Our results are indicative of a well-informed cohort, implementing appropriate protective measures, with the majority reporting social media as their main source of COVID-19 information. This demonstrates that social media is an impactful way of collecting data and delivering information, even though there may be inherent limitations to this study design. We plan to use this platform again to determine if such changes are maintained with the continuation of the pandemic.

Acknowledgments

KFA designed the study, analyzed data, and wrote the manuscript; SW designed the survey, analyzed the data, and edited the manuscript; MHJ and MA contributed to survey distribution and editing of the manuscript; and SLA assisted in study design, analysis of data, and writing of the manuscript. All authors approved this version of the manuscript. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[PDF File (Adobe PDF File), 349 KB - jmir_v22i9e19913_app1.pdf]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus InvestigatingResearch Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. Coronavirus disease (COVID-2019) situation reports. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-04-11]
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
4. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. J Med Internet Res 2020 Jun 30;22(6):e19636 [FREE Full text] [doi: 10.2196/19636] [Medline: 32544071]
5. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020 May;109:102433 [FREE Full text] [doi: 10.1016/j.jaut.2020.102433] [Medline: 32113704]
6. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar 28;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
7. Liu Y, Wang Z, Ren J, Tian Y, Zhou M, Zhou T, et al. A COVID-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study. J Med Internet Res 2020 Jun 29;22(6):e19786 [FREE Full text] [doi: 10.2196/19786] [Medline: 32540845]
8. Mikami T, Miyashita H, Yamada T, Harrington M, Steinberg D, Dunn A, et al. Risk Factors for Mortality in Patients with COVID-19 in New York City. J Gen Intern Med 2020 Jun 30:e [FREE Full text] [doi: 10.1007/s11606-020-05983-z] [Medline: 32607928]
9. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
10. Coronavirus disease (COVID-2019) advice for public. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-07-02]
11. Nelson LM, Simard JF, Oluyomi A, Nava V, Rosas LG, Bondy M, et al. US Public Concerns About the COVID-19 Pandemic From Results of a Survey Given via Social Media. JAMA Intern Med 2020 Apr 07:e [FREE Full text] [doi: 10.1001/jamainternmed.2020.1369] [Medline: 32259192]
12. Surani Z, Hirani R, Elias A, Quisenberry L, Varon J, Surani S, et al. Social media usage among health care providers. BMC Res Notes 2017 Nov 29;10(1):654 [FREE Full text] [doi: 10.1186/s13104-017-2993-y] [Medline: 29187244]
13. The Understanding America Study Coronavirus in America "COVID" Survey. University of South California. 2020. URL: https://uasdata.usc.edu/index.php [accessed 2020-03-22]
14. Q&A on coronaviruses (COVID-19). World Health Organization. 2020. URL: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses [accessed 2020-04-12]

15. MOH_Bahrain. Twitter. URL: https://twitter.com/MOH_Bahrain [accessed 2020-04-12]
16. mohbahrain. Instagram. URL: https://www.instagram.com/mohbahrain [accessed 2020-04-12]
17. BMJ GH Blogs. Sex, gender and COVID-19: Disaggregated data and health disparities. BMJ. 2020 Mar. URL: https://blogs.bmj.com/bmjgh/2020/03/24/sex-gender-and-covid-19-disaggregated-data-and-health-disparities [accessed 2020-04-12]
18. James S, Abate D, Abate K, Abay S, Abbafati C, Abbasi N, et al. . [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
19. Tyrer S, Heyman B. Sampling in epidemiological research: issues, hazards and pitfalls. BJPsych Bull 2016 Apr;40(2):57-60 [FREE Full text] [doi: 10.1192/pb.bp.114.050203] [Medline: 27087985]

Abbreviations

CDC: Centers of Disease Control and Prevention
RNA: ribonucleic acid
WHO: World Health Organization

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by I Basheti, A Benetoli; comments to author 23.06.20; revised version received 04.07.20; accepted 26.07.20; published 08.09.20.

Please cite as:
Ali KF, Whitebridge S, Jamal MH, Alsafy M, Atkin SL
Perceptions, Knowledge, and Behaviors Related to COVID-19 Among Social Media Users: Cross-Sectional Study
J Med Internet Res 2020;22(9):e19913
URL: http://www.jmir.org/2020/9/e19913/
doi:10.2196/19913
PMID:32841153

Original Paper

Real-World Implications of a Rapidly Responsive COVID-19 Spread Model with Time-Dependent Parameters via Deep Learning: Model Development and Validation

Se Young Jung[1,2*], MD, MPH; Hyeontae Jo[3*], PhD; Hwijae Son[4], MSc; Hyung Ju Hwang[4], PhD

[1]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[2]Department of Family Medicine, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[3]Basic Science Research Institute, Pohang University of Science and Technology, Pohang, Republic of Korea

[4]Department of Mathematics, Pohang University of Science and Technology, Pohang, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyung Ju Hwang, PhD
Department of Mathematics, Pohang University of Science and Technology
77, Cheongam-ro, Nam-gu, Pohang-si
Gyeongsangbuk-do
Pohang, 37673
Republic of Korea
Phone: 82 542792056
Email: hjhwang@postech.ac.kr

Abstract

Background: The COVID-19 pandemic has caused major disruptions worldwide since March 2020. The experience of the 1918 influenza pandemic demonstrated that decreases in the infection rates of COVID-19 do not guarantee continuity of the trend.

Objective: The aim of this study was to develop a precise spread model of COVID-19 with time-dependent parameters via deep learning to respond promptly to the dynamic situation of the outbreak and proactively minimize damage.

Methods: In this study, we investigated a mathematical model with time-dependent parameters via deep learning based on forward-inverse problems. We used data from the Korea Centers for Disease Control and Prevention (KCDC) and the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University for Korea and the other countries, respectively. Because the data consist of confirmed, recovered, and deceased cases, we selected the susceptible-infected-recovered (SIR) model and found approximated solutions as well as model parameters. Specifically, we applied fully connected neural networks to the solutions and parameters and designed suitable loss functions.

Results: We developed an entirely new SIR model with time-dependent parameters via deep learning methods. Furthermore, we validated the model with the conventional Runge-Kutta fourth order model to confirm its convergent nature. In addition, we evaluated our model based on the real-world situation reported from the KCDC, the Korean government, and news media. We also crossvalidated our model using data from the CSSE for Italy, Sweden, and the United States.

Conclusions: The methodology and new model of this study could be employed for short-term prediction of COVID-19, which could help the government prepare for a new outbreak. In addition, from the perspective of measuring medical resources, our model has powerful strength because it assumes all the parameters as time-dependent, which reflects the exact status of viral spread.

(J Med Internet Res 2020;22(9):e19907) doi:10.2196/19907

KEYWORDS

epidemic models; SIR models; time-dependent parameters; neural networks; deep learning; COVID-19; modeling; spread; outbreak

Introduction

Similar to the 1918 influenza pandemic that occurred more than 100 years ago, the COVID-19 pandemic has created major disruptions worldwide. At the end of World War I, the 1918 influenza pandemic wreaked havoc globally; it killed more than 40 million people, more than 2% of the world's population [1]. During the outbreak, preventive measures such as social distancing and wearing masks were recommended to curb the spread of the virus [2]. Unfortunately, these measures were insufficient. The 1918 influenza pandemic exhibited an unusual bimodal or trimodal peak in the United States, lasting for almost two years [3]. Thus, by analogy, we can infer that decreases in infection rates of COVID-19 do not guarantee continuity of the trend. Therefore, it is necessary to develop a precise spread model of COVID-19 that responds promptly to the dynamic situation of the outbreak. If the model accurately measures the effectiveness of COVID-19–related preventative measures and provides reasonable information about the spreading trend in the next few days, it will be possible to proactively minimize damage by taking effective actions against recurring outbreak situations. Furthermore, we assessed the potential roles of a number of public health measures acting in advance based on the developed model to reduce contact rates and thereby reduce transmission of the virus in the absence of a COVID-19 vaccine [4].

Recently, there have been numerous studies on developing models to find a mathematical description of a system and translate it to the current situation of COVID-19. These studies typically introduce the susceptible-infected-recovered (SIR) model or its derivatives. In some of these studies, the model parameters are considered as constants due to the complexity of modeling. For instance, a previous study proposed a conceptual model that includes individual behavioral reactions and government actions, while another study reviewed the basic reproduction number of COVID-19 with constant parameters [5,6]. However, the reproduction number (R) innately assumes time-dependent variables. R is a function of three primary parameters; two of these are biological constants (the infectiousness of the pathogen and the duration of contagiousness after a person becomes infected), and the other is a sociobehavioral and environmental variable (the contact rate) [7]. The contact rate causes the reproduction number to fluctuate through human-to-vector or human-to-human interactions varying over time or space. Thus, it is more reasonable to define mathematical parameters in a model as time-dependent variables. However, previous representative studies did not use this method. A previous study divided the phase manually and considered the parameters as time-varying piecewise constants [6]. Other studies considered the parameters as partial functions of time and proposed methods to approximate the time-varying parameters [8,9]. More recently, a method to quantify the effects of quarantine control using a neural network was proposed. Although the authors considered the strength of quarantine control as a time-dependent parameter, the other parameters were still considered as constants [10]. Overall, most previous studies partially adopted time-variant parameters due to technical difficulties. In general, parameters of the deterministic SIR model with constant parameters can be estimated after solving the solutions of the model. However, this approach has a limitation when the model has time-dependent parameters. In previous studies related to COVID-19 [11,12], it was already recognized that parameters will change at a specific moment, such as the early phase of the epidemic, enforcement of the quarantine policy, or supply of medical equipment. As a result, piecewise constant parameters emerged depending on the artificially divided time intervals. In contrast, we suggested a new method to calculate the time-varying parameters without any artificial setting. This method enables us to analyze the times when unusual events occur and to evaluate the quarantine policy. This is the starting point of this research. We aimed to develop a model that was more precise and sensitive than previous models by introducing as many time-dependent parameters as possible to reflect that the current situation is changing on a daily basis. We adopted the SIR model with the concept of the forward-inverse problem. Furthermore, we approximated outcome variables and parameters in the model with neural networks to compute the infection rate, recovery rate, and reproduction numbers more accurately.

Methods

Methodological Overview

Mathematical modeling is a process that aims to find a mathematical description of a system and translate it into a relational expression. When a system (eg, an infectious disease) continuously changes over time, differential equations, which may include parameters, can be used to model it. The process of finding the parameters that best fit the given data from the system is called an inverse problem. In this study, we aimed to analyze COVID-19 spread in South Korea using the SIR model. We approximated each outcome variable (S, I, and R) and parameter (β and γ) in the model using deep learning. Moreover, to address the shortcomings of previous studies, we considered the parameters as functions of time, which allowed us to compute the infection rate, the recovery rate, and the time-dependent reproduction number, R_{TD}. This approach is more interpretable because $\beta(t)$, $\gamma(t)$, and R_{TD} can be obtained as functions of time, and the overall dynamics of the actual data can also be obtained. We hypothesized that R_{TD} could be used as a surrogate marker to indicate the pressure on health care resources in a region. This is because the number of available beds for patients with COVID-19 in an area decreases when the infection rate increases or when the recovery rate stagnates or decreases.

Additionally, unlike in other models, such as the growth model, we do not assume any distribution type for the modeling. In the traditional growth model, the growth rate is considered as a piecewise constant function to compute the effective reproduction number. However, this assumption is not realistic in many cases, as the reproduction number can dramatically change. In contrast, our model is an appropriate solution for such problems due to its time-dependent nature. Furthermore, we provide numerical simulation results that guarantee the convergence of our deep learning approach. Finally, our

methodology is applicable to many areas involving differential equations, and it can be easily implemented without a deep understanding of the model.

Terminology

The reproduction number has several variants. The basic reproduction number (R_0) is defined as the expected number of cases directly generated by one case in a population, assuming all individuals are susceptible to infection. Compared to R_0, the effective reproduction number (R_t) does not assume complete susceptibility of the population [7]. Strictly speaking, all reproduction numbers after the first date of introduction of new pathogens should be regarded as R_t. In this study, we wanted to develop a time-dependent effective reproduction number that is a variant of R_t; we designated this number as R_{TD}.

Data

We collected our data from the Korea Centers for Disease Control and Prevention (KCDC) and the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The data consisted of the cumulative numbers of tested people (T), confirmed cases (I or I_{pos}), negative cases (I_{neg}), and recovered or deceased cases (R) from February 7 to March 30, 2020, for South Korea and from March 5 to March 30, 2020, for the administrative provinces of Seoul, Busan, Daegu, and Gyeonggi. The data are available at the KCDC website [13]. Although data for South Korea are available from January 29, the numbers of negative, recovered, and deceased cases are not available for the first few weeks; therefore, we began our data range on February 7, 2020. The complete data, including numbers of negative, recovered, and deceased cases, for each administrative province are available from March 5; therefore, we used all data up to March 30, 2020. We set t=0 as March 5, 2020, when data for each province became available, and February 7, 2020 corresponds to t=–26.6 (Figures 1 and 2).

Figure 1. Cumulative numbers of infected and recovered COVID-19 cases in South Korea (left) and cumulative numbers of negative cases and tested people (right).

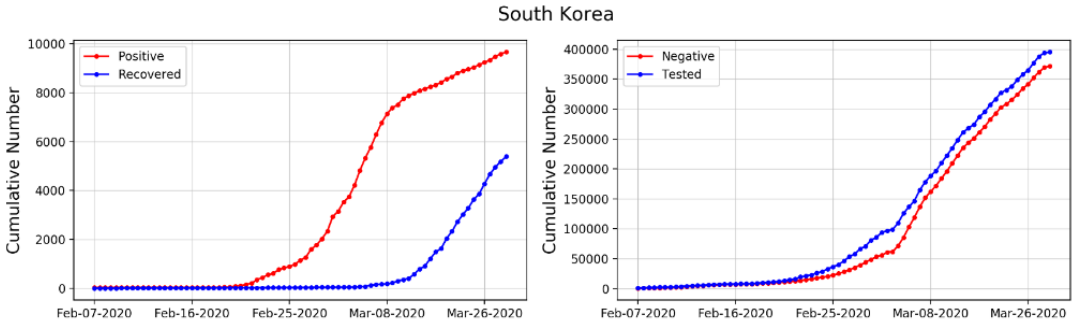
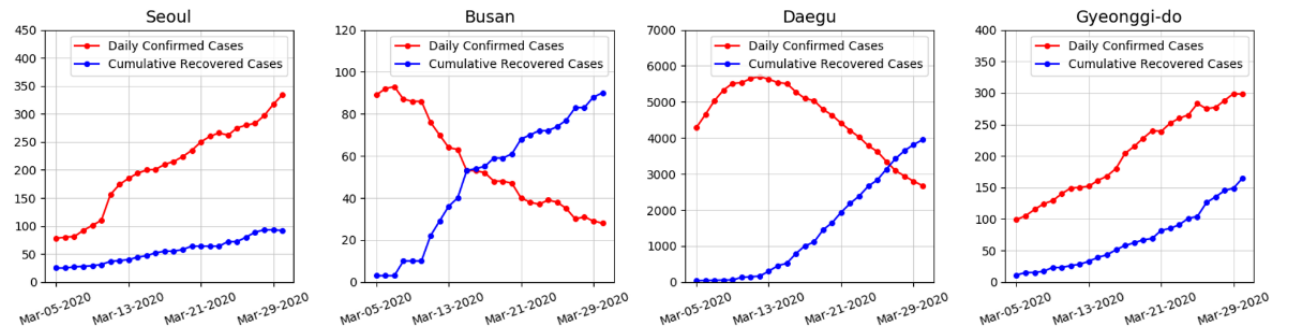


Figure 2. Daily numbers of confirmed cases and cumulative numbers of recovered cases for Seoul, Busan, Daegu, and Gyeonggi-do.

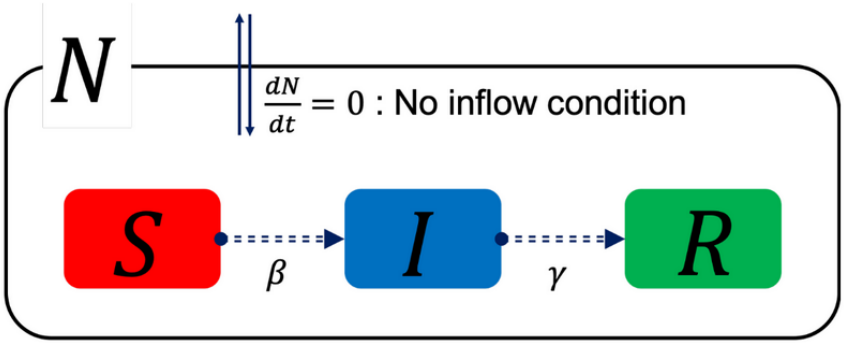


This study received an exemption from informed consent by the institutional review board committee of the Seoul National University Bundang Hospital because we used public data provided by the KCDC.

SIR Model

Infectious disease modeling in mathematics can capture an epidemic of a given infectious disease and aid public health interventions. Modeling usually requires disease-related statistical data, calculation of model parameters, and analysis of the epidemic. We adopted the SIR model, which is suitable for our data (see Figure 3). For a fixed time t≥0, let S(t), I(t), R(t), and N(t) denote the numbers of susceptible, infected, and recovered (or removed) cases and the sum of these three populations, respectively. Moreover, we applied a scaled SIR model (divided by N for each outcome variable S, I, and R) and time-varying parameters (β and γ) to the final SIR model. We also assumed that the total number of the population is time-invariant, that is, $S(t) + I(t) + R(t) = 1$. The mathematical formula of the SIR model is provided in detail in Multimedia Appendix 1.

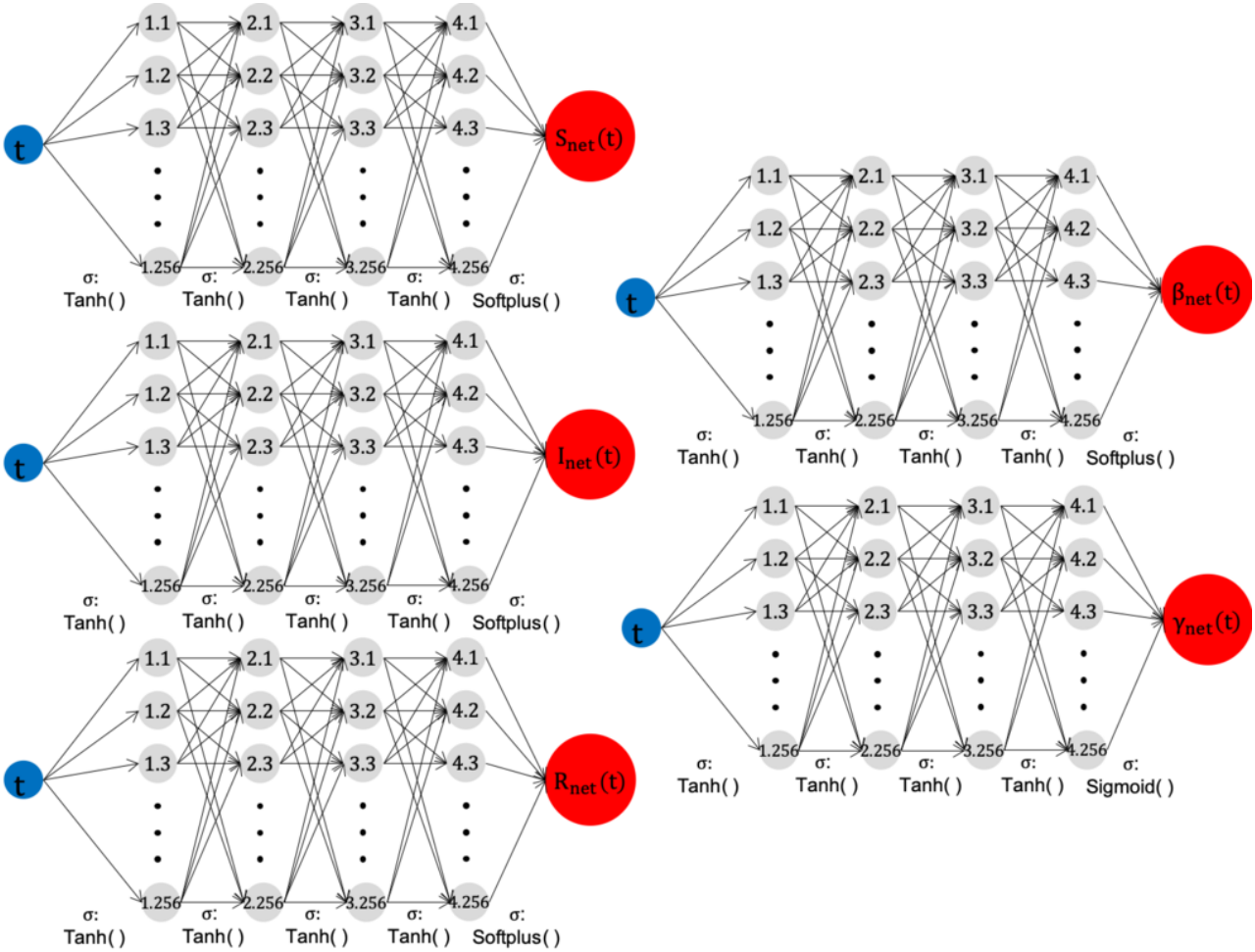
Figure 3. Illustration of the SIR model. I: infected; R: recovered; S: susceptible.



Deep Learning

We constructed five neural network models for S, I, R, β, and γ, denoted by S_{net}, I_{net}, R_{net}, β_{net}, and γ_{net}, respectively. The concrete model structures are presented in Figure 4. We applied similar training methods to solve forward and inverse problems, as introduced in previous studies [14,15]. The detailed deep learning methodology is provided in Multimedia Appendix 1.

Figure 4. Forward-Inverse SIR model networks. Each network contains 1 input node (time t), one1output node (value), 4 hidden layers, and 256 nodes in each hidden layer. The hyperbolic tangent tanh(x) is used in the activation functions except for the last layer. The Softplus and Sigmoid functions are used in the last hidden layer to meet the constraints S, I, R, β>0, and 0<γ<1. I: infected; R: recovered; S: susceptible.



We conducted simulations for four provinces: Seoul, Gyeonggi-do, Busan, and Daegu. We applied five deep neural network (DNN) models to derive the parameters S_{net}, I_{net}, R_{net}, β_{net}, and γ_{net}. For a more accurate evaluation of the model parameters, we also provided a numerical solution called Runge-Kutta fourth order (RK4) using the estimated parameters. RK4 is one of the most well-known and theoretically proven algorithms that converges to analytic solutions. In contrast, the neural network-based methodology of this study has a weak theoretical background for convergence. Therefore, we aimed to show how close the time-dependent parameters found by DNN are to the actual solution through RK4.

For the RK4 method, we set a step size of $h=10^{-3}$, with 26 observations used for Seoul, Busan, Daegu, and Gyeonggi and

77 observations used for South Korea. The observations are presented in Multimedia Appendix 1.

Results

Estimating the Parameters of the SIR model with DNN

We estimated the model parameters (β and γ) and outcome variables (S, I, and R) in the SIR model via DNNs for South Korea, Seoul, Busan, Daegu, and Gyeonggi. The results for South Korea are presented in Figure 5. The results for Seoul, Busan, Daegu, and Gyeonggi are provided in Multimedia Appendix 1 (Figures SM1 to SM4, respectively).

We also estimated R_{TD} for South Korea (Figure 6).

Figure 5. SIR model target values and relative errors from February 7 (t=–26.6) to March 30 (t=25.0), 2020, in South Korea. The red lines in the top three graphs denote the S^{net}, I^{net}, and R^{net} values for each graph, the green lines in the top and middle three graphs denote the observations, and the blue lines in the middle three graphs denote the RK4 results with the parameters β^{net} and γ^{net}. The population ratio is the number of people in each group (S, I, and R) divided by the total number of people (N). Relative errors were defined as (|observed value − Network [or RK4] value|/|observed value|) × 100 and were calculated for each parameter. I: infected; R: recovered; RK4: Runge-Kutta fourth order method; S: susceptible.

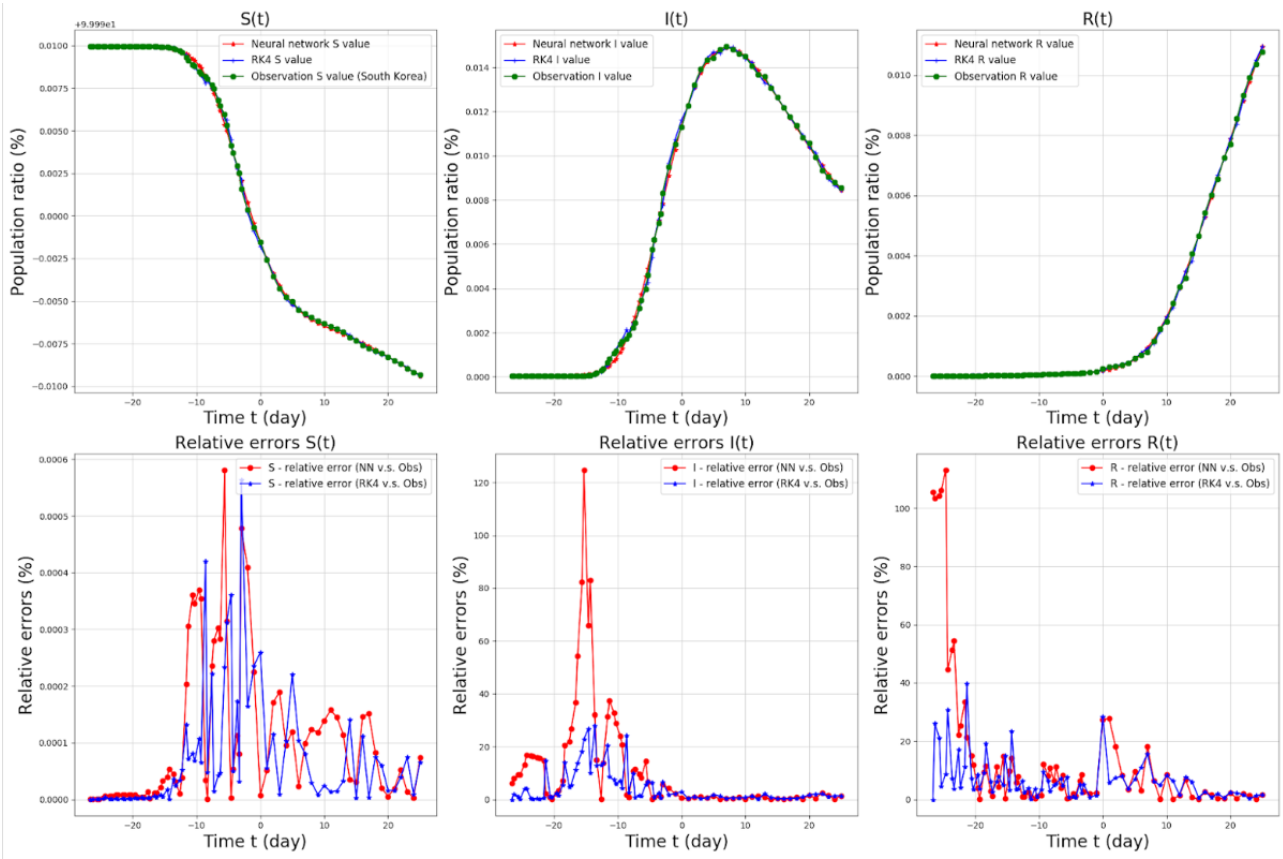
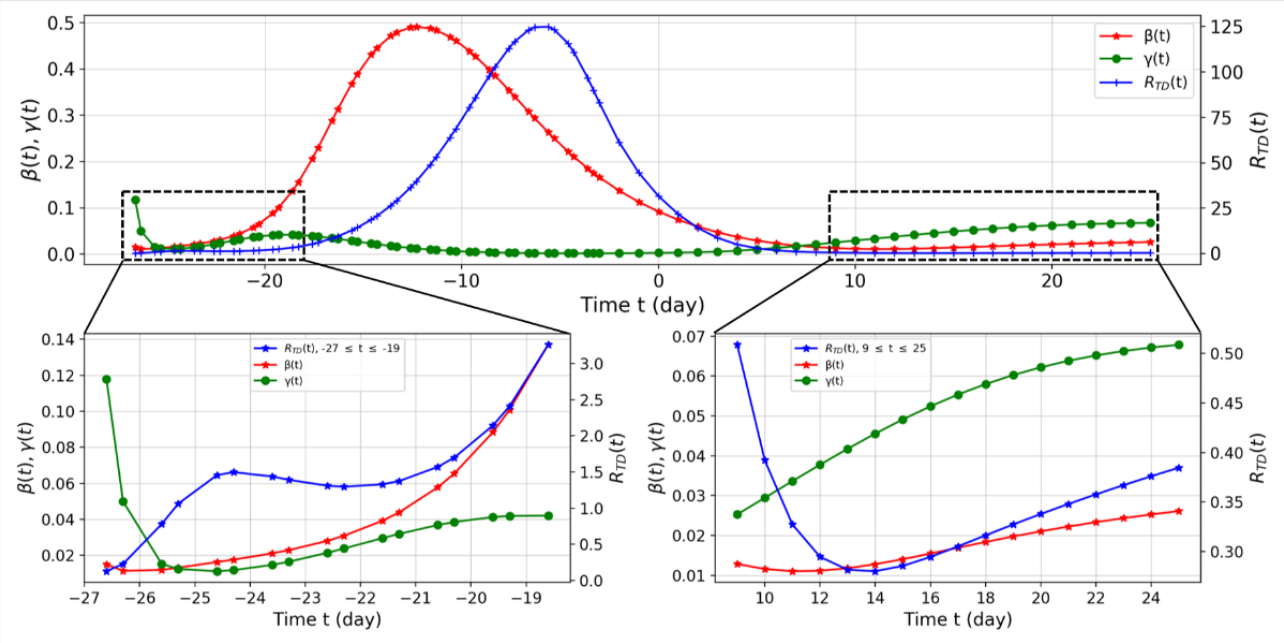


Figure 6. Suspected-infected-recovered model parameter network values and R_{TD} values from February 7 (t=–26.6) to March 30 (t=25.0), 2020, for South Korea. We divided the range of R_{TD} into two parts, shown at bottom left (–26.6≤t≤–19) and bottom right (9≤t≤25.0). On February 18 (t=–15.3), the first case was confirmed to be related to Shincheonji, which was the starting point of the outbreak in Daegu.



We summarized the overall trend by analyzing R_{TD} (t). First, on February 8 (t=–25.3) in South Korea, R_{TD}=1.0610 implies the spread of COVID-19. Starting from February 18, R_{TD} (t) increased dramatically (t=–15.6), and it reached its peak (R_{TD} (t)=124.8454) on February 28 (t=–5.6). After March 13 (t=8.0), R_{TD} decreased below 1 again, signaling a decreasing trend in the spread of COVID-19 from an epidemiological viewpoint. However, R_{TD} began to increase again from March 19 (t=14.0). From February 7 to March 30, the average values of β and γ were 0.1656 and 0.0253, respectively.

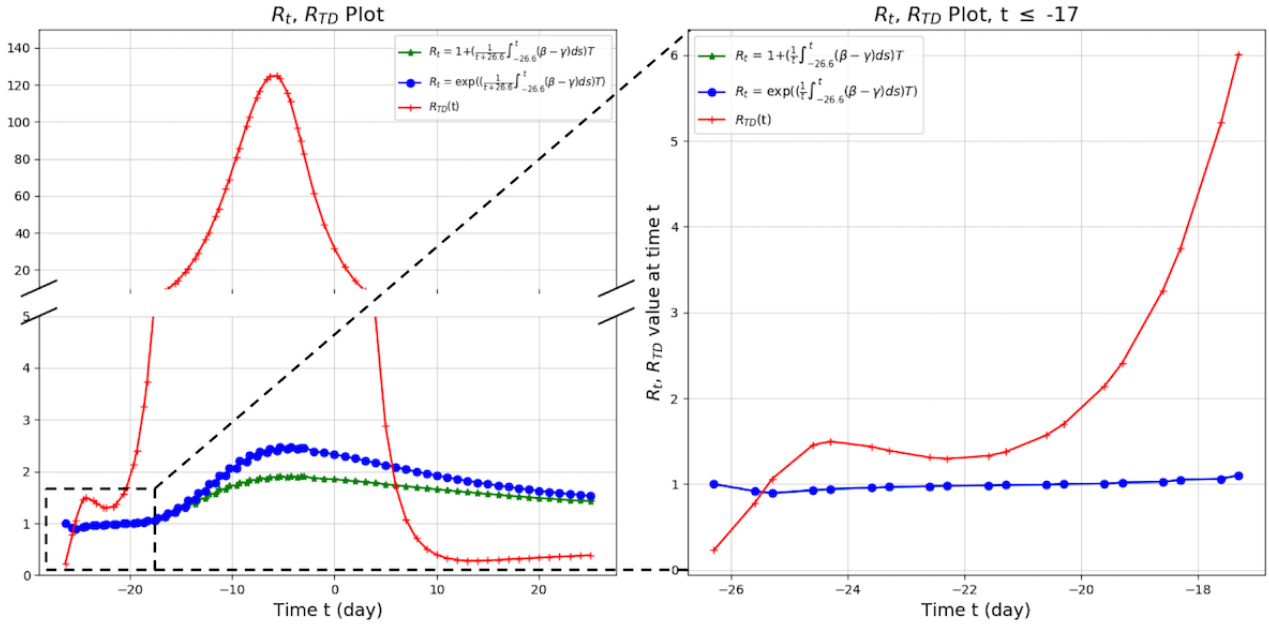
In the second case, Seoul, up to March 9 (t=4), β reached 0.2306 while γ only reached 0.0192, resulting in a maximum value of 12.0405 for R_{TD} in this period. After March 16 (t=11), R_{TD} decreased to 3.1244 but then increased again, reaching 3.8255 on March 30 (t=25). This indicates that effective control of the spread of COVID-19 was not achieved. The average values of β and γ were 0.0705 and 0.0140, respectively. In the third case, Busan, on March 5 (t=0), at the beginning of the observation, β was 0.1300 (Supplementary Table), while R_{TD} was 156.7965. This is because R(t), the recovery group, did not change in the initial stage, whereas γ was estimated to be 0.0008 due to the constraint γ>0. On March 8 (t=3), R_{TD} was 0.0908 because of the change in R(t), reaching 0.5401 on March 30 (t=25). The average values of β and γ were 0.0253 and 0.0670, respectively.

In the final case of Daegu, similar to Busan, R_{TD} was 521.9075 at the beginning of the observation on March 5 (t=0). After March 11 (t=6), the recovery rate γ began to increase faster than the infection rate β, with R_{TD} having its lowest value of 0.1224 on March 24 (t=19). After March 24, R_{TD} increased, reaching 0.2409 on March 30 (t=25) (see Figure SM3 in Multimedia Appendix 1). The average values of β and γ were 0.0191 and 0.0387, respectively. The results for other provinces are presented in figures and tables in Multimedia Appendix 1.

Time-Dependent Effective Reproduction Number (RTD)

Because R_{TD} is the ratio of β(t) to γ(t), R_{TD} can have a large value when γ is small compared to β. This situation can be observed in the early stage of COVID-19 spread in South Korea, excluding Seoul and Busan (eg, the Shincheonji cult cases). However, following the computation of the basic reproduction number in a previous study, we obtained the effective reproduction number R_t in the usual range found in previous studies [16]. In the SIR model, we approximated S as 1 because S was sufficiently large compared to I. The detailed formula is presented in Multimedia Appendix 1. R_{TD} responded more sensitively than R_t to the real-world situation from t=–26 to t=–18 (right side of Figure 7).

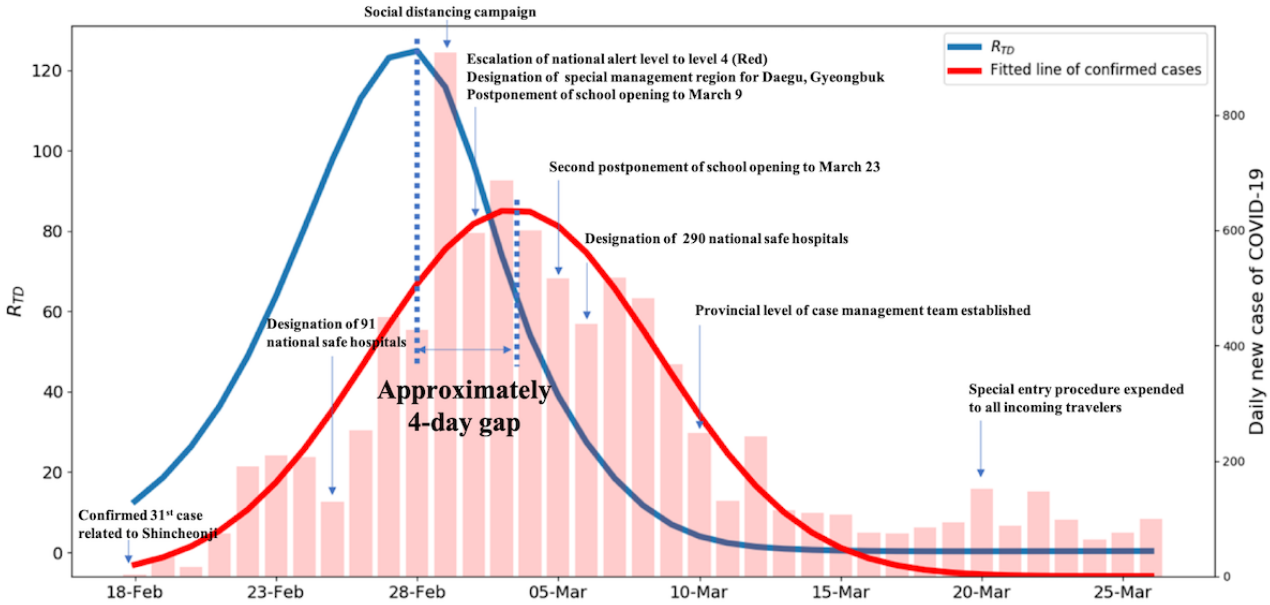
Figure 7. Comparison of R_{TD} and R_t for South Korea. R_t was computed based on the growth model. R_t: effective reproduction number; R_{TD}: time-dependent effective reproduction number.



Characteristics of RTD

R_{TD} is a more sensitive and responsive marker than R_t, and it reflects subtle changes of situations over time. Especially at the starting point of an outbreak, we can detect increasing trends more accurately with R_{TD} (Figure 7). Furthermore, R_{TD} is an indicator that precedes real-world changes. Looking at the real-world data, there is a time delay of 4 days between the peak of R_{TD} and the peak of confirmed cases (Figure 8) [17]. We also observed this pattern of time delay between the peak of R_{TD} and the peak of confirmed cases in other countries (Multimedia Appendix 1, Characteristics of R_{TD}).

Figure 8. Comparison of R_{TD} and real-world confirmed cases. R_{TD}: time-dependent effective reproduction number.



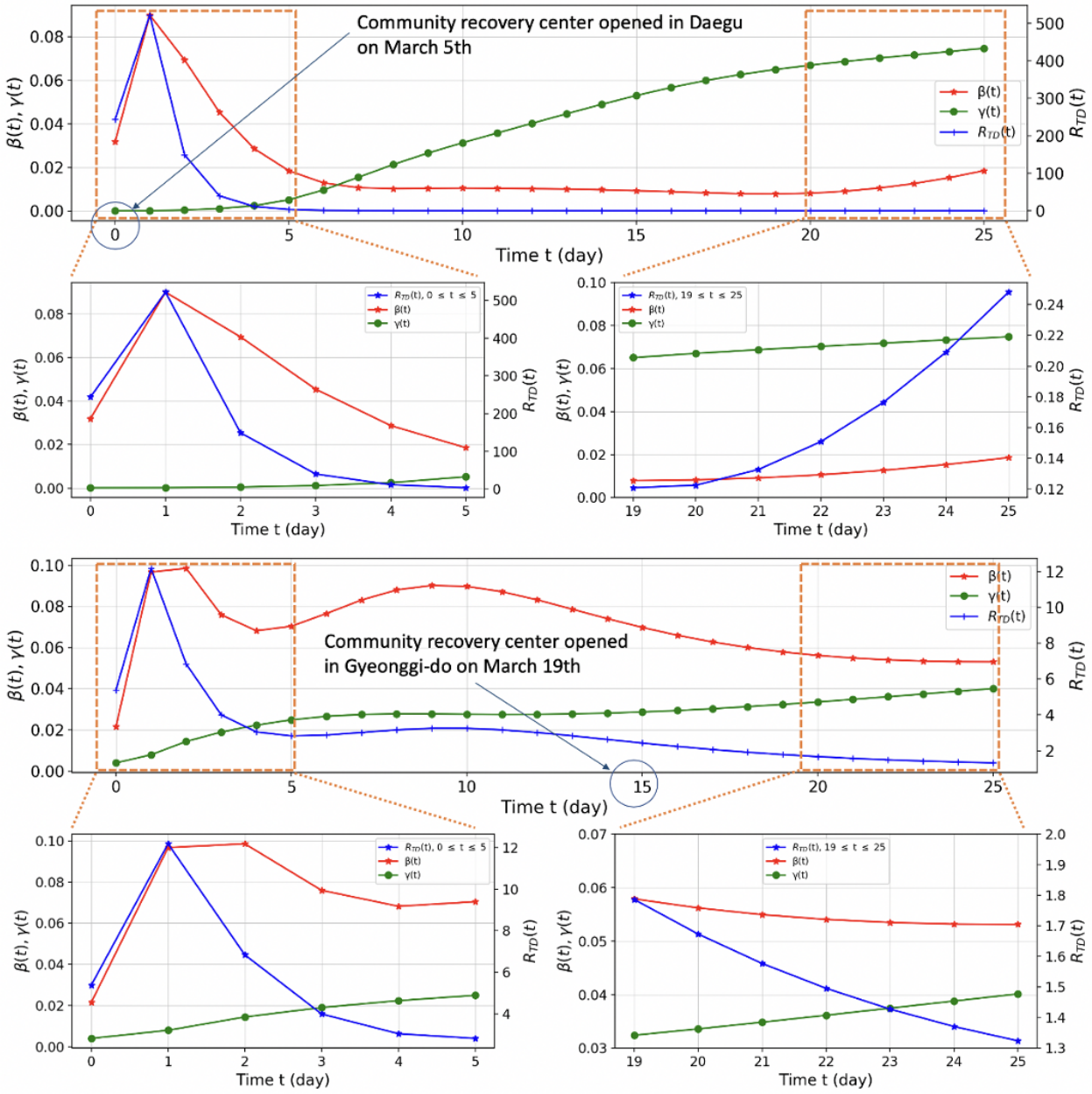
Real-World Implications of RTD

The R_{TD} we developed has important real-world implications for measuring the current status of the viral spread and the effectiveness of interventions. By setting the infection rate and recovery rate as time-dependent parameters, it is possible to accurately evaluate the pressure of depletion of health resources on the community. Indeed, after March 5, when R_{TD} exceeded 500, Daegu was in danger of total depletion of medical resources [18,19]. Patients who were self-isolating at home while waiting for hospitalization died, and a previously secured negative pressure room became full and could not continuously accept severely ill patients. In response to this situation, the Korean government opened the Community Treatment Center (CTC) to care for patients with mild illness in Daegu in early March [20]. The CTC was staffed with seven physicians, five nurses,

and several paramedic workers who monitored and cared for low-risk patients with COVID-19. The government would have been able to preemptively enact drastic policies if it had observed the changes in R_{TD} that preceded the trend of confirmed cases by approximately 4 days without any sacrifice of patients (Figure 8).

Compared to Daegu, Gyeonggi-do intervened more proactively. The R_{TD} of Daegu at the first opening of the CTC was over 500; however, that of Gyeonggi-do was 2.6. The local government in Gyeonggi-do, which closely monitored the situation in Daegu, prevented the exhaustion of medical resources by providing optimal medical services for each risk group of patients with COVID-19 in cooperation with the central government, along with general policies such as public disclosure of mobile routes of infected people, encouragement of social isolation, and wearing of masks (Figure 9).

Figure 9. Comparison of β(t), γ(t), and R_{TD} in Daegu and Gyeonggi-do. Note the differences in the vertical scales of R_{TD}. R_{TD}: time-dependent effective reproduction number.



Discussion

Novelty of the Model

We developed an entirely new SIR model with time-dependent parameters via deep learning methods. Furthermore, we validated the model with the conventional RK4 model to confirm its convergent nature. In addition, we evaluated our model based on real-world data reported by the KCDC, the Korean government, and news media.

Compared to previous studies, this research has the following three technical advantages. First, previous studies only dealt with the infected cases under certain assumptions, such as the cumulative number of infected cases increasing exponentially [21]. In our method, we can compute the effective and

time-dependent reproduction numbers without any assumptions. Moreover, we computed the entire dynamics for S, I, and R simultaneously; therefore, the analysis is more precise. Secondly, in another previous study, the authors manually divided the phase of COVID-19 spread according to the preventative and control measures to overcome the limitation of the constant reproduction number [11]. In our method, however, we did not need to artificially divide the phases because the results including S, I, and R and the parameters are naturally time-dependent. Thirdly, rather than using statistical inference techniques as in previous research, we applied a neural network to solve the forward-inverse problem consisting of the SIR model and its parameters [9]. Therefore, our method gives deterministic and more accurate values without any statistical uncertainty. Furthermore, by leveraging the neural network, our method can capture richer structures in the data and SIR model compared to the filtering techniques used in prior research [8].

Implications for Real-World Intervention

In the situation of a novel virus pandemic, it is crucial for every central and local government to maintain appropriate medical resources in readiness for unexpected penetration of the new disease. South Korea saw one of the most disastrous outbreaks of COVID-19 during the first few weeks of March 2020. In Daegu especially, the entire local medical system was on the brink of collapse. However, the Korean government soon developed a preemptive policy for each local government by learning from the situation in Daegu. The government solved its acute hospital bed shortage by revising the triage criteria more than seven times and implementing CTCs all over the country. Since then, lives were saved by reserving beds for the most acutely ill patients with COVID-19 and placing patients with less severe disease in CTCs [22].

In a country such as Korea, where there is no interregional blockade, the spread of the virus can be exacerbated in a few days due to movement of the virus across regions. In fact, the number of COVID-19 cases started increasing again from March 19 (t=14.0), indicating that the containment of COVID-19 cannot be realized without achieving herd immunity or developing therapeutics.

Furthermore, we require a tool that can monitor virus outbreaks simultaneously across regions in the shortest time span.

The same principle applies even if we broaden our view from the spread of viruses between regions to the spread among countries. In the current COVID-19 pandemic, the world must work together to prevent the spread of the virus. This is because the entire world is socially, culturally, and economically intertwined through advanced transportation. Therefore, there is an urgent need for a tool that can respond sensitively over time, provide information about the current virus outbreak, and evaluate the effectiveness of interventions. The methodology and new model of this study could be employed for proactive intervention. In addition, from the perspective of measuring medical resources, our model has powerful strength because it assumes all the parameters as time-dependent, which reflects the exact status of viral spread. Furthermore, the methodology and modeling approach are scalable and universal; therefore, they can be applied to other new infectious disease pandemics if real-world data are available.

Limitations

This research has several limitations. First, the time-dependent model of this study was validated only with COVID-19 data from South Korea. However, this model can be easily applied to data from another outbreak because the modeling process and methodology are disclosed fully in this article. To crossvalidate our strategies, we provide results of similar analyses of outbreaks in Italy, Sweden, and the United States in Multimedia Appendix 1 (see Figures SM5 to SM10). Secondly, because of the nature of deep learning, the results of the model may have been overfitted to South Korean data. However, with the new approach of this research, it is more feasible and reasonable for every researcher to adopt the modeling methodology and apply the model by training it with local data that reflect local situations. In this case, an overfitted model can be reinterpreted as a model that is appropriately fitted to the local situation or that reflects the characteristics of the region.

Acknowledgments

HJ.H was supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (Nos. 2017R1E1A1A03070105 and NRF-2019R1A5A1028324) and by an Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2019-0-01906, Artificial Intelligence Graduate School Program (POSTECH)).

Authors' Contributions

SYJ and HTJ drafted the entire manuscript as first authors. HJS contributed to the discussion of the data. HJH supervised the entire process as the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mathematical formulation of the SIR model and supplemental figures.

[DOCX File , 2745 KB - jmir_v22i9e19907_app1.docx]

References

1. Spanish flu. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=Spanish_flu&oldid=949335031 [accessed 2020-04-06]
2. Kassraie A. Spanish Flu: How America Fought a Pandemic a Century Ago. AARP. 2020 Mar 20. URL: http://www.aarp.org/politics-society/history/info-2020/spanish-flu-pandemic.html [accessed 2020-04-15]
3. Strochlic N, Champine RD. How some cities 'flattened the curve' during the 1918 flu pandemic. National Geographic. 2020 Mar 27. URL: https://www.nationalgeographic.com/history/2020/03/how-cities-flattened-curve-1918-spanish-flu-pandemic-coronavirus/ [accessed 2020-04-15]
4. Ferguson N, Laydon D, SSP PR, Casella F, Nedjati Gilani G, Imai N, et al. Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. MRC Centre for Global Infectious Disease Analysis 2020 Mar 16:online. [doi: 10.25561/77482]
5. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020 Mar 13;27(2):5-20 [FREE Full text] [doi: 10.1093/jtm/taaa021] [Medline: 32052846]
6. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis 2020 Apr;93:211-216 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.058] [Medline: 32145465]
7. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number (R0). Emerg Infect Dis 2019 Jan;25(1):1-4 [FREE Full text] [doi: 10.3201/eid2501.171901] [Medline: 30560777]
8. 8 CY, Lu P, Chang C. A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons. arXiv q-bio. 2020 Feb 28. URL: http://arxiv.org/abs/2003.00122 [accessed 2020-09-02]
9. You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. Int J Hyg Environ Health 2020 Jul;228:113555 [FREE Full text] [doi: 10.1016/j.ijheh.2020.113555] [Medline: 32460229]
10. Dandekar R, Barbastathis G. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning. medRxiv 2020 Apr 06:preprint [FREE Full text] [doi: 10.1101/2020.04.03.20052084]
11. Wang H, Wang Z, Dong Y, Chang R, Xu C, Yu X, et al. Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. Cell Discov 2020 Feb 24;6(1):10 [FREE Full text] [doi: 10.1038/s41421-020-0148-0] [Medline: 32133152]
12. Feng X, Chen J, Wang K, Wang L, Zhang F, Jin Z, et al. Phase-adjusted estimation of the COVID-19 outbreak in South Korea under multi-source data and adjustment measures: a modelling study. Math Biosci Eng 2020 May 18:3637-3648. [doi: 10.3934/mbe.2020205]
13. Korea Centers for Disease Control and Prevention. Webpage in Korean. URL: http://www.cdc.kr/ [accessed 2020-09-02]
14. Raissi M, Perdikaris P, Karniadakis G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 2019 Feb;378:686-707. [doi: 10.1016/j.jcp.2018.10.045]
15. 14 JH, Son H, Hwang H. Deep Neural Network Approach to Forward-Inverse Problems. arXiv math. 2020 Jul 27. URL: http://arxiv.org/abs/1907.12925 [accessed 2020-09-02]
16. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci 2007 Feb 22;274(1609):599-604 [FREE Full text] [doi: 10.1098/rspb.2006.3754] [Medline: 17476782]
17. Flattening the curve on COVID-19: How Korea responded to a pandemic using ICT. Korea Ministry of the Interior and Safety. 2020 Apr 16. URL: https://mois.go.kr/eng/bbs/type002/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000022&nttId=76748 [accessed 2020-04-20]
18. Park HN. Measures on Daegu hospital shortage coming: Prime minister. The Korea Herald. 2020 Mar 01. URL: http://www.koreaherald.com/view.php?ud=20200301000210 [accessed 2020-04-22]
19. Thousands wait for hospital beds in South Korea as coronavirus cases surge. The Japan Times. 2020 Mar 04. URL: https://www.japantimes.co.jp/news/2020/03/04/asia-pacific/thousands-wait-hospital-beds-south-korea-coronavirus-cases-surge/ [accessed 2020-04-22]
20. Kim YM. Community treatment centers save medical resources in pandemic. Korea Biomedical Review. 2020 Apr 01. URL: http://www.koreabiomed.com/news/articleView.html?idxno=7897 [accessed 2020-04-22]
21. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. Int J Infect Dis 2020 Apr;93:339-344 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.031] [Medline: 32198088]
22. Yoon D. How South Korea Solved Its Acute Hospital-Bed Shortage. The Wall Street Journal. URL: https://www.wsj.com/articles/how-south-korea-solved-its-acute-hospital-bed-shortage-11584874801 [accessed 2020-04-15]

Abbreviations

CSSE: Center for Systems Science and Engineering

CTC: Community Treatment Center
DNN: deep neural network
KCDC: Korea Centers for Disease Control and Prevention
R: reproduction number
R_0: basic reproduction number
R_t: effective reproduction number
R_{TD}: time-dependent effective reproduction number
RK4: Runge-Kutta fourth order
SIR: susceptible-infected-recovered

Edited by G Eysenbach, G Fagherazzi, C Lovis; submitted 06.05.20; peer-reviewed by YR Park, S Pit, Y Zhong; comments to author 29.06.20; revised version received 23.07.20; accepted 01.09.20; published 09.09.20.

Please cite as:
Jung SY, Jo H, Son H, Hwang HJ
Real-World Implications of a Rapidly Responsive COVID-19 Spread Model with Time-Dependent Parameters via Deep Learning: Model Development and Validation
J Med Internet Res 2020;22(9):e19907
URL: http://www.jmir.org/2020/9/e19907/
doi:10.2196/19907
PMID:32877350

Original Paper

Patient Satisfaction With Telemedicine During the COVID-19 Pandemic: Retrospective Cohort Study

Ashwin Ramaswamy[1], MD, MPP; Miko Yu[1], MA; Siri Drangsholt[1], MD; Eric Ng[1], MS; Patrick J Culligan[1], MD; Peter N Schlegel[1], MD; Jim C Hu[1], MD, MPH

Department of Urology, Weill Cornell Medicine, New York, NY, United States

Corresponding Author:
Ashwin Ramaswamy, MD, MPP
Department of Urology
Weill Cornell Medicine
525 East 68th Street
New York, NY, 10065
United States
Phone: 1 646 962 9600
Email: asr9066@nyp.org

Abstract

Background: New York City was the international epicenter of the COVID-19 pandemic. Health care providers responded by rapidly transitioning from in-person to video consultations. Telemedicine (ie, video visits) is a potentially disruptive innovation; however, little is known about patient satisfaction with this emerging alternative to the traditional clinical encounter.

Objective: This study aimed to determine if patient satisfaction differs between video and in-person visits.

Methods: In this retrospective observational cohort study, we analyzed 38,609 Press Ganey patient satisfaction survey outcomes from clinic encounters (620 video visits vs 37,989 in-person visits) at a single-institution, urban, quaternary academic medical center in New York City for patients aged 18 years, from April 1, 2019, to March 31, 2020. Time was categorized as pre–COVID-19 and COVID-19 (before vs after March 4, 2020). Wilcoxon-Mann-Whitney tests and multivariable linear regression were used for hypothesis testing and statistical modeling, respectively.

Results: We experienced an 8729% increase in video visit utilization during the COVID-19 pandemic compared to the same period last year. Video visit Press Ganey scores were significantly higher than in-person visits (94.9% vs 92.5%; $P<.001$). In adjusted analyses, video visits (parameter estimate [PE] 2.18; 95% CI 1.20-3.16) and the COVID-19 period (PE 0.55; 95% CI 0.04-1.06) were associated with higher patient satisfaction. Younger age (PE –2.05; 95% CI –2.66 to –1.22), female gender (PE –0.73; 95% CI –0.96 to –0.50), and new visit type (PE –0.75; 95% CI –1.00 to –0.49) were associated with lower patient satisfaction.

Conclusions: Patient satisfaction with video visits is high and is not a barrier toward a paradigm shift away from traditional in-person clinic visits. Future research comparing other clinic visit quality indicators is needed to guide and implement the widespread adoption of telemedicine.

(J Med Internet Res 2020;22(9):e20786) doi:10.2196/20786

KEYWORDS

telemedicine; medicine; pandemics; patient satisfaction; remote consultation; disruptive technology; medical informatics; health care delivery; practice patterns; physicians; health policy; health services research; health care reform; COVID-19

Introduction

New York City was the world's COVID-19 epicenter in early 2020 [1]. As of May 8, 2020, the five boroughs had 175,997 confirmed cases and 14,381 deaths, comprising 14% of all confirmed cases and 19% of all deaths from COVID-19 in the United States [2]. Health care providers postponed elective surgeries, expanded intensive care unit (ICU) capacity, deployed nursing and physician staff, and rapidly transitioned most clinic encounters to telemedicine (defined here as synchronous video visits) [3].

Historically, telemedicine focused on rural medicine [4] and/or moved forward incrementally through institutional initiatives [5]. The widespread adoption of telemedicine associated with

the COVID-19 pandemic was unprecedented and may have a significant and durable impact on health care delivery. Telemedicine has not commonly been tested in disaster settings [6]. It was an essential component of the medical response to COVID-19 by reducing demand on strained health care infrastructure and enabling health care needs to be met at home while reducing exposure for patients and medical staff [7,8]. Patient demand for telemedicine outstripped the ability of health care providers to supply it [9]. In early March, the Centers for Medicare and Medicaid Services established telemedicine payment parity with in-person visits, suspended licensure and malpractice insurance restrictions, and waived HIPAA (Health Insurance Portability and Accountability Act) regulations regarding video visits [10] to limit barriers to widespread adoption of telemedicine.

We examined patient acceptance of video visits by comparing Press Ganey patient satisfaction scores for video vs in-person visits at an urban, quaternary referral, academic medical center from April 1, 2019, to March 31, 2020. We hypothesized that there would be no difference in Press Ganey patient satisfaction scores between video and in-person visits. We captured one month of clinic visits during the COVID-19 pandemic and sought to determine the factors associated with patient satisfaction during this time frame.

Methods

NewYork-Presbyterian/Weill Cornell Medical Center (NYP/WCM) is a large nonprofit academic medical center located in New York City. As of May 8, 2020, NYP/WCM has admitted a total of 1443 COVID-19 patients. At our institution, inpatient services are provided at NYP, while outpatient services are provided predominately at WCM facilities; both institutions share the same providers.

Data Source

We used a customized version of the Press Ganey Outpatient Medical Practice Survey to evaluate patient satisfaction following clinic encounters at WCM from April 1, 2019, to March 31, 2020. The Press Ganey survey is used by more than 26,000 health care organizations, including over 60% of all US hospitals. It is the most commonly used, validated tool for assessing patient satisfaction in the outpatient setting [11]. The data contained deidentified patient-level data with the following variables: date of survey, visit type, patient age, gender, first visit (yes vs no), and Press Ganey satisfaction score (0%-100%). Visit type in Press Ganey is the same as the outpatient visit category coded in the outpatient electronic health record (Epic). There were over 200 visit types and 40 specialties represented. All video visits were synchronous video-based provider-patient visits scheduled and accessed through the outpatient enterprise electronic health record. WCM has been reimbursed with telemedicine payment parity since 2018.

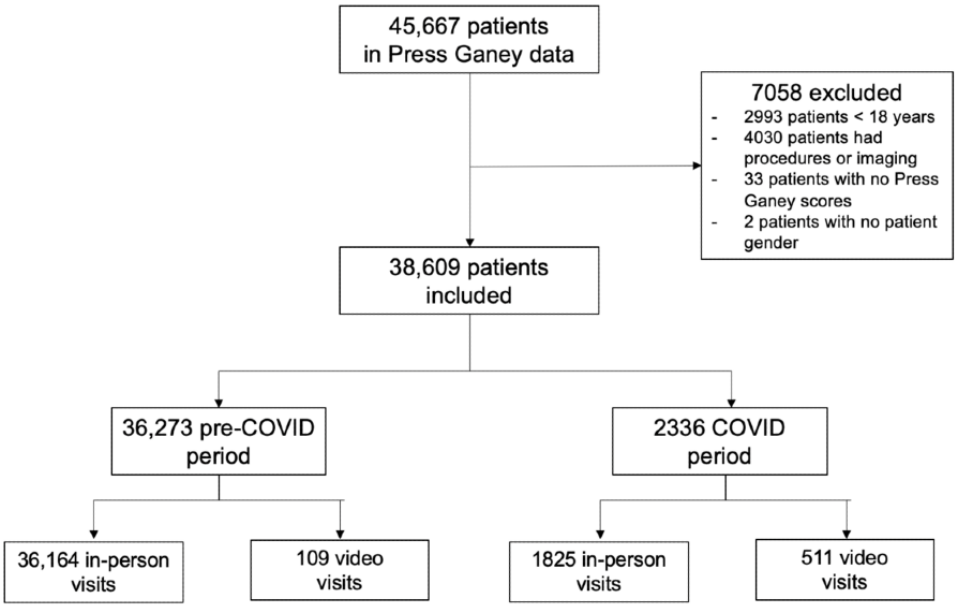
The WCM Press Ganey Medical Practice Survey contains 31 items assessed on a 5-point Likert scale (ie, very poor, poor, fair, good, very good) to evaluate seven domains of patient care: Background Questions (3 items), Access (8 items), Moving Through Your Visit (4 items), Nurse/Assistant (3 items), Care Provider (6 items), Personal Issues (4 items), and Overall Assessment (3 items). WCM surveyed patients with 19 items from the standardized Press Ganey Outpatient Medical Practice Survey; the remaining 12 were by WCM from the Press Ganey item bank. The same survey instrument was used across all specialties and providers without variation for video vs in-person visits. Press Ganey sent the survey instrument 2-3 days after completion of the outpatient visit or video visit. Press Ganey then reported deidentified satisfaction scores to WCM without linkage to the patient's electronic health record to maintain confidentiality.

Study Population

We performed a retrospective study of patients aged 18 years and older. In order to adjust for the COVID-19 pandemic, we categorized visits after March 3, 2020, as the *COVID-19 period*. This timing corresponds with a WCM mandate to shift the majority of outpatient care from in-person to video visits.

Our total data included 45,667 outpatient visits across 210 visit categories with 2670 (5.8%) outpatients visits during the COVID-19 period. We defined the study group as consisting of video visits, which were identified if the visit type contained the words "Video Visit." A total of 7058 outpatient visits were excluded: 4030 outpatient visits involving procedures (eg, surgery, venipuncture) or imaging, 2993 visits comprising pediatric patients <18 years of age, 33 visits that did not have Press Ganey scores, and 2 visits that did not include patient gender. The comparison group consisted of in-person outpatient visits. After applying our exclusion criteria, our final WCM Press Ganey data included 38,609 visits across 88 visit categories; of these, 620 (1.6%) video visits constituted the study population (Figure 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram detailing the inclusion and exclusion criteria for our study.



Statistical Analysis

Independent variables were compared by pre–COVID-19 vs COVID-19 period using the paired *t* test and the chi-squared test. Hypothesis testing was conducted using the nonparametric Wilcoxon-Mann-Whitney test, comparing Press Ganey satisfaction scores between in-person and video visits across the study period. Additionally, we compared in-person and video visits in the pre–COVID-19 period to in-person and video visits during the COVID-19 period, respectively.

The dependent variable in our study was the Press Ganey patient satisfaction score. We fit a multivariable linear regression model with the following covariables: video visit (vs in-person), gender, age (18-25, 26-39, 40-59, 60-79, and 80 years [reference]), COVID-19 period (yes vs no), and new vs established visit.

Significance was established at $P<.05$. Statistical analysis was performed in R (version 4.0.0, The R Project for Statistical Computing) and STATA 14.0 (StataCorp LLC). The study was approved by the WCM Institutional Review Board and patient consent was not required for this study.

Results

NYP/WCM experienced an 8729% increase in video visit use during the COVID-19 period compared to the pre–COVID-19 period.

The mean age of the total study population was 58.8 years (SD 16.5 years). Patients were slightly older in the pre–COVID-19 period than COVID-19 period (59.85 years vs 59.08 years; $P=.03$) (Table 1). There were no differences in gender or other visit characteristics between the two time periods.

In the pre–COVID-19 period, very few outpatient visits types were video visits (0.3%). During the COVID-19 period, video visits comprised 21.9% of outpatient visits. The proportion of all in-person visit types decreased except for postoperative visits. The visit type “follow-up” showed the greatest decline in share of outpatient visits, decreasing from 50.0% to 36.5% over the past year. Internal medicine outpatient visits constituted the greatest proportion of visits in both time periods.

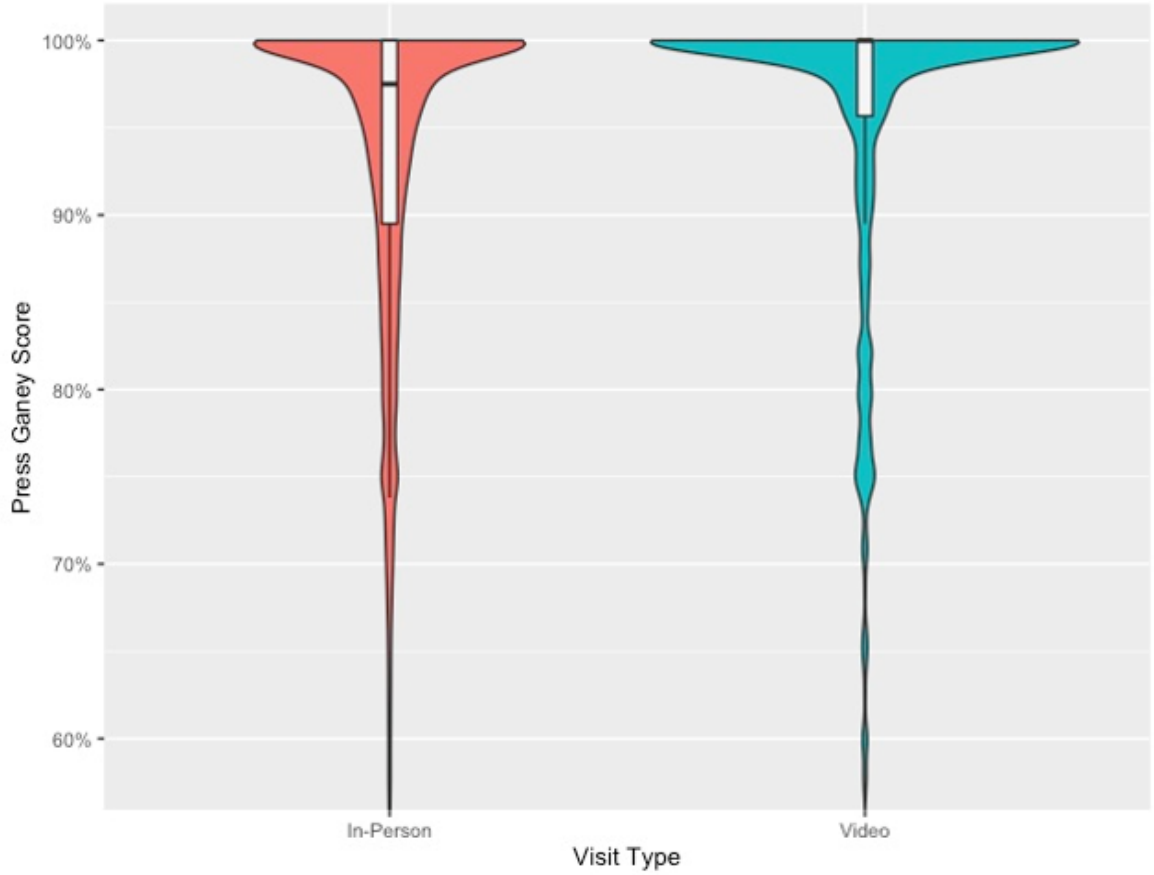
Table 1. Baseline characteristics of clinic visits before and after COVID-19.

Characteristic	Pre–COVID-19 (April 1, 2019 to March 3, 2020) (n=36,273)	COVID-19 (March 4-31, 2020) (n=2336)	*P* value
Press Ganey score (%), mean (SD)			
All visits	92.47 (11.25)	93.43 (10.51)	<.001
Telemedicine	95.01 (8.65)	94.87 (10.22)	.31
In-person	92.46 (11.26)	93.02 (10.56)	.004
Age (years)			
Median (IQR)	63 (48-72)	62 (47-72)	—[a]
Mean (SD)	59.85 (16.49)	59.08 (16.15)	.03
Category, n (%)			<.001
18-25	853 (2.35)	46 (1.97)	
26-39	4827 (13.31)	315 (13.48)	
40-59	9526 (26.26)	713 (30.52)	
60-79	17,911 (49.38)	1089 (46.62)	
≥80	3156 (8.70)	173 (7.41)	
Sex, n (%)			.29
Male	14,444 (39.82)	904 (38.70)	
Female	21,829 (60.18)	1432 (61.30)	
Type of visit, n (%)			<.001
Video	109 (0.30)	511 (21.88)	
Follow-up	18,131 (49.98)	852 (36.47)	
New patient	7606 (20.97)	393 (16.82)	
Established well visit	3204 (8.83)	157 (6.72)	
Consultation	1224 (3.37)	74 (3.17)	
New well visit	1215 (3.35)	77 (3.30)	
Follow-up (complex)	1132 (3.12)	55 (2.35)	
Physical	772 (2.13)	44 (1.88)	
Post-op	390 (1.08)	30 (1.28)	
Other	2490 (6.86)	143 (6.12)	
Visit type, n (%)			.15
New	9816 (27.06)	665 (28.47)	
Existing	26,457 (72.94)	1671 (71.53)	
Specialty			<.001
Internal medicine	7427 (20.47)	626 (26.80)	
Obstetrics/gynecology	3287 (9.06)	206 (8.82)	
Cardiology	3015 (8.31)	163 (7.00)	
Ophthalmology	2863 (7.89)	144 (6.16)	
Otolaryngology	2611 (7.20)	133 (5.69)	
Hematology/oncology	1934 (5.33)	119 (5.09)	
Dermatology	1646 (4.54)	84 (3.60)	
Other	13,491 (37.19)	861 (36.86)	

[a]Not applicable.

Press Ganey patient satisfaction scores were significantly higher in the COVID-19 period when compared to the pre–COVID-19 period (93.4% vs 92.5%, *P*<.001). Notably, across the study period, patient satisfaction with video was significantly higher than in-person visits (94.9% vs 92.5%, *P*<.001) (Figure 2); this association was consistent during the pre–COVID-19 (95.0% vs 92.5%, *P*<.001) and COVID-19 periods (94.9% vs 93.0%, *P*<.001). While Press Ganey scores with video visits did not change across time periods (95.0% vs 94.9%, *P*=.31), the scores for in-person visits increased in the COVID-19 period (92.5% vs 93.0%, *P*=.004).

Figure 2. Violin and box-and-whiskers plot depicting the unadjusted distribution of in-person vs video visits.



Overall, all of our covariables were statistically significant (Table 2). In adjusted analyses, video visits (parameter estimate [PE] 2.18; 95% CI 1.20-3.16), the COVID-19 period (PE 0.55; 95% CI 0.04-1.06), and the age category 60-79 years (PE 0.70; 95% CI 0.29-1.11) were associated with higher Press Ganey scores. Female gender (PE –0.73; 95% CI –0.96 to –0.50) and new visit type (PE –0.75; 95% CI –1.00 to –0.49) were associated with lower Press Ganey satisfaction scores.

Table 2. Multivariable linear regression for variables predicting Press Ganey scores.

Variable	Press Ganey score		*P* value
	Parameter estimate	95% CI	
Telemedicine	2.18	1.20 to 3.16	<.001
Female	–0.73	–0.96 to –0.50	<.001
Age (years) (reference: ≥80 years)			
18-25	–2.05	–2.88 to –1.22	<.001
26-39	–1.95	–2.45 to –1.46	<.001
40-59	–0.66	–1.10 to –0.22	.003
60-79	0.70	0.29 to 1.11	.001
COVID-19 period	0.55	0.04 to 1.06	.04
New visit	–0.75	–1.00 to –0.49	<.001

Discussion

Traditionally, health care encounters between a provider and a patient have occurred face to face in a physical location. Over the past 20 years, the internet and technology have made it possible for health care to be delivered digitally, providing new avenues for medicine to improve the value of care. Bridging gaps in time and distance, video visits have enabled providers to remotely care for patients with acute stroke [5], requiring intensive care [12], and located in rural areas or prison [13]. Half the hospitals in the United States report providing telehealth-based services [14]. Two years ago, the concept of a "medical virtualist" was created to describe a new specialty in which physicians primarily deliver care digitally [15]. The strengths of telemedicine have made it an indispensable tool in the clinical response to the COVID-19 pandemic. With the removal of financial disincentives and privacy barriers that limited widespread adoption, use of telemedicine has grown substantially in the United States during the COVID-19 pandemic. The 8729% increase in video visit utilization at our academic medical center is akin to the 4345% increase at New York University Langone Health [16] and the 4000% increase at Partners Healthcare [17].

Telemedicine is a new, and potentially disruptive, innovation and must be shown to be safe, effective, patient-centered, timely, efficient, and equitable [18]. Clinical consultations conducted through video visits are associated with high patient satisfaction [19,20] and lower costs [21-23] without a difference in clinical outcomes [13,24-27] compared to in-person consultations. However, most of these conclusions are based on evidence from small studies focused on remote telemedicine in sparsely populated locations or highly specific patient populations (eg, stroke care, rural ICUs, prisons) [5,12,13] not relevant to the care delivered during the COVID-19 pandemic [28]. Of all studies evaluating quality in telemedicine, four systematic reviews demonstrate there is limited published evidence to evaluate patient satisfaction as a metric for comparison to in-person visits [29-32]. This is problematic because patient satisfaction has been cited as the most important factor in the success of telemedicine initiatives [33]. Patient satisfaction as a measure of quality of care is a valid outcome [34] and a key component to value-based care [35]. Patient satisfaction is also associated with treatment plan adherence [36], reduced surgical readmissions [37], and patient retention [38].

Using the most current Press Ganey satisfaction scores, we found that video visits were associated with greater patient satisfaction when compared to in-person visits, which was not what we initially hypothesized. However, our results do not justify the use of telemedicine in lieu of in-person visits if both were equally accessible given the limitations outlined below. Furthermore, we observed that overall patient satisfaction is higher in the COVID-19 period, and that younger age, female gender, and new visit type were associated with lower patient satisfaction. To our knowledge, this is the largest study of patient satisfaction comparing video to in-person visits. The results of our study have particular relevance due to the unprecedented public health crisis that has necessitated the widespread adoption of video visits for patient safety and practicality.

Our study must be interpreted in the context of the study design. First, this is a retrospective study that prevents us from establishing causality between video visits and increased patient satisfaction. However, our data captures a rapid transition of outpatient care to video visits associated with the COVID-19 pandemic, and our analysis was adjusted for time. Second, our deidentified data did not capture patient-level variables that may influence Press Ganey scores, such as race, income, education, comorbidities, and other characteristics as they were not reported to WCM by Press Ganey to maintain patient confidentiality. Third, the same Press Ganey survey items were used for both video and in-person visits even though Press Ganey developed a new telemedicine version of the Medical Practice Survey in 2018 that may better characterize unique aspects pertinent to video visits [39]. Fourth, we were unable to assess for nonresponder bias; respondents to Press Ganey surveys have been shown to be more satisfied and more willing to respond than nonrespondents [40]. Fifth, what constitutes a clinically significant difference in patient satisfaction using Press Ganey survey scores is not well established. However, that said, there is mounting evidence in recent years that Press Ganey patient satisfaction score increases of 1%-10% can be viewed as clinically relevant [41-44].

In conclusion, we demonstrate that patient satisfaction with video visits compared favorably with in-person visits over the past year and during the COVID-19 pandemic. Our findings support the use of video visits as a viable alternative to traditional in-person visits. The New York City experience may offer insights into the future use of video visits as a new paradigm for health care delivery generally and in times of public health crisis.

Conflicts of Interest

PJC is a stockholder in Origami Surgical and has received consulting fees from Intuitive Surgical, Hologic, Coloplast. JCH receives research support from the Frederick J and Theresa Dow Wallace Fund of the New York Community Trust. PNS reports consulting fees from Roman Health, Gilead Sciences, and Theralogix. The remaining authors report no further disclosures related to this work.

References

1. Why New York is the epicenter of the American coronavirus outbreak. CNN. 2020 Mar. URL: https://www.cnn.com/2020/03/26/us/new-york-coronavirus-explainer/index.html [accessed 2020-05-10]

2. Johns Hopkins Coronavirus Resource Center. COVID-19 United States Cases by County. URL: https://coronavirus.jhu.edu/us-map [accessed 2020-05-09]
3. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
4. Telehealth Parity Laws. Health Affairs. URL: https://www.healthaffairs.org/do/10.1377/hpb20160815.244795/full/ [accessed 2020-05-07]
5. Levine SR, Gorman M. "Telestroke" : the application of telemedicine for stroke. Stroke 1999 Feb;30(2):464-469 [FREE Full text] [doi: 10.1161/01.str.30.2.464] [Medline: 9933289]
6. Lurie N, Carr BG. The Role of Telehealth in the Medical Response to Disasters. JAMA Intern Med 2018 Jun 01;178(6):745-746. [doi: 10.1001/jamainternmed.2018.1314] [Medline: 29710200]
7. Ross SW, Lauer CW, Miles WS, Green JM, Christmas AB, May AK, et al. Maximizing the Calm before the Storm: Tiered Surgical Response Plan for Novel Coronavirus (COVID-19). J Am Coll Surg 2020 Jun;230(6):1080-1091.e3 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2020.03.019] [Medline: 32240770]
8. Bloem BR, Dorsey ER, Okun MS. The Coronavirus Disease 2019 Crisis as Catalyst for Telemedicine for Chronic Neurological Disorders. JAMA Neurol 2020 Apr 24:e. [doi: 10.1001/jamaneurol.2020.1452] [Medline: 32329796]
9. Hong Y, Lawrence J, Williams D, Mainous III A. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. JMIR Public Health Surveill 2020 Apr 07;6(2):e18961 [FREE Full text] [doi: 10.2196/18961] [Medline: 32250963]
10. President Trump Expands Telehealth Benefits for Medicare Beneficiaries During COVID-19 Outbreak. CMS. 2020 May. URL: https://www.cms.gov/newsroom/press-releases/president-trump-expands-telehealth-benefits-medicare-beneficiaries-during-covid-19-outbreak [accessed 2020-05-06]
11. Hopkins B, Patel M, Yamaguchi J, Cloney M, Dahdaleh N. Predictors of patient satisfaction and survey participation after spine surgery: a retrospective review of 17,853 consecutive spinal patients from a single academic institution. Part 1: Press Ganey. J Neurosurg Spine 2019 Jan 04;30(3):382-388. [doi: 10.3171/2018.8.SPINE18594] [Medline: 30611140]
12. Avdalovic MV, Marcin JP. When Will Telemedicine Appear in the ICU? J Intensive Care Med 2019 Apr 13;34(4):271-276. [doi: 10.1177/0885066618775956] [Medline: 29756527]
13. Arora S, Thornton K, Murata G, Deming P, Kalishman S, Dion D, et al. Outcomes of Treatment for Hepatitis C Virus Infection by Primary Care Providers. N Engl J Med 2011 Jun 09;364(23):2199-2207. [doi: 10.1056/nejmoa1009370]
14. Jain S, Khera R, Lin Z, Ross JS, Krumholz HM. Availability of Telemedicine Services Across Hospitals in the United States in 2018: A Cross-sectional Study. Annals of Internal Medicine 2020 Apr 30:e. [doi: 10.7326/m20-1201]
15. Nochomovitz M, Sharma R. Is It Time for a New Medical Specialty?: The Medical Virtualist. JAMA 2018 Feb 06;319(5):437-438. [doi: 10.1001/jama.2017.17094] [Medline: 29181539]
16. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
17. Schwamm LH, Erskine A, Licurse A. A digital embrace to blunt the curve of COVID19 pandemic. NPJ Digit Med 2020 May 4;3(1):64 [FREE Full text] [doi: 10.1038/s41746-020-0279-6] [Medline: 32377575]
18. Schwamm LH. Telehealth: seven strategies to successfully implement disruptive technology and transform health care. Health Aff (Millwood) 2014 Feb;33(2):200-206. [doi: 10.1377/hlthaff.2013.1021] [Medline: 24493761]
19. Sabesan S, Simcox K, Marr I. Medical oncology clinics through videoconferencing: an acceptable telehealth model for rural patients and health workers. Intern Med J 2012 Jul;42(7):780-785. [doi: 10.1111/j.1445-5994.2011.02537.x] [Medline: 21627743]
20. Dorsey ER, Venkataraman V, Grana MJ, Bull MT, George BP, Boyd CM, et al. Randomized controlled clinical trial of "virtual house calls" for Parkinson disease. JAMA Neurol 2013 May 01;70(5):565-570 [FREE Full text] [doi: 10.1001/jamaneurol.2013.123] [Medline: 23479138]
21. Baker LC, Johnson SJ, Macaulay D, Birnbaum H. Integrated telehealth and care management program for Medicare beneficiaries with chronic disease linked to savings. Health Aff (Millwood) 2011 Sep;30(9):1689-1697 [FREE Full text] [doi: 10.1377/hlthaff.2011.0216] [Medline: 21900660]
22. Nord G, Rising KL, Band RA, Carr BG, Hollander JE. On-demand synchronous audio video telemedicine visits are cost effective. Am J Emerg Med 2019 May;37(5):890-894. [doi: 10.1016/j.ajem.2018.08.017] [Medline: 30100333]
23. Jennett PA, Affleck Hall L, Hailey D, Ohinmaa A, Anderson C, Thomas R, et al. The socio-economic impact of telehealth: a systematic review. J Telemed Telecare 2003 Dec;9(6):311-320. [doi: 10.1258/135763303771005207] [Medline: 14680514]
24. Su D, Zhou J, Kelley MS, Michaud TL, Siahpush M, Kim J, et al. Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials. Diabetes Res Clin Pract 2016 Jun;116:136-148 [FREE Full text] [doi: 10.1016/j.diabres.2016.04.019] [Medline: 27321329]
25. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. The Lancet 2017 Sep;390(10098):959-968. [doi: 10.1016/s0140-6736(17)31327-2]

26. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan 01;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]
27. Portnoy JM, Waller M, De Lurgio S, Dinakar C. Telemedicine is as effective as in-person visits for patients with asthma. Ann Allergy Asthma Immunol 2016 Sep;117(3):241-245. [doi: 10.1016/j.anai.2016.07.012] [Medline: 27613456]
28. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
29. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The Use and Effects of Electronic Health Tools for Patient Self-Monitoring and Reporting of Outcomes Following Medication Use: Systematic Review. J Med Internet Res 2018 Dec 18;20(12):e294 [FREE Full text] [doi: 10.2196/jmir.9284] [Medline: 30563822]
30. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 03;320(7248):1517-1520 [FREE Full text] [doi: 10.1136/bmj.320.7248.1517] [Medline: 10834899]
31. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013 Aug 19;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]
32. Kraai I, Luttik M, de Jong R, Jaarsma T, Hillege H. Heart failure patients monitored with telemedicine: patient satisfaction, a review of the literature. J Card Fail 2011 Aug;17(8):684-690. [doi: 10.1016/j.cardfail.2011.03.009] [Medline: 21807331]
33. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
34. Kupfer JM, Bond EU. Patient satisfaction and patient-centered care: necessary but not equal. JAMA 2012 Jul 11;308(2):139-140. [doi: 10.1001/jama.2012.7381] [Medline: 22782413]
35. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. N Engl J Med 2016 Feb 11;374(6):504-506. [doi: 10.1056/nejmp1511701]
36. Zolnierek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
37. Lobo Prabhu K, Cleghorn MC, Elnahas A, Tse A, Maeda A, Quereshy FA, et al. Is quality important to our patients? The relationship between surgical outcomes and patient satisfaction. BMJ Qual Saf 2018 Jan 03;27(1):48-52. [doi: 10.1136/bmjqs-2017-007071] [Medline: 29101291]
38. Safran DG, Montgomery JE, Chang H, Murphy J, Rogers WH. Switching doctors: predictors of voluntary disenrollment from a primary physician's practice. J Fam Pract 2001 Feb;50(2):130-136. [Medline: 11219560]
39. Press Ganey Introduces New Telemedicine Solution for Medical Practices. Press Ganey. 2018 Apr. URL: https://www.pressganey.com/about/news/press-ganey-introduces-new-telemedicine-solution-for-medical-practices [accessed 2020-05-07]
40. Cleveland Clinic Orthopaedic Arthroplasty Group. Press Ganey Administration of Hospital Consumer Assessment of Healthcare Providers and Systems Survey Result in a Biased Responder Sample for Hip and Knee Arthroplasties. J Arthroplasty 2019 Nov;34(11):2538-2543. [doi: 10.1016/j.arth.2019.06.045] [Medline: 31331703]
41. Hwang C, Lipman G, Kane M. Effect of an emergency department fast track on Press-Ganey patient satisfaction scores. West J Emerg Med 2015 Jan 01;16(1):34-38 [FREE Full text] [doi: 10.5811/westjem.2014.11.21768] [Medline: 25671005]
42. Chandra A, Sieck S, Hocker M, Gerardo CJ, Villani J, Harrison D, et al. An Observation Unit May Help Improve an Institution s Press Ganey Satisfaction Score. Crit Pathw Cardiol 2011;10(2):104-106. [doi: 10.1097/hpc.0b013e31821c5da8]
43. Siddiqui Z, Berry S, Bertram A, Allen L, Hoyer E, Durkin N, et al. Does Patient Experience Predict 30-Day Readmission? A Patient-Level Analysis of HCAHPS Data. J Hosp Med 2018 Sep;13(10):681-687. [doi: 10.12788/jhm.3037] [Medline: 30261085]
44. Newgard CD, Fu R, Heilman J, Tanski M, Ma OJ, Lines A, et al. Using Press Ganey Provider Feedback to Improve Patient Satisfaction: A Pilot Randomized Controlled Trial. Acad Emerg Med 2017 Sep 16;24(9):1051-1059 [FREE Full text] [doi: 10.1111/acem.13248] [Medline: 28662281]

Abbreviations

ICU: intensive care unit
NYP/WCM: NewYork-Presbyterian/Weill Cornell Medical Center
PE: parameter estimate

Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by J Portnoy, K Lawrence; comments to author 13.07.20; revised version received 23.07.20; accepted 06.08.20; published 09.09.20.

Please cite as:
Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, Hu JC
Patient Satisfaction With Telemedicine During the COVID-19 Pandemic: Retrospective Cohort Study
J Med Internet Res 2020;22(9):e20786
URL: http://www.jmir.org/2020/9/e20786/
doi:10.2196/20786
PMID:32810841

Short Paper

App-Based Tracking of Self-Reported COVID-19 Symptoms: Analysis of Questionnaire Data

Martin Zens[1], MD, PhD; Arne Brammertz[2], Dipl-Vw; Juliane Herpich[1], MD; Norbert Südkamp[3*], MD, PhD; Martin Hinterseer[1*], MD

[1]Department of Medicine, Kliniken Ostallgaeu-Kaufbeuren, Fuessen, Germany

[2]DESIGN-IT GmbH, Frankfurt, Germany

[3]University Medical Center Freiburg, Freiburg, Germany

[*]these authors contributed equally

Corresponding Author:
Martin Zens, MD, PhD
Department of Medicine
Kliniken Ostallgaeu-Kaufbeuren
Stadtbleiche 1
Fuessen, 87629
Germany
Phone: 49 8362 500 726
Email: martin.zens@me.com

Abstract

Background: COVID-19 is an infectious disease characterized by various clinical presentations. Knowledge of possible symptoms and their distribution allows for the early identification of infected patients.

Objective: To determine the distribution pattern of COVID-19 symptoms as well as possible unreported symptoms, we created an app-based self-reporting tool.

Methods: The COVID-19 Symptom Tracker is an app-based daily self-reporting tool. Between April 8 and May 15, 2020, a total of 22,327 individuals installed this app on their mobile device. An initial questionnaire asked for demographic information (age, gender, postal code) and past medical history comprising relevant chronic diseases. The participants were reminded daily to report whether they were experiencing any symptoms and if they had been tested for SARS-CoV-2 infection. Participants who sought health care services were asked additional questions regarding diagnostics and treatment. Participation was open to all adults (≥18 years). The study was completely anonymous.

Results: In total, 11,829 (52.98%) participants completed the symptom questionnaire at least once. Of these, 291 (2.46%) participants stated that they had undergone an RT-PCR (reverse transcription-polymerase chain reaction) test for SARS-CoV-2; 65 (0.55%) reported a positive test result and 226 (1.91%) a negative one. The mean number of reported symptoms among untested participants was 0.81 (SD 1.85). Participants with a positive test result had, on average, 5.63 symptoms (SD 2.82). The most significant risk factors were diabetes (odds ratio [OR] 8.95, 95% CI 3.30-22.37) and chronic heart disease (OR 2.85, 95% CI 1.43-5.69). We identified chills, fever, loss of smell, nausea and vomiting, and shortness of breath as the top five strongest predictors for a COVID-19 infection. The odds ratio for loss of smell was 3.13 (95% CI 1.76-5.58). Nausea and vomiting (OR 2.84, 95% CI 1.61-5.00) had been reported as an uncommon symptom previously; however, our data suggest a significant predictive value.

Conclusions: Self-reported symptom tracking helps to identify novel symptoms of COVID-19 and to estimate the predictive value of certain symptoms. This aids in the development of reliable screening tools. Clinical screening with a high pretest probability allows for the rapid identification of infections and the cost-effective use of testing resources. Based on our results, we suggest that loss of smell and taste be considered cardinal symptoms; we also stress that diabetes is a risk factor for a highly symptomatic course of COVID-19 infection.

(J Med Internet Res 2020;22(9):e21956) doi:10.2196/21956

KEYWORDS

COVID-19; self-reporting; symptom; tracking; app; surveillance; distribution; digital tool; screening

Introduction

COVID-19 was initially characterized as an acute respiratory infection with significant virulence and mortality. Following reports of the first cases in Wuhan, China, in December 2019, the virus spread globally with 9,738,374 confirmed cases as of June 26, 2020 [1]. Previous research has revealed that COVID-19, which is caused by the novel coronavirus SARS-CoV-2, is a systemic disease rather than an isolated acute respiratory illness [2,3]. The diversity in symptoms is the result of SARS-CoV-2 attacking various organs.

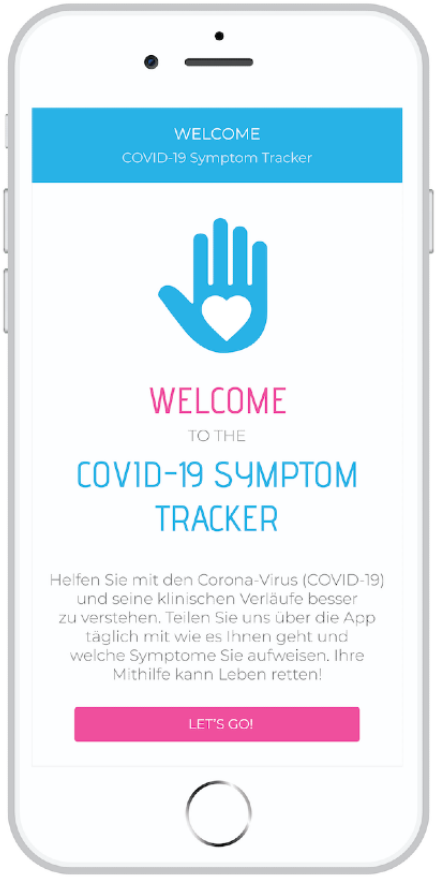
The aim of this study was to identify further symptoms and to investigate whether certain symptoms may be used as for screening to differentiate COVID-19 infection from other diseases. A powerful and reliable clinical screening tool may help to identify and isolate SARS-CoV-2–infected individuals and slow down disease prevalence.

Methods

Study Setting and Participants

The COVID-19 Symptom Tracker was developed by DESIGN-IT GmbH, in collaboration with the University Medical Center Freiburg and Kliniken Ostallgaeu-Kaufbeuren, Fuessen Hospital (Figure 1). The first version was released for Apple iOS on April 8, 2020, and for Google Android on April 20, 2020. Within 5 weeks, the app was downloaded 22,327 times. The app was released in German, English, French, and Spanish and advertised through public media (TV, radio, newspaper, online), following a nationwide press release. An initial questionnaire asked for demographic information (age, gender, postal code) and past medical history with relevant chronic diseases. The participants are notified daily by push notifications to report whether they were experiencing symptoms and if they had been tested for SARS-CoV-2. In addition to a binary response regarding the prevalence of symptoms, for some items (eg, fever, cough) we collected additional details like temperature range or expectoration. If users sought professional health care advice (eg, visited a hospital, a private practice, etc), additional questions regarding diagnostics and treatment were asked. Participation was open to all adults (≥18 years). The study was anonymous; a formal consent for participation was not required due to this fact but was asked for regardless.

Figure 1. The COVID-19 Symptom Tracker.



Statistical Analysis

Data were downloaded from the database server. Only data sets of individuals who live in Germany and had completed the entire symptom questionnaire at least once were included in further analysis. Participants were presented as counts, age as mean and standard deviation derived from an age span, and questionnaire items as counts and percentages. For each risk factor and symptom, odds ratios were calculated using binary

logistic regression. Data from untested participants were not included in this analysis.

Ethics

Approval of the study design was obtained by the Ethics Committee at the University Medical Center Freiburg (EK 337/20). The committee was very supportive of the project.

Data and Code Availability

The anonymous data collected by the COVID-19 Symptom Tracker app can be shared with researchers upon request with a research protocol or due to a question of public interest, if permitted by the ethics committee. The app code is also available upon request. Requests can be sent to the corresponding author. The Laravel PHP script for SQL data extraction is publicly available [4].

Results

Between April 8 and May 15, 2020, 22,327 individuals installed the COVID-19 Symptom Tracker on their mobile device. In total, 11,829 (52.98%) participants completed the symptom questionnaire at least once. Of these, 291 (2.46%) stated that they had undergone an RT-PCR (reverse transcription-polymerase chain reaction) test for SARS-CoV-2; 65 (0.55%) reported a positive test result and 226 (1.91%) a negative one. The mean number of reported symptoms in the group of untested participants was 0.81 (SD 1.85). On average, individuals with a negative result reported 4.26 symptoms (SD 2.52) and those with a positive result reported 5.63 symptoms (SD 2.82). All participants were asked to fill out an initial questionnaire on intrinsic risk and demographic data. The self-reported prevalence of risk factors (eg, diabetes: 917/11,538, 7.95%; hypertension: 3268/11,538, 25.19%; and smoking: 3679/11,538, 31.89%) showed a significant correlation with the data provided regularly by the German Federal Office of Statistics [5] and the Ministry of Health [6] (eg, diabetes: 7.2%; hypertension: 24.6%), which suggests that the study cohort is representative of the German population (age group: 18-65 years). Elderly individuals are underrepresented due to the methodology used in this study.

The daily symptom questionnaire requested information on known or suggested symptoms as of April 1, 2020 [7-10] and permitted users to enter additional symptoms as free text. Odds ratios (OR) were calculated for all risk factors and reported symptoms. According to our data, the most significant risk factors are diabetes (OR 8.95, 95% CI 3.30-22.37) and chronic heart disease (OR 2.85, 95% CI 1.43-5.69) (Figure 2). We identified chills, fever, loss of smell, nausea and vomiting, and shortness of breath as the top five strongest predictors for a COVID-19 infection (Figure 2). The OR for loss of smell was 3.13 (95% CI 1.76-5.58); of comparable significance were chills (OR 4.48, 95% CI 2.51-8.01) and fever (OR 4.37, 95% CI 2.44-7.81). Chills, fever, and shortness of breath have been identified and communicated as common symptoms in the mainstream media. Nausea and vomiting (OR 2.84, 95% CI 1.61-5.00) has been reported as an uncommon symptom; however, our data suggest a significant predictive value. The characteristics of the entire cohort are displayed in Table 1.

Figure 2. (a) Association between reported symptoms and the odds ratio for a positive SARS-CoV-2 test result; (b) association between reported risk factors and the odds ratio for a positive SARS-CoV-2 test result. Both analyses were carried out in a population of 291 participants who tested via reverse transcription-polymerase chain reaction. Error bars represent 95% CI.

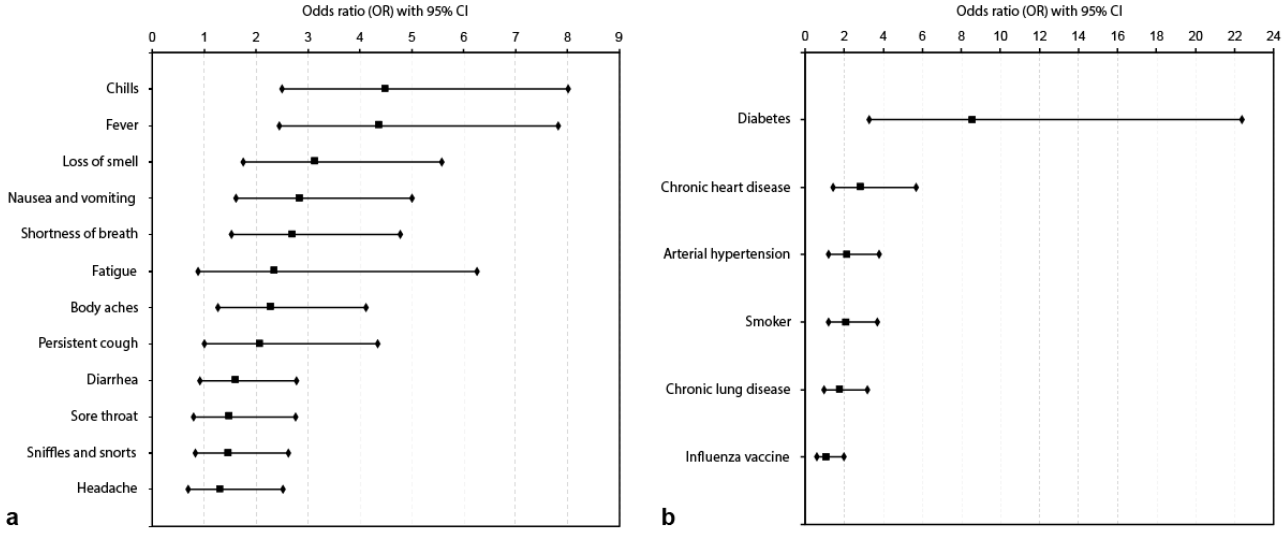


Table 1. Characteristics of participants (N=11,829).

Characteristic	Tested for SARS-CoV-2		Not tested for SARS-CoV-2
	Positive test result	Negative test result	
Participants, n (%)	65 (0.55)	226 (1.91)	11,538 (97.54)
Female, n (%)	33 (50.77)	109 (48.23)	4360 (37.79)
Age (years), mean (SD)	42.65 (13.33)	41.04 (12.88)	44.47 (15.41)
Smoker, n (%)	33 (50.77)	74 (32.74)	3679 (31.89)
Influenza vaccine, n (%)	21 (32.31)	69 (30.53)	3268 (28.32)
Comorbidities, n (%)			
Diabetes	14 (21.54)	7 (3.10)	917 (7.95)
Arterial hypertension	26 (40.00)	54 (23.89)	2906 (25.19)
Chronic lung disease	22 (33.85)	51 (22.57)	1352 (11.72)
Chronic heart disease	17 (26.15)	25 (11.06)	1020 (8.84)
Symptoms, n (%)			
Loss of smell	31 (47.69)	51 (22.57)	325 (2.82)
Fatigue	60 (92.31)	189 (83.63)	1560 (13.52)
Shortness of breath	40 (61.54)	84 (37.17)	554 (4.80)
Fever	36 (55.38)	50 (22.12)	296 (2.57)
Persistent cough	55 (84.62)	164 (72.57)	1290 (11.18)
Diarrhea	34 (52.31)	92 (40.71)	670 (5.81)
Chills	38 (58.46)	54 (23.89)	393 (3.41)
Headache	50 (76.92)	162 (71.68)	1388 (12.03)
Sniffling and snorting	43 (66.15)	129 (57.08)	1339 (11.61)
Nausea and vomiting	34 (52.31)	63 (27.88)	428 (3.71)
Body aches	45 (69.23)	112 (49.56)	1388 (12.03)
Sore throat	48 (73.85)	148 (65.49)	1185 (10.27)

Discussion

Principal Findings

The COVID-19 pandemic has triggered numerous research endeavors that have focused on procuring a better understanding of the disease caused by the novel SARS-CoV-2. Agile and dynamic projects led to contemporaneous approaches with similar and comparable study designs. Menni et al [11] report of real-time symptom tracking tool to predict COVID-19 infections. The approach is based on a smartphone app available in the United Kingdom and United States; 2,618,862 participants were included and analyzed in order to design a prediction model. Simultaneously, a total of 22,327 participants reported potential COVID-19 symptoms on a smartphone-based app in Germany. However, unlike the cohort discussed in this work, the British cohort did not represent the general population (eg, overrepresentation of female participants).

Menni et al [11] suggested that loss of smell and taste should be included in routine screening for COVID-19. We strongly agree with this opinion as our data also suggest a strong predictive value. Although chills and fever are found in COVID-19 patients with a higher probability, in a clinical setting a loss of smell and taste is unique and allows a better differentiation from other infectious diseases since many are accompanied by chills and fever. In the present situation, we suggest that loss of smell and taste, especially in combination with other symptoms, ought to be considered a red flag and should result in immediate testing for SARS-CoV-2 as well as isolation of the patient until the test result is obtained.

Furthermore, our work suggests that an increased awareness of gastrointestinal symptoms, namely nausea and vomiting, is needed as these have a stronger predictive value for a COVID-19 infection than symptoms such as sore throat or persistent cough, which are commonly considered as typical.

Apart from that, a significant association was seen between every symptom mentioned in this paper and a positive test result for SARS-CoV-2. This finding is explained by the preselection of symptoms that were reported to be associated with COVID-19 in prior publications.

Manual screening of the answers to the open question which asked for additional symptoms revealed an accumulation of the following symptoms: vertigo, painful ears or eyes, burning

sensation of the tongue, and thoracic pain. A static analysis of these reported symptoms is not yet possible due to insufficient data but corresponding questions have been added to the questionnaire in an updated version of the smartphone app.

Diabetes was identified as a major risk factor for a symptomatic course of COVID-19 infection. Other studies reported an increased risk, rapid progression, and a worse prognosis in patients with diabetes mellitus [12,13]. The mechanism remains unclear and requires further investigation.

Limitations

The limitations of this study are mainly due to the self-reporting nature of our methodology for data retrieval. The design does not allow for the verification of the reported symptoms or test results. Apart from that, the participants are not invited or preselected and may not represent the general population. The use of a smartphone device may have resulted in an underrepresentation of older adults. Another possible limitation is the small number of participants that had been tested for SARS-CoV-2 infection. A possible correlation between age, gender, and postal code (demographic information) and individual symptoms was not considered in the univariate analysis of symptoms.

Conclusions

Self-reported symptom tracking may help to identify novel symptoms of COVID-19 and estimate the predictive value of certain symptoms [14]. This may aid in the development of reliable screening tools. Clinical screening with a high pretest probability allows for the rapid identification of infections and serves as a cost-effective use of testing resources. Our data stress the necessity for an awareness of loss of smell and taste as cardinal symptoms.

Acknowledgments

The authors would like to express their sincere thanks to all the participants who voluntarily entered the study and shared information on their personal health and well-being. We thank the Ministerium fuer Wissenschaft, Forschung und Kunst, Baden-Wuerttemberg for an initial funding of the research platform, which was the foundation for the COVID-19 Symptom Tracker app. Furthermore, we would like to thank J Spieler, data protection officer at the University Medical Center Freiburg, for his support. The article processing fee was funded by the Baden-Wuerttemberg Ministry of Science, Research and Art and the University of Freiburg through the Open Access Publishing funding program in cooperation with a grant from Arbeitsgemeinschaft Leitende Kardiologische Krankenhausärzte eV.

Authors' Contributions

AB designed the app and database. MZ, JH, and MH developed the study questionnaires. MZ and NPS analyzed the data and plotted the results of the statistical analysis. MZ and JH wrote the manuscript. All authors revised the manuscript.

Conflicts of Interest

AB is the CEO of DESIGN-IT GmbH.

References

1. COVID-19 Coronavirus Pandemic. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1? [accessed 2020-06-26]
2. Zheng Y, Ma Y, Zhang J, Xie X. COVID-19 and the cardiovascular system. Nat Rev Cardiol 2020 May 5;17(5):259-260 [FREE Full text] [doi: 10.1038/s41569-020-0360-5] [Medline: 32139904]
3. Ronco C, Reis T. Kidney involvement in COVID-19 and rationale for extracorporeal therapies. Nat Rev Nephrol 2020 Jun 9;16(6):308-310 [FREE Full text] [doi: 10.1038/s41581-020-0284-7] [Medline: 32273593]
4. martinzens / covid-19-st. GitHub. 2020. URL: https://github.com/martinzens/covid-19-st [accessed 2020-08-26]
5. Statistisches Bundesamt. German Federal Office of Statistics. URL: https://www.destatis.de [accessed 2020-08-26]
6. COVID-19 in Deutschland. Robert Kock Institute. URL: https://www.rki.de/DE/Home/homepage_node.html [accessed 2020-08-26]
7. Lake MA. What we know so far: COVID-19 current clinical knowledge and research. Clin Med (Lond) 2020 Mar 05;20(2):124-127 [FREE Full text] [doi: 10.7861/clinmed.2019-coron] [Medline: 32139372]
8. Gane S, Kelly C, Hopkins C. Isolated Sudden Onset Anosmia in COVID-19 Infection. A Novel Syndrome? Rhin 2020 Jun 01;58(3):299-301. [doi: 10.4193/rhin20.114]
9. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). J Gen Intern Med 2020 May 4;35(5):1545-1549 [FREE Full text] [doi: 10.1007/s11606-020-05762-w] [Medline: 32133578]
10. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]

11. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
12. Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes Metab Res Rev 2020 Mar 31:e3319 [FREE Full text] [doi: 10.1002/dmrr.3319] [Medline: 32233013]
13. Muniyappa R, Gubbi S. COVID-19 pandemic, coronaviruses, and diabetes mellitus. Am J Physiol Endocrinol Metab 2020 May 01;318(5):E736-E741 [FREE Full text] [doi: 10.1152/ajpendo.00124.2020] [Medline: 32228322]
14. Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nat Med 2020 May;26(5):634-638 [FREE Full text] [doi: 10.1038/s41591-020-0857-9] [Medline: 32273611]

Abbreviations

OR: odds ratio
RT-PCR: reverse transcription-polymerase chain reaction

Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by I Rodriguez, P Innominato; comments to author 24.07.20; revised version received 25.07.20; accepted 26.07.20; published 09.09.20.

Please cite as:
Zens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M
App-Based Tracking of Self-Reported COVID-19 Symptoms: Analysis of Questionnaire Data
J Med Internet Res 2020;22(9):e21956
URL: http://www.jmir.org/2020/9/e21956/
doi:10.2196/21956
PMID:32791493

Original Paper

Assessing the Impact of the COVID-19 Pandemic in Spain: Large-Scale, Online, Self-Reported Population Survey

Nuria Oliver[1], PhD, ING; Xavier Barber[2], BSc, MSc, Prof Dr; Kirsten Roomp[3], BSc, PhD; Kristof Roomp[4]

[1]The Institute for Human(ity)-Centric Artificial Intelligence, ELLIS Unit Alicante Foundation, Alicante, Spain

[2]Center of Operations Research, Universidad Miguel Hernández, Elche, Spain

[3]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg

[4]Microsoft, Redmond, WA, United States

Corresponding Author:
Nuria Oliver, PhD, ING
The Institute for Human(ity)-Centric Artificial Intelligence
ELLIS Unit Alicante Foundation
Avenida Universidad
Alicante, 03690
Spain
Phone: 34 630726085
Email: nuria@alum.mit.edu

Abstract

Background: Spain has been one of the countries most impacted by the COVID-19 pandemic. Since the first confirmed case was reported on January 31, 2020, there have been over 405,000 cases and 28,000 deaths in Spain. The economic and social impact is without precedent. Thus, it is important to quickly assess the situation and perception of the population. Large-scale online surveys have been shown to be an effective tool for this purpose.

Objective: We aim to assess the situation and perception of the Spanish population in four key areas related to the COVID-19 pandemic: social contact behavior during confinement, personal economic impact, labor situation, and health status.

Methods: We obtained a large sample using an online survey with 24 questions related to COVID-19 in the week of March 28-April 2, 2020, during the peak of the first wave of COVID-19 in Spain. The self-selection online survey method of nonprobability sampling was used to recruit 156,614 participants via social media posts that targeted the general adult population (age >18 years). Given such a large sample, the 95% CI was ±0.843 for all reported proportions.

Results: Regarding social behavior during confinement, participants mainly left their homes to satisfy basic needs. We found several statistically significant differences in social behavior across genders and age groups. The population's willingness to comply with the confinement measures is evident. From the survey answers, we identified a significant adverse economic impact of the pandemic on those working in small businesses and a negative correlation between economic damage and willingness to stay in confinement. The survey revealed that close contacts play an important role in the transmission of the disease, and 28% of the participants lacked the necessary resources to properly isolate themselves. We also identified a significant lack of testing, with only 1% of the population tested and 6% of respondents unable to be tested despite their doctor's recommendation. We developed a generalized linear model to identify the variables that were correlated with a positive SARS-CoV-2 test result. Using this model, we estimated an average of 5% for SARS-CoV-2 prevalence in the Spanish population during the time of the study. A seroprevalence study carried out later by the Spanish Ministry of Health reported a similar level of disease prevalence (5%).

Conclusions: Large-scale online population surveys, distributed via social media and online messaging platforms, can be an effective, cheap, and fast tool to assess the impact and prevalence of an infectious disease in the context of a pandemic, particularly when there is a scarcity of official data and limited testing capacity.

(J Med Internet Res 2020;22(9):e21319) doi:10.2196/21319

KEYWORDS

COVID-19; SARS-CoV-2; public health authorities; large-scale online surveys; infectious disease; outbreak; public engagement; disease prevalence; impact; survey; spain; public health; perception

Introduction

Background

The first cases of COVID-19 were reported in Wuhan, China in December 2019. Since then, the SARS-CoV-2 virus has spread worldwide, infecting over 24 million people and causing over 825,000 deaths worldwide as of August 27, 2020 [1]. This virus has caused significantly more infections and deaths, compared with previous outbreaks of other coronaviruses causing severe acute respiratory syndrome and Middle East respiratory syndrome. The World Health Organization declared a global COVID-19 pandemic on March 11, 2020, and to date has been unable to predict the duration of the pandemic [2].

The first confirmed case of COVID-19 in Spain was reported on January 31, 2020, when a German tourist tested positive in the Spanish Canary Islands. However, this was an isolated imported case. It was not until February 24 when Spain confirmed several new COVID-19 cases related to a recent SARS-CoV-2 outbreak in the north of Italy. Since that date, the number of COVID-19 cases grew exponentially in Spain so that by March 30, 2020, there were over 85,199 confirmed cases, 16,780 recoveries, and the staggering figure of 7424 deaths, according to the official numbers. On March 25, 2020, the death toll attributed to COVID-19 in Spain surpassed that of mainland China, and it was only surpassed by the death toll in Italy. The economic and social impact of the COVID-19 pandemic in Spain is without precedent.

To combat the pandemic, the Spanish Government implemented a series of social distancing and mobility restriction measures. First, all classes at all educational levels were cancelled in the main hot spots of the disease on March 10, in the Basque Country and on March 11, 2020, in the Madrid and La Rioja regions. All direct flights from Italy to Spain were cancelled on March 10. On March 12, the Catalan Government quarantined four municipalities that were particularly affected by the virus. On March 13, the Government of Spain declared a state of emergency for 2 weeks across the entire country. Since the state of emergency was established, all schools and university classes were cancelled, large-scale events and nonessential travel were forbidden, and workers were encouraged to tele-work. Despite these efforts, the daily growth rate in the number of confirmed COVID-19 cases continued to grow. Thus, on March 30, new mobility restriction and social distancing measures were implemented; all nonessential labor activity was to be interrupted for a 2-week period. Moreover, the Spanish Government extended the state of emergency first until April 11 and then renewed on a biweekly basis until June 21. Although these interventions put a halt to the normal daily lives of most people in Spain, their impact on people's economic, physical, and mental well-being were unknown at the time, as was the actual prevalence of the disease.

Given the growth rate in the number of confirmed COVID-19 cases, rapid assessments of the population's situation and perceptions of the infection are of paramount importance. Traditional methods, such as population-representative household surveys are slow to design and deploy [3]. Phone surveys are generally faster to conduct, yet they are labor intensive and often yield low response rates (as low as 10% or less [4]). Moreover, the resulting sample might be biased and difficult to reweight [5]. Given the limitations of these traditional methods and given the need for rapid data collection, large-scale online surveys can be a valuable method to quickly assess and longitudinally monitor the situation and perceptions of the population in the context of a pandemic [6]. Thus, to shed light on important, yet unknown, questions related to COVID-19, we designed a 24-question online survey, called the *Covid19Impact* survey, to be targeted to the Spanish population. The survey became viral 12 hours after its publication, yielding over 140,000 answers. It is one of the largest surveys in the world carried out in the context of the COVID-19 pandemic [7].

Population Surveys During the COVID-19 Pandemic

Other efforts to collect data from the population regarding the COVID-19 pandemic have been deployed in multiple countries. The largest study to date involved the *Methods* smartphone app, with 2,618,862 participants who self-reported symptoms in the United States and the United Kingdom [8]. The study asked questions focused on risk factors and symptoms, and described a predictive model of COVID-19 based on these variables. In Canada, *FLATTEN* [9] has gathered data from respondents and asks simple health and demographic-related questions to help monitor the spread of the virus in an anonymous manner. The *International Survey on Coronavirus* asks questions focusing on the psychological impact of the crisis [10]. There were three main findings from the analysis of this survey's answers: many respondents found that both the population and their governments' response to the COVID-19 pandemic was insufficient, this insufficient response was associated with lower mental well-being, and a strong government response was associated with an improvement in respondents' views of other people and their government together with better mental well-being. The *COVID-19:CH Survey* in Switzerland aims to collect personal data related to COVID-19 testing with additional health- and potential exposure–related information [11]. The data collected is presented to the public in a visual format, giving information on, among other things, demographics, comorbidities, and symptoms. In Israel, the Weizmann Institute and the Ministry of Health are collecting data on basic demographics, health, and potential exposure [12]. The project aims to predict the location of COVID-19 outbreaks by analyzing information collected about the virus symptoms and public behavior in real time [6,13,14].

Numerous efforts with smaller numbers of respondents have also taken place or are ongoing. In China, an early study was conducted between January 27 and February 1, 2020, which relied on the Chinese social media and traditional media outlets asking about knowledge, attitudes, and practices toward COVID-19 [15]. Among its many findings, the authors reported that most respondents felt that China could win the battle against the virus. An early international project was run from February 23 to March 2, 2020, collected data from the United Kingdom and the United States using an online platform managed by Prolific Academic Ltd, and asked about knowledge and perceptions of COVID-19 [16,17]. The survey provided potential information to guide public health. In mid-March and over 48

hours, responses were collected in the United States; the survey had been posted on 3 social media platforms (Twitter, Facebook, and Nextdoor) and collected data on symptoms, concerns, and individual actions [18]. They showed that 95.7% of respondents made lifestyle changes, including handwashing, avoiding social gatherings, social distancing, etc. In the United Kingdom, data was collected attempting to identify sociodemographic adoption of social-distancing measures, ability to work from home, and both the willingness and ability to self-isolate [19], providing potential information to policy makers. An online survey (*FEEL-COVID*) used the snowball sampling method to collect data in India and found that almost one-third of respondents were negatively psychologically impacted by the pandemic [20].

Our work complements these previous related efforts by focusing on Spain (one of the most affected countries by the COVID-19 pandemic) and by addressing four areas of people's experiences during the confinement: their social contact behavior, economic impact, labor situation, and health status.

This Study

Despite the availability of data regarding the number of confirmed COVID-19 cases, hospitalized and intensive care patients, and deaths in the early stages of the COVID-19 pandemic, there was a scarcity of high-quality data about important questions related to the population's experience.

First, there is the issue of underreporting confirmed cases and COVID-19–related deaths. Work by the Imperial College COVID-19 Response Team [21] estimated that 15% of the Spanish population could be infected by SARS-CoV-2. However, this figure was estimated to be much lower at around 5.3% by the preliminary results of a seroprevalence study carried out by the Spanish Ministry of Health [22,23]. Assessing the percentage of infected individuals is of utmost importance to build accurate epidemiological models and to assist policy makers in their decisions.

Second, there are unknowns regarding the sources of infection. Are people being infected by friends, family members, relatives, and coworkers, or are they being infected when shopping in supermarkets or at the bakery? The effectiveness of different government interventions will depend on the answers to these questions.

Third, the economic impact that the COVID-19 crisis will have on people's lives is yet to be quantified. According to the latest figures from the Spanish Industry, Commerce and Tourism Ministry, only 0.2% of Spanish companies have 250 or more employees, 44.6% of companies are micro (1-9 employees) or small (10-49 employees), and 54.4% of companies consist of the self-employed [24]. Small businesses are generally unprepared to confront such a crisis. Moreover, tourism represents 14.6% of Spanish gross domestic product (GDP) and 2.8 million jobs, and these are threatened by the COVID-19 pandemic [24]. Measuring the impact that the pandemic is having on people's finances is of great value to policy makers. Finally, there is the personal experience related to having to be confined in the home for weeks. How much longer will the population be able to sustain this situation?

In this paper, we describe the *Covid19Impact* survey, which was designed to answer these questions. We present the methodology that we followed to gather a large-scale sample via an online survey, followed by the analysis of the resulting answers and the main insights derived from them. Finally, we describe our conclusions and lines of future work.

Methods

Sampling and Data Collection

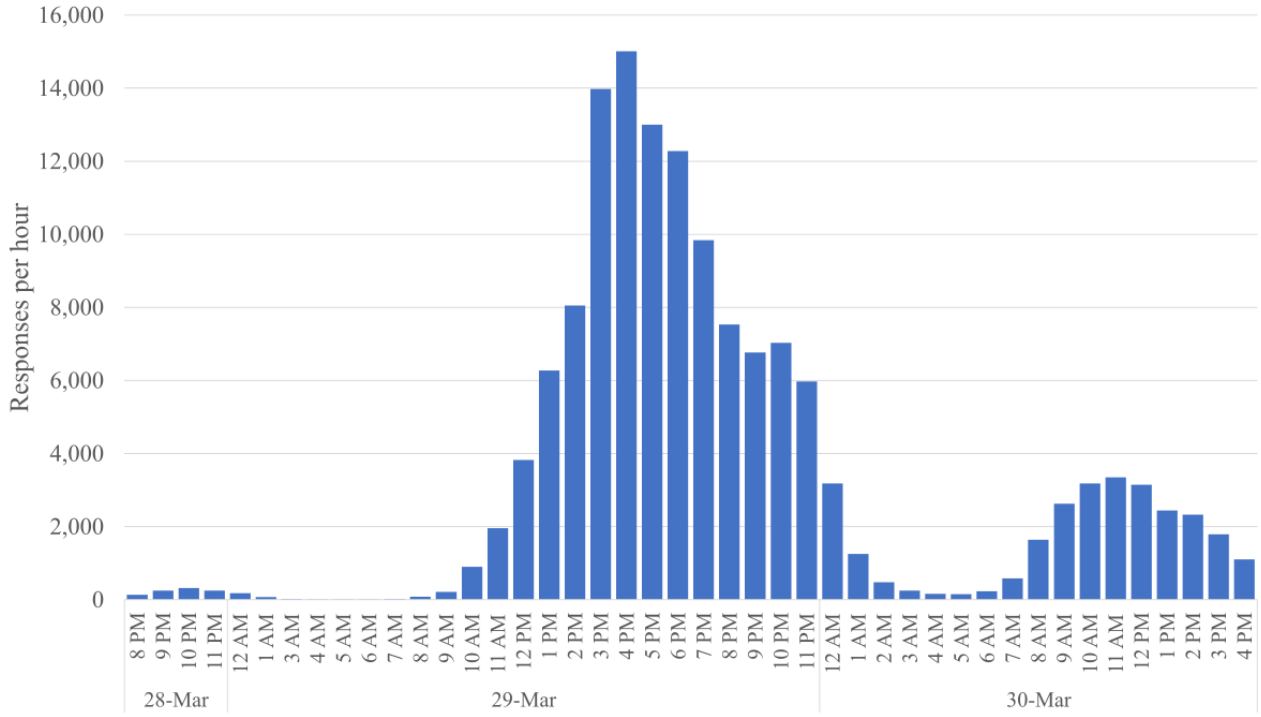
To answer the previously formulated questions, we designed a 24-question anonymous online survey that we refer to as the *Covid19Impact* survey (Multimedia Appendix 1). The survey is divided in 4 sections that address four different dimensions related to the population's experience during the COVID-19 crisis: their social contact in the last 2 weeks, the economic impact of the pandemic, their workplace and labor situation, and their health status. Moreover, the survey collects basic demographic (age range, gender, postal code) and home (type of home and number and ages of people in the home) data.

We used the self-selection online survey method of nonprobability sampling to recruit participants via social network posts (mainly Twitter and WhatsApp), asking the Spanish population (18 years or older) to answer the survey. This sampling method is particularly suitable during a confinement situation where the mobility and social contact of the population is greatly reduced. Thus, the online distribution of the survey enabled fast access to it by large numbers of people.

In addition to distributing the survey on Twitter and WhatsApp, we used snowball sampling [25]. The goal was to collect as large of a sample as possible in a short amount of time, as the COVID-19 situation was rapidly evolving, and new government measures might be required. The objective is to gather a snapshot of people's experiences regarding the four sections previously described.

Anticipating the start of new mobility restriction and social distancing measures on Monday, March 30, 2020, we deployed the survey on Saturday, March 28 at 8 PM. Via social media (Twitter and WhatsApp) and snowball sampling, we distributed the survey to a wide set of highly connected users who, in turn, distributed it to their contacts. The survey was also distributed by professional organizations, town halls, civil groups, and associations. In the 12 hours that followed, the survey went viral in Spain, and by the afternoon of Monday, March 30, we had collected over 140,000 answers. Figure 1 illustrates the growth in the number of answers over time, and the peak was reached in the time frame between 4 PM and 5 PM on Saturday, March 29, with more than 15,000 answers in 1 hour.

Figure 1. Number of answers collected by the Covid19Impact survey in its first two days, reported in one hour intervals.



The initial version of the survey was delivered via Google Forms, which allowed us to write and deploy the survey in an anonymous, scalable, and free manner within hours. The URL to the Google Forms was shared via bit.ly, such that we could estimate how many times the link had been shared. After reaching 140,000 answers, we began to hit scale limitations in Google Forms, so on March 30, 2020, we moved the survey to Survey123 [26] for future editions of the data collection.

Questionnaire Structure

All questions were anonymized to preserve privacy and no personal information was collected. In addition, the snowball sampling methodology enabled the anonymous distribution of the survey. The survey can be found online [27].

First, the survey obtained explicit consent from the users. Only when consent was granted and respondents confirmed they were adults could respondents continue to the rest of the questions.

The first section (question Q1-Q4) gathers basic demographics: country, age range, gender, and postal code. Next, there are 3 questions (Q5-Q7) related to the home situation: type of home, number of people in the home, and their ages. The following 7 questions (Q8-Q14) address the social contact behavior of the respondents during the last 2 weeks. This is an important section of the survey as we aim to understand the level of social interaction that people had despite the confinement and social distancing measures. The questions asked about having had contact with infected individuals, whether children were taken care of outside the home, if they had an external person coming to their house (eg, house cleaner), for what types of activities had they left their home, and what transportation means had they used. The last two questions intend to capture people's perceptions of the confinement measures: if they thought the measures were enough to contain the pandemic and for how long they would be able to tolerate the containment situation.

Personal economic impact is assessed with questions Q15 and Q16, followed by three questions (Q17-Q19) related to their workplace situation. Finally, the last 5 questions (Q20-Q24) address their health state to assess how many people might be infected by the virus, determine the ability of participants to self-isolate, and collect feedback regarding testing availability and testing results.

None of the questions, except for the consent question, were compulsory, and all the health-related questions included "I prefer not to answer" as a choice.

Credibility and Validity

Before widely deploying the survey, we carried out a pilot study to validate its content and proper anonymization with a small sample of participants. The questions were written in Spanish and English. Once all the bugs were fixed and minor feedback about the wording of the questions was addressed, we proceeded to widely deploy the survey.

Ethical Approval of the Research Protocol and Instruments

Before its deployment, the research protocol and instrument were reviewed and approved by the cabinet of the President of the Valencian Region of Spain. The findings of this survey have been regularly used and shared by the Valencian Government to assist their policy making during the COVID-19 pandemic [28].

Data Exclusion, Cleansing, and Reweighting

From a total of 156,614 answers, we eliminated all answers with blank or invalid postal codes. Moreover, we only analyzed responses with nonblank answers related to age, gender, province, and profession (including those who reported not working), yielding a final data set of 141,865 answers.

Thus, we report the results of analyzing these 141,865 answers collected between 8 PM GMT of March 28 and 11:59 PM GMT on April 2, 2020. With such a large sample, this survey is one of the largest population surveys on COVID-19 and the largest in Spain published to date [7].

All questions were binary or categorical. Thus, we report the percentage of participants who selected each response. Because our gender, age, geographic location, and profession distributions were not proportional to those of the general population of Spain, we computed a weighting factor, such that the resulting sample had similar demographic, geographic, and profession distributions as those of Spain, reported by the Spanish National Institute of Statistics (INE). To reduce biases, we used the reweighted data for all statistical inferences. The user and home situation statistics presented in the next section correspond to the raw data without reweighting. However, the rest of the sections regarding the statistical analysis of questions Q8 to Q24 correspond to analyzing the reweighted data. Multimedia Appendix 2 contains both the raw answers and the reweighted values of the univariate tables for each of the questions. Multimedia Appendix 3 contains the 141,865 responses as a text file, with zip code and time stamp information removed to protect the anonymity of the participants.

Statistical Analyses

The sampling error, after reweighting the samples, was 0.43. This small sampling error, due to the large sample, yields a narrow 95% CI of ±0.8428 for all proportions reported.

We use the Z test to compare two proportions, considering that the data comes from a survey and as such, the variance of each proportion is different to that of an infinite population test. We use a chi-square test to compare the independence between two questions [29]. Differences between answers greater than 0.85 were statistically significant with $P<.001$.

We measured the association between nominal variables using Cramér's V for RxC tables and Pearson phi for 2x2 tables [30]. We used weighted logistic regression to compute the odds ratio for a multivariate model using a quasi-binomial distribution family [31].

Results

User Statistics and Home Situation (Q1-Q7)

Geographically, most respondents were from the Valencian Region (102,021/141,865, 71.9%). However, there were also many answers from other regions of Spain including 10,365 answers from Madrid and 5691 from Catalonia, as shown in Table 1. Multimedia Appendix 2 contains the univariate tables corresponding to all the questions in the survey, including information about the participants' type of home and the number of people and ages of those living in their home.

Table 1. Age, gender, and geographical distribution of survey respondents (raw, unweighted data).

Sex, Autonomous community	Age ranges (years), n								Total (n=141,865)
	18-20 (n=3324)	21-29 (n=14,128)	30-39 (n=25,719)	40-49 (n=38,726)	50-59 (n=34,762)	60-69 (n=19,551)	70-79 (n=5093)	≥80 (n=562)	
Female									
Valencia	1687	6455	11,433	17,002	15,109	7697	1681	193	61,257
Madrid	75	563	1235	1739	1482	702	193	26	6015
Andalucía	105	411	700	883	746	335	61	4	3245
Catalonia	48	344	593	757	726	437	107	15	3027
Rest of Spain	322	1346	2474	3157	2512	1196	238	30	11,275
Total	2237	9119	16,435	23,538	20,575	10,367	2280	268	84,819
Male									
Valencia	873	3571	6453	10,620	10,181	6793	2064	209	40,764
Madrid	42	355	774	1237	1097	597	230	18	4350
Andalucía	27	209	397	667	617	410	109	10	2446
Catalonia	18	184	350	552	513	305	121	18	2061
Rest of Spain	127	690	1310	2112	1779	1079	289	39	7425
Total	1087	5009	9284	15,188	14,187	9184	2813	294	57,046

Given the gender, age and location biases in the raw data, we reweighted the data to match the distribution of the Spanish population according to the latest census [32], as reflected in Multimedia Appendix 2.

Almost all of the 141,865 respondents (n=141,807, 98.8%) lived in an apartment (n=93,060, 65.6%) or a single-family home (n=46,975, 33.1%). Most of the participants lived in a home

with 2 (n=42,513, 30.0%), 3 (n=36,879, 26.0%), or 4 (n=38,265, 27.0%) people, which is consistent with Spain's demography.

The rest of the reported statistics in this paper correspond to analyzing the reweighted sample to match in gender, age, province, and profession the distribution in Spain according to the latest data published by the Spanish INE.

Given that COVID-19's fatality rates are largest for older adults [33], we analyzed the age distribution of the homes with older adults: 11.8% of respondents older than 50 years lived with an older adult (age>60 years) and 19.9% of respondents lived in homes inhabited only by older adults. Intergenerational homes are particularly important for the transmission of SARS-CoV-2 [34].

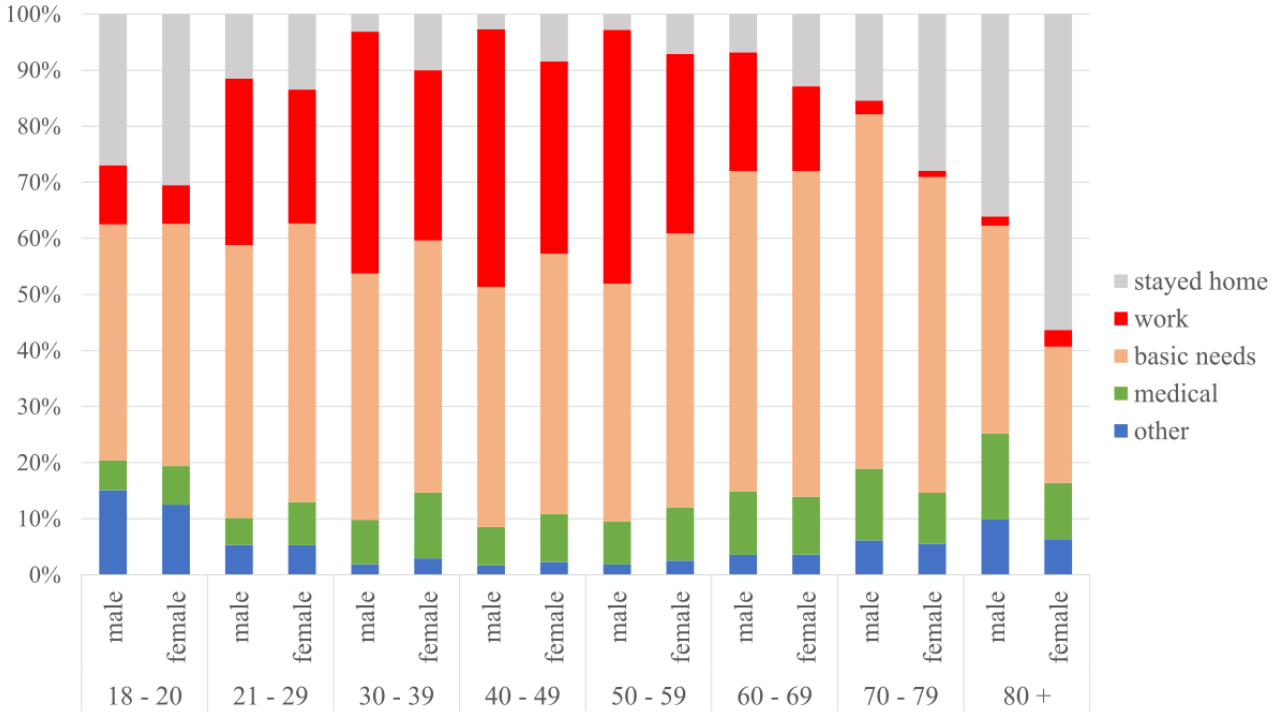
Social Contact Behavior (Q8-Q14)

With respect to social contact behavior with individuals with a confirmed SARS-CoV-2 infection (Q8), 17.3% of respondents reported having had close contact with a person who was infected with COVID-19 (n=140,008). The most common social context was a coworker (6.2%), a household member (6.1%), or a friend or relative (5.4%). In the case of having been in close contact with a confirmed infected person who was a patient of the participant, a gender-centric analysis revealed a significant (P<.001) difference between male and female respondents: 60.7% of the respondents were female vs 39.3% male. This large difference is partially due to the larger percentage of women (72.5%) who work in the health care sector vs men (27.5%) in Spain [32].

When asked if an outside person regularly visited the home (Q10), we identified a significant difference (P<.001) between older adults (age >70 years) and younger respondents (n=141,365): 21.2% of older respondents regularly had a person coming to their home versus only 13.6% in the case of younger adults (age <60 years). This is an important finding as special measures might need to be taken to protect the 21.2% of older adults who regularly receive external people in their homes.

Respondents (n=140,686) left their homes during the social distancing period for a variety of purposes (Q11), as shown in Figure 2: covering basic needs (supermarkets, bakery, and pharmacy) was the most common reason, reported by 47.8% of respondents, followed by going to work (31.3% of respondents). We identified statistically significant differences (P<.001) regarding age and gender. Older respondents (age >60 years) were more likely than younger participants (age <60 years) to stay entirely at home (14.9% older vs 7.6% for younger), and to leave their home to go to the pharmacy (11.5% vs 10.8%) and newspaper stand (9.7% vs 3.9%).

Figure 2. Reasons for leaving the home by gender and age.



Conversely, younger respondents (age<60 years) were more likely to leave their home to help others than older respondents (age >60 years; 81.0% vs 71.8%). Interestingly, the youngest respondents (aged 18-29 years, n=17,416) were also more likely to stay entirely at home versus respondents 30 years or older (23.1% vs 8.2%).

Regarding gender, among all female respondents, 14.8% reported not leaving the home versus 6.5% among male respondents. This difference was statistically significant (P<.001). The same pattern is found with respect to leaving the home to go to work, where 26.0% of all female participants versus 36.7% of all male respondents selected this option.

The main means of transportation (Q12) used by respondents was individual (84.5%; by foot, individual car, motorcycle, scooter) versus shared (5.9%; public transport, shared car, taxi). In this question, we observed the same gender patterns as in Q11 (n=140,308): among female respondents, 13.0% reported not leaving the home versus 6.2% among male respondents.

The last two questions in this section (Q13 and Q14) concerned the personal experience of respondents regarding the containment measures; 50.4% of participants (n=141,481) believed that the government should implement more measures to contain the pandemic, and only 2.2% thought that the measures were too severe. There was a significant difference (P<.001) in the support of the measures by age group. Despite being at a lower risk of death, 50.0% of younger people (age <60 years) believed measures should be stronger versus 37.1% of older people (age >59 years).

Q14 (n=138,155) explored how sustainable participants consider the social distancing measures to be. The most popular answer by respondents was that they could continue in this confined state for 1 additional month (44.1%), and a nonnegligible 32.4% reported being able to continue in confinement for 3-6 months. An interesting gender difference was found for those who responded that they could stay in confinement for 6 months: among female participants, 8.0% reported this to be the case versus 12.9% among male participants (P<.001). This might be due to the fact that women in Spain see their workload increased during the weeks of social distancing and mobility restriction, as reported in [35].

Personal Economic Impact and Workplace Situation (Q15-Q19)

An inevitable consequence of the COVID-19 pandemic is its economic and labor impact. Spain is a country with mostly small businesses, many of which are family owned. Q15-Q19 aim to shed light on the individual experiences and fears of people regarding their financial and employment situation.

When asked about the economic impact that the COVID-19 crisis is having on respondents' lives (Q15, n=139,008), 43.0% felt that the crisis had not yet significantly affected them economically. Moreover, 29.1% reported that their employer or company was undergoing financial problems, and 7.7% reported having lost a significant part of their savings or their job. These results need to be taken with great caution, as the sample was collected at the end of March and early April, when the devastating economic impact of the pandemic was not yet evident.

Among the respondents who had worked in the last month, there were significant differences in the distribution of work activities, as shown in Table 2. The most affected professions included hospitality and construction. The least affected were education and public administration.

Table 2. Distribution of jobs between respondents who had or were in danger of losing their job/business vs those who were not (Cramér's V=0.252).

Job categories[a]	Lost job or business, %[b]	Not lost job or business, %[b]
Administrative services	6.0	7.8
Retail	9.3	5.3
Communications	1.2	1.8
Construction	18.7	7.3
Domestic services	0.2	1.5
Education	3.1	14.2
Entertainment/arts	0.8	0.7
Essential services	2.2	9.7
Finance	1.3	4.8
Food production	3.2	3.8
Health and social services	1.8	2.8
Hospitality	29.3	13.8
Manufacturing	6.2	4.2
Other	8.8	7.4
Professional/technical/science	1.5	2.2
Public administration	0.3	5.3
Sanitation	2.7	4.4
Transportation	3.2	2.8

[a]The job categories are defined by the Spanish labor department (for the survey we only included categories with more than 1% representation in the population).

[b]Percentages are based off the weighted sample.

Small businesses have so far borne the brunt of the economic impact. For respondents (n=24,386) working in larger companies (≥100 employees), 80.1% reported that they had not yet been significantly affected versus only 42.7% of workers (n=39,052) at the smallest companies (1-9 workers) being unaffected. Among those working in small companies, 19.4% reported their companies were facing bankruptcy even at this early stage of the pandemic.

Again, there is a gender-based statistically significant difference (P<.001, n=139,008). In terms of having lost their jobs or savings, this option was selected by 8.3% among female participants versus 5.9% among male respondents.

With respect to the labor situation of our respondents (Q16, n=141,865), the majority (71.2%) reported working in the last month. A small fraction (5.9%) of respondents were students.

Q17 (n=98,740) focused on whether respondents had gone to work in the last week. The answers were split between the three available options: 38.3% did not go to work, 28.7% tele-worked, and 33.0% went to work.

Statistically significant gender differences (P<.001, n=98,740) were observed regarding working participants who did not go to work (42.0% among female participants vs 34.9% among male participants) and those who did go to work (29.1% among female participants vs 36.6% among male participants). No significant gender difference was found for those who tele-worked (28.9% among female participants vs 28.5% among male participants). In sum, female workers were significantly more likely to stay home than male workers.

Moreover, we found that the economic impact was a key factor in determining how much longer participants believed that they could continue in confinement. To explore the relationship between economic impact, age, and the willingness to stay in confinement, we built a multivariate weighted logistic regression model with *willingness and ability to stay in confinement* as a dependent variable (answers from Q14, divided into two values: 0, corresponding to answering that "at most I could continue in confinement for one week," and 1, corresponding to answering that "I could continue in confinement for longer than one week"). As covariate variables, we used sex, age, and the answers to question Q15 (economic impact). The logistic regression model revealed a clear impact of severe economic damage on willingness to stay in confinement; those who reported not having enough money to buy food had on average more than twice the probability of reporting not willing to continue in confinement for longer than 1 week (OR 2.23, 95% CI 1.81-2.77), and those who report being unable to pay their mortgage were on average 1.54 times more likely to also report not willing to continue in confinement for longer than 1 week (OR 1.54, 95% CI 1.29-1.83). Age also mattered; according to the model, respondents younger than 21 years had on average over twice the probability to report not willing to continue in confinement for longer than 2 weeks than those 21 years and older (OR 2.06, 95% CI 1.73-2.45).

Health State (Q20-Q24)

In the last section, Q20-Q24 asked respondents about their health. Regarding risk factors (Q20, n=135,583), we obtained a similar split between those who reported having at least one risk factor (48.3%) versus none of the listed risk factors (46.9%). In addition, 4.9% of respondents were health care workers. The risk factors that we asked participants about were hypertension, diabetes, cardiovascular disease, respiratory illness, immunosuppression, cancer, smoker (former), smoker (current), pregnancy, and health care worker.

Q21 (n=141,313) aimed to evaluate the ability of respondents to isolate themselves were they to be diagnosed with COVID-19. This is an important question given the relevance of implementing effective quarantine measures. Whereas 72.3% of respondents reported having the ability to properly isolate themselves, a nonnegligible 27.7% of respondents acknowledged not having the necessary resources to implement a proper quarantine.

In terms of age, 34.9% of respondents younger than 50 years reported not having the appropriate quarantine resources versus 21.0% of those older than 50 years. This might be due to the presence of other adults or children in the home. Indeed, 96.7% of respondents living alone (n=13,820) reported being able to self-isolate versus 68.6% of those living with other people (n=127,493), and only 5.7% of respondents younger than 50 years reported living alone when compared to 18.3% of adults 50 years and older. Moreover, when we looked at the impact of having children in the home, we observed that 41.1% of adults with children in the home (n=28,139) responded not being able to properly isolate versus 28.0% of adults without children in the home (n=67,659, P<.001). Among those living with older adults (n=15,124), 10.8% reported not having appropriate quarantine infrastructure at home.

To shed light on the percentage of the population that might currently be infected by SARS-CoV-2, Q22 asked respondents if they currently had any of the following symptoms that were unusual for them: difficulty breathing, dry cough, fever, headache, productive cough, anosmia, muscle pain, and sore throat; 16.8% of respondents (n=136,386) responded having at least one of the symptoms. Regarding gender, a larger percentage of women (19.0%) versus men (14.5%) reported having symptoms. This difference is statistically significant (P<.001). The age group who most reported having symptoms was the 30-39 years age group (n=24,839, 20.9%).

Finally, when asked whether respondents had been tested for COVID-19 (n=138,023), 87.4% felt they did not need to be tested; 6.1% were told by their doctor they should be tested, but no tests were available; 0.7% had tested negative; 0.3% had tested positive; and 0.2% were waiting for their outcomes, resulting in an overall test rate of 1.2%. We found statistically significant (P<.001) differences between those who exhibited any of the three symptoms (difficulty breathing, dry cough, and fever) and those who did not, and their answers regarding testing: 93.1% of those who did not have symptoms considered testing not necessary versus only 58.1% for those who had such symptoms. Table 3 depicts the responses from these two groups.

Table 3. Testing needs, depending on the presence of symptoms.[a]

Testing	Difficulty breathing, dry cough, or fever, %[b]	Other or no symptoms, %[b]
Negative	2.3	0.6
No need	58.1	93.1
No test available	32.5	4.9
Positive	3.9	0.2
Waiting for results	1.2	0.1
No, but need one due to being caretaker of person at risk	2.1	1.1

[a]All differences between the symptoms/no symptoms groups were statistically significant (P<.001), and Cramér's V=0.327.

[b]Percentages are based off the weighted sample.

When looking at Q8 (whether respondents had close contact with an infected individual) together with Q23 (whether they had been tested for coronavirus and the results of the test), we identified an interesting pattern. Among those who had tested positive and answered Q8 (n=414), 80.9% had close contact with a known infected individual; of these, 32.4% had been through a member of the household, friend, or relative, 26.6% through a patient (health care workers), 11.1% at work, and only 1.7% through a client. Thus, over 80% of respondents with COVID-19 knew their likely source of infection. This finding is partly explained by the fact that the survey was answered during a period of confinement with reduced mobility and social contact.

Finally, we observed in the data a nonlinear relationship between testing positive and age, gender, and the ability to self-isolate (Q21). Thus, we carried out a multivariate weighted logistic regression analysis to study the relationship between these variables and found a three-way interaction between them. Females 70 years or older who reported not being able to properly isolate had on average almost twice the probability of testing positive than otherwise (OR 1.91, 95% CI 1.18-3.073).

Prevalence

One of the goals of the survey was to make a rapid estimate of the COVID-19 prevalence in the Spanish population. The rapid rise in deaths from mid-March onward made it clear that there were far more infections than what the official case numbers reported. Model-based prevalence estimates, such as the report from Imperial College on March 30, 2020, estimated that 15% of the population had been infected to that date in Spain [21]. Large-scale, online surveys have been shown to be a useful tool to quickly estimate the percentage of currently infected individuals in the population and identify risk factors to help design measures to contain the epidemic [8,13,14]. The goal is not to replace the golden standard of seroprevalence surveys but to assess the value of a cheap and fast large-scale online survey to infer COVID-19 prevalence at a time when there was data scarcity and limited testing capacity.

Q24 asked respondents about their coronavirus test result, which provided the ground truth for building a model to predict prevalence. Among respondents reporting confirmed test results, 426 of 1345 reported testing positive, which after reweighing would translate to 235,000 positive tests in Spain's population of 47 million. The official number of reported positive cases on April 2, 2020, was 181,859. This difference of 30% could be due to several factors, including selection bias, underreporting, or delays in the release of official test statistics [36-38].

Since at the time it was not clear which symptoms or other factors were most indicative of COVID-19 (anosmia was reported as an important symptom by March 20 [39]), we created a generalized linear model to infer the likelihood of testing positive for SARS-CoV-2 from the survey answers. In addition to symptoms, we included as independent variables gender, age, and the presence of a person with a positive coronavirus test result in the home. Our target variable was given by Q24; we selected those answers corresponding to participants who reported having tested positive (coded as 1) and those who reported having tested negative (coded as 0) for coronavirus (n=1345).

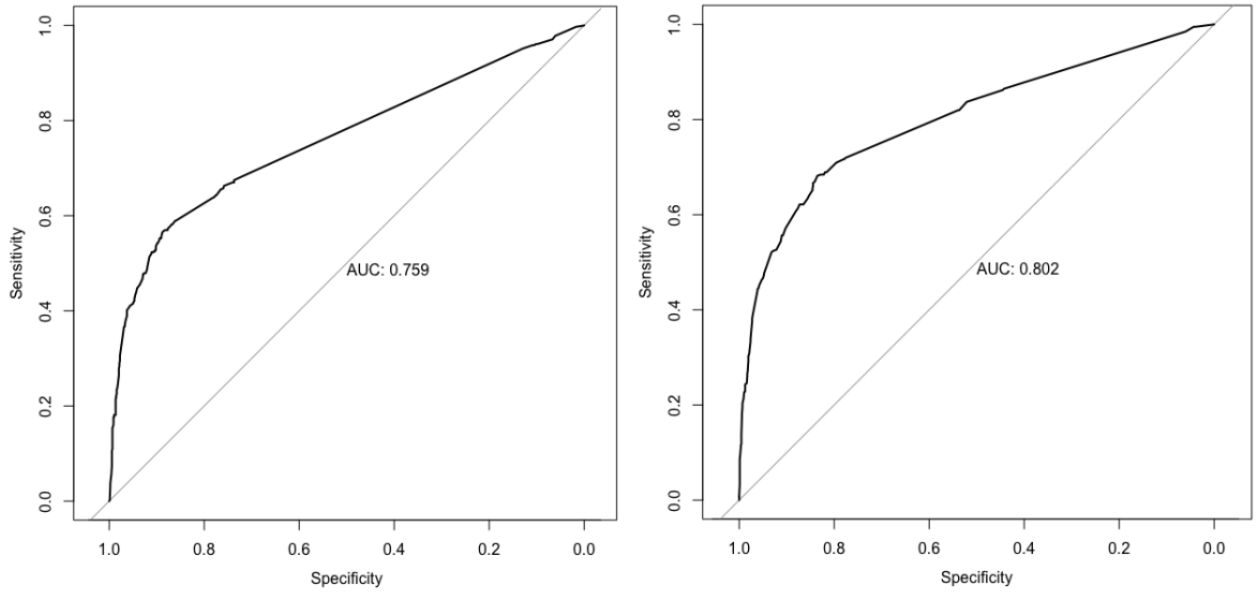
Moreover, we performed feature selection to select the features that yielded the best performing model. The selected independent variables are depicted in Table 4: a subset of the reported symptoms (Q22), whether the household already had an infected member (Q8), gender, and whether the participant's age was older than 70 years.

Table 4. Selected variables and coefficients of the generalized linear model.

Variable	Estimate	SE	T value	*P* value
(Intercept)	0.12510	0.01636	7.647	<.001
Member of home infected	0.28555	0.03137	9.103	<.001
Fever	0.18569	0.03724	4.986	<.001
Dry cough	0.05834	0.02808	2.078	.04
Productive cough	–0.09785	0.03676	–2.662	.008
Muscle pain	0.07208	0.03603	2.000	.046
Loss of sense of smell	0.45410	0.03409	13.319	<.001
Age >70 years	0.17487	0.06785	2.577	.01
Male	0.08038	0.02306	3.485	<.001

The obtained values from the generalized linear model were converted into probabilities by means of the logistic function (exp(x)/(1+exp(x)). The variables and parameters of the model are shown in Table 4. The final model had a sensitivity of 0.77 and a specificity of 0.80. Figure 3 (right) shows the receiver operating characteristic of this model. Although we experimented with more sophisticated machine learning models, for the purpose of this paper, we wanted to show that we could arrive at a reasonable estimate using a simple and easily reproducible method. A similar approach has been described in [6] for US and UK data.

Figure 3. Receiver operating characteristic of the symptom-only model (left) and full model (symptom, sex, age >70 years, and household infected member model; right). AUC: area under the curve.



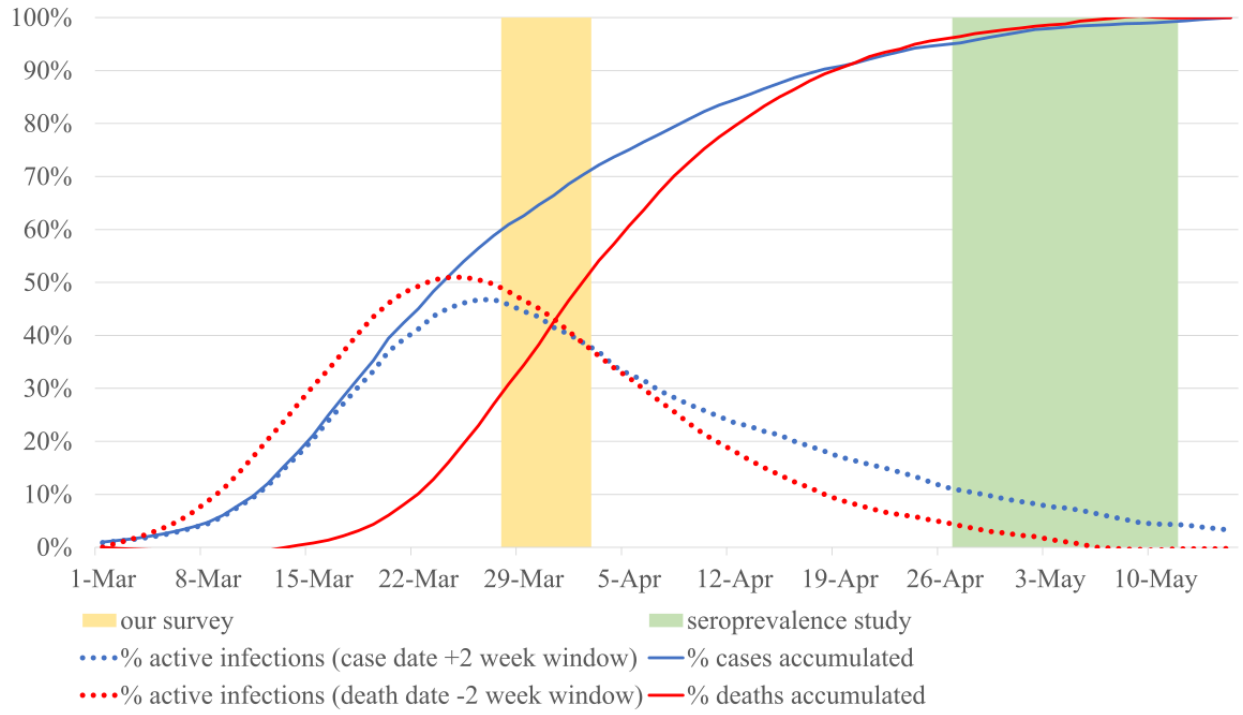
Based on the model with the coefficients below, we estimated the number of SARS-CoV-2 positive individuals among all the respondents to the survey. Geographically, we aggregated the results by the 17 autonomous communities in Spain, making it easier to compare with official data, since each autonomous community has its own health care system, and official figures are always reported by autonomous community.

According to this model, 5798 of the respondents had likely SARS-CoV-2 infections (of which 40% were asymptomatic), which would lead to a prevalence of 5.0% (95% CI ±1.1; after rebalancing by region), suggesting that official tests were only identifying 10% of the infected individuals.

On May 13, 2020, more than a month after we had carried out our analysis, the Spanish INE published the initial results of a nationwide seroprevalence study performed between April 27 and May 11 [22,23]. This study provides ground truth data to assess the estimates of our study, both for the countrywide estimate and at a regional level.

To be able to compare our prevalence estimates with the seroprevalence study, we needed to estimate the proportion of infected individuals at the time of our survey in relation to those detected by the seroprevalence study, since the latter would capture the infected to date and, thus, would include many more individuals than those infected at the time of our survey. We performed this estimation using two different approaches, illustrated in Figure 4.

Figure 4. Two methods for estimating the proportion of active coronavirus infections during the time of our study in relation to those identified by the seroprevalence study. Red dotted line based on the Mortality Monitoring System deaths, assuming infection started 2 weeks prior to death. Blue dotted line based on reported positive cases, assuming infection ended 2 weeks after the case was reported. Cases and deaths from the Carlos III Health Institute in Spain.



In our first estimate of the proportion, we used the excess mortality estimates provided by the Spanish Mortality Monitoring System known as MoMo [40] and made two assumptions: the number of deaths was proportional to the overall number of infected individuals and all infected individuals would have been infected for at least 2 weeks prior to death [41].

Our second estimate used the number of officially reported positive cases and assumed that individuals were infected for at least 2 weeks after their diagnosis.

As shown in Figure 4, both estimates gave similar results; according to the former, 47% (red dotted line) and, according to the latter, 45% (blue dotted line) of individuals that were detected by the seroprevalence study would have had an active SARS-CoV-2 infection during the time of our study. Note that our survey responses have a significant skew toward the beginning of our study, as 73% of the responses of our survey were collected between March 28 and 29, 2020.

In addition, we created a second model based only on symptoms. Although it had a lower area under the curve (see Figure 3, left), this model had a better defined time frame, since it only captured the people who had symptoms during the time of our study—as opposed to also capturing as-of-yet uninfected members of the household who might be infected in the future. Thus, the proportions computed previously would apply better to the symptom-only model.

Table 5 shows the prevalence estimations of each of the models based on our survey answers (symptom-only model and full model) and of the seroprevalence study for each of the 17 autonomous communities in Spain. The symptom-only model estimated a prevalence 40% lower than that of the seroprevalence study.

Table 5. Comparison of inferred prevalence by two models based on the survey answers and the seroprevalence study.

Autonomous community	Participants, n	Symptom-only model, % (95% CI)	Full model, % (95% CI)	Seroprevalence survey, % (95% CI)
Andalucía	5691	2.2 (±0.3)	4.4 (±0.5)	2.7 (2.2-3.2)
Aragón	1463	2.0 (±0.3)	3.1 (±0.9)	4.9 (3.8-6.3)
Asturias	655	1.5 (±0.3)	4.0 (±1.5)	1.8 (1.3-2.5)
Balearic Islands	1222	1.9 (±0.3)	4.2 (±1.1)	2.4 (1.6-3.5)
Canarias	1052	1.4 (±0.2)	3.0 (±1.0)	1.8 (1.1-2.8)
Cantabria	497	2.8 (±0.3)	4.6 (±1.8)	3.2 (2.1-5.0)
Castilla y León	1994	3.7 (±0.4)	6.1 (±1.0)	7.2 (6.3-8.1)
Castilla-La Mancha	3469	8.0 (±0.3)	10.4 (±1.0)	10.8 (9.3-12.4)
Catalonia	5088	2.8 (±0.3)	4.8 (±0.6)	5.9 (4.9-6.9)
Valencia	102,021	1.6 (±0.3)	3.4 (±0.1)	2.5 (1.9-3.2)
Extremadura	656	2.3 (±0.4)	4.4 (±1.6)	3.0 (2.2-4.1)
Galicia	2257	1.3 (±0.3)	2.6 (±0.7)	2.1 (1.7-2.6)
Madrid	10,365	6.1 (±0.4)	8.8 (±0.5)	11.3 (9.8-13.0)
Murcia	3566	1.5 (±0.3)	3.2 (±0.6)	1.4 (0.8-2.4)
Navarra	580	3.6 (±0.4)	5.5 (±1.9)	5.8 (4.3-7.6)
País Vasco	1007	1.9 (±0.4)	3.9 (±1.2)	4.0 (3.1-5.2)
Rioja, La	220	1.8 (±0.4)	5.0 (±2.9)	3.3 (2.4-4.4)
National	141,803	3.0 (±0.3)	5.0 (±1.1)	5.0 (4.7-5.4)

The prevalence estimates by the full model are closer to the estimates provided by the seroprevalence study. This finding might be explained by the fact that 40% of the identified likely infections by the full model were not based on symptoms, but instead based on the variable that captures if the respondents shared their home with an infected individual. Due to the harsh nature of the lockdown in Spain at that time, including the banning of all mobility except essential labor and basic needs between March 30 and April 9, 2020 (both included) [42], many of the new infections in the period between the end of our study and the end of the seroprevalence survey (ie, between April 4 and May 17) may have been household members captured by our model.

Discussion

Principal Findings

Through the survey answers, we identified several patterns and implications for the design of public policies in the context of the COVID-19 pandemic.

First, our work highlights the value of involving the population and carrying out large-scale online surveys for a quick assessment of the situation and perceptions during a pandemic. We were overwhelmed by the response to the survey. Mayors in large and small towns got involved and shared it with their employees and residents, professional and civic associations disseminated it among their members, individuals advertised it among their contacts, and a few media organizations gave it visibility via articles and posts. This outstanding response by people might reflect a societal need to have more information about the impact of the COVID-19 pandemic in our lives and is an example of citizen's science and people's willingness to help by contributing with their answers to achieve more data-driven decision-making processes. Although the sample has some biases, we used reweighting to mitigate them.

Second, we empirically corroborate the impact that close contacts play in the transmission of the disease. Over 16% of respondents reported having had close contact with someone who was infected by SARS-CoV-2. This percentage was much higher (80.9%) among those who had tested positive for coronavirus. According to this finding, those testing positive were likely infected by someone they knew and had close contact with, rather than, for example, an unknown infected stranger in a supermarket. This finding could have implications for contact tracing strategies.

Third, gender matters. Several statistically significant differences were found between male and female respondents, with a pattern of placing women in situations of higher vulnerability or exposure when compared to men. As in other aspects of society, gender-based differences exist in the context of a pandemic. It is a socially important factor that needs to be considered.

Fourth, age also matters. We identified statistically significant differences in the social contact behavior questions between older participants (age >60 years) and younger participants (age <60 years). Older respondents were almost twice as likely to stay entirely at home than younger participants. There were also different aged-based attitudes toward the containment measures; younger participants were significantly more supportive of stronger measures than older participants, while they were more likely to report not being able to stand the confinement any more (4.9%) versus older adults (0.8%). We also found

associations between age, gender, the ability to self-isolate, and the probability of testing positive; older females without the capacity to isolate themselves were almost twice as likely to test positive than otherwise.

Participants demanded more measures, as 50.4% of respondents were supportive of implementing additional social distancing measures. This result might reflect the worry in people's minds regarding the exponential progression of the pandemic and the lack of clear signs of flattening the curve at the time of answering the survey.

Moreover, the majority of respondents (76.5%) were willing to remain in confinement for a month or more, and 32.4% of respondents reported being able to do so for 3-6 additional months.

Even at the end of March, the economic impact of the pandemic was evident, particularly for those working in small companies, 19.4% of which reported to be facing bankruptcy. Moreover, over 47.3% of participants who worked in small companies reported having been impacted by the pandemic. In terms of professions, hospitality, construction, and retail were the most affected. Hospitality represents 6.2% of the Spanish GDP [43] and construction 5.6% [44]. We expect the economic impact to be significantly larger as the pandemic progresses.

Among those who were working, 28.7% of respondents reported tele-working and one-third leaving the home to go to work. The tele-work figure is lower than in other countries. For example, in the United States, it is estimated that 56%-62% of the workforce could work remotely. Moreover, on March 31, 2020, the government established labor mobility restrictions for all nonessential professions. Given that 71.2% of respondents (n=141,865) reported having worked in the last month, our expectation is that about 23% of the population would have been impacted by such measures. Regarding workplace infections, we found that 11.1% of those who tested positive (and did not work in the health care sector) had close contact with someone at work who had tested positive for coronavirus.

Close, known contacts seem to play a large role on infections, as 80.9% of those who had tested positive responded having had close contact with a known infected individual. This finding is relevant in the design of contact tracing, testing, and isolation strategies.

Quarantine infrastructure might be needed, as over 27.7% of respondents reported not having the appropriate infrastructure to isolate themselves at home. Effective quarantine measures for asymptomatic or lightly symptomatic patients are key to control the spread of the pandemic. Thus, developing the needed infrastructure might be key to slowdown the transmission of the disease.

The number of SARS-CoV-2 infected individuals in Spain in March and April was certainly larger than what has been officially reported. In our survey, over 16.8% of respondents reported having at least one possibly COVID-19–related symptom. We show how the prevalence of a rapidly spreading disease such as COVID-19 can be estimated using a large-scale population survey. From the answers to two of the questions (Q22: symptoms and Q8: contact with coronavirus-infected individual in the household) plus demographic information, we built a generalized linear model and reported SARS-CoV-2 prevalence estimations that are on par with those carried out by a seroprevalence study in Spain. As shown in related studies [6], when public policy decisions need to be made rapidly in a situation of data scarcity and limited testing capacity, a large-scale population survey might be of great value to make fast assessments of prevalence, as it can be deployed rapidly and enable the collection of results within hours.

Finally, in the context of Spain, our survey revealed a lack of tests; over 6.1% of respondents reported not being able to do the test despite their doctor's recommendation. Moreover, a significant difference was found between those who had at least one of three COVID-19 symptoms, namely, dry cough, fever, and difficulty breathing, and those who did not regarding the impact of testing unavailability; 32.5% of the symptomatic individuals reported that tests were not available despite their doctor's recommendation. Thus, we found that there was a need for more tests.

Limitations

Although the sample size in our study is large, our methodology is not exempt of limitations. First, there are several sources of bias in our study: selection bias, given that all participants volunteered to fill out the survey without any incentive; self-reported bias; and sampling bias as we used a nonprobability sampling technique. Geographically, we lacked representation of several geographic regions in Spain and particularly rural areas. In our analysis, we tried to mitigate some of these biases by correcting for gender, age, location, and profession via reweighting using the Spanish INE census data. However, reweighting does not eliminate the risk of selection bias. Second, this is an in-the-wild study, and thus, people could have provided untruthful answers. We addressed this limitation by filtering entries without proper zip codes and entries that had inconsistencies in them. In addition, our study provides a snapshot over a 5-day time period, such that the results are only representative of this time period. We have addressed this limitation by deploying the study on a weekly basis since its first deployment at the end of March [28].

Finally, our prevalence estimates have several limitations, including overestimation of symptoms due to the existence of other flu-like illnesses at the time, and selection bias for symptomatic individuals who might have been more motivated to respond or forward the survey. We also lacked detail regarding which SARS-CoV-2 test respondents had taken, the reliability of these tests, and the relative timeframe of when the tests were taken versus symptoms. However, as noted in the manuscript, our goal is to show the value of large-scale, online, self-reported population studies to quickly and cheaply approximate the prevalence of COVID-19 when testing capacity is limited and data is scarce, as was the case in Spain at the time of the study and as might be the case in other countries in the future.

Conclusions

The COVID-19 pandemic is undoubtedly having a major impact on the lives of people worldwide. Although there is data

regarding the number of reported cases, hospitalizations and intensive care patients, and deaths, there is a scarcity of data about the individual experiences of people; their personal, financial and labor situations; their health state; and their attitudes toward the confinement measures. This paper reports the first results of analyzing a large-scale, rich data set of self-reported information regarding the social contact, economic impact, working situation, and health status of over 140,000 individuals in Spain. It is the largest population survey carried out in Spain in the context of an infectious disease pandemic.

The data is extremely rich and multifaceted. Thus, it offers numerous avenues of future work and deeper analysis according to different dimensions, including location (at a zip code level), which we have not covered in this paper.

We have launched successive versions of the *Covid19Impact* survey [27] in consecutive weeks throughout the COVID-19 pandemic to assess the COVID-19 situation from the perspective of the population in Spain over time and identify changes in people's situations and perceptions regarding the pandemic.

Acknowledgments

We thank the thousands of participants who volunteered to fill out this survey and shared it with their contacts. Their generosity and enthusiasm have enabled this valuable data set to be collected. Kristof Roomp would like to thank his employer, Microsoft, for letting him volunteer his time to help on this effort. This project has been carried out in collaboration with the Valencian Government of Spain via the position of NO as Commissioner to the President of the Valencian Region on Artificial Intelligence Strategy and Data Science against COVID-19. It was partially funded by the research project "CD4COVID" awarded by "FONDO Supera COVID-19" from Banco Santander, the Spanish National Research Council and Conference of Rectors of Spanish Universities.

Authors' Contributions

NO and Kristof Roomp conceptualized the study, interpreted and analyzed the data, drafted the manuscript, and provided supervision. Kristof Roomp deployed the survey and parsed the responses. XB reweighted and analyzed the data. Kirsten Roomp conceptualized the study, carried out the literature review and provided feedback to the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions translated from Spanish.
[DOCX File , 18 KB - jmir_v22i9e21319_app1.docx]

Multimedia Appendix 2
Univariate tables.
[DOCX File , 62 KB - jmir_v22i9e21319_app2.docx]

Multimedia Appendix 3
A .txt file is provided with the 141,865 survey answers used in this paper. To protect the anonymity of the survey respondents, only province-level geographic location information is included, the timestamp has been removed, and the answers have been randomly sorted to prevent deanonymization.
[TXT File , 44646 KB - jmir_v22i9e21319_app3.txt]

References

1. Coronavirus disease 2019 (COVID-19) situation reports. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
2. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
3. Gong W, Taighoon Shah M, Firdous S, Jarrett BA, Moulton LH, Moss WJ, et al. Comparison of three rapid household survey sampling methods for vaccination coverage assessment in a peri-urban setting in Pakistan. Int J Epidemiol 2019 Apr 01;48(2):583-595. [doi: 10.1093/ije/dyy263] [Medline: 30508112]
4. Kennedy C, Hartig H. Response rates in telephone surveys have resumed their decline. Pew Research Center. 2019 Feb 27. URL: https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/
5. Abraham K, Presser S, Helms S. How social processes distort measurement: the impact of survey nonresponse on estimates of volunteer work in the United States. AJS 2009 Jan;114(4):1129-1165. [doi: 10.1086/595945] [Medline: 19824303]

6. Allen WE, Altae-Tran H, Briggs J, Jin X, McGee G, Raghavan R, et al. Population-scale longitudinal mapping of COVID-19 symptoms, behavior, and testing identifies contributors to continued disease spread in the United States. medRxiv 2020 Jun 11:A. [doi: 10.1101/2020.06.09.20126813] [Medline: 32577674]
7. Surveys. Oxford Supertracker: The Global Directory for COVID Policy Trackers and Surveys. URL: https://supertracker.spi.ox.ac.uk/surveys/
8. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
9. Flatten. 2020. URL: https://flatten.ca/
10. Fetzer T, Witte M, Hensel L, Jachimowicz J, Haushofer J, Ivchenko A, et al. Perceptions of an insufficient government response at the onset of the COVID-19 pandemic are associated with lower mental well-being. PsyArXiv 2020 Apr 16:A [FREE Full text] [doi: 10.31234/osf.io/3kfmh]
11. COVID-19:CH Survey. 2020. URL: https://covid19survey.ethz.ch/survey
12. coronaisrael.org. 2020. URL: https://coronaisrael.org/en/
13. Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nat Med 2020 May;26(5):634-638 [FREE Full text] [doi: 10.1038/s41591-020-0857-9] [Medline: 32273611]
14. Bagheri SHR, Asghari AM, Farhadi M, Shamshiri AR, Kabir A, Kamrava SK, et al. Coincidence of COVID-19 epidemic and olfactory dysfunction outbreak. medRxiv 2020 Mar 27:A. [doi: 10.1101/2020.03.23.20041889]
15. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
16. Geldsetzer P. Knowledge and perceptions of COVID-19 among the general public in the United States and the United Kingdom: a cross-sectional online survey. Ann Intern Med 2020 Jul 21;173(2):157-160 [FREE Full text] [doi: 10.7326/M20-0912] [Medline: 32196071]
17. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
18. Nelson LM, Simard JF, Oluyomi A, Nava V, Rosas LG, Bondy M, et al. US public concerns about the COVID-19 pandemic from results of a survey given via social media. JAMA Intern Med 2020 Apr 07:1020-1022 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1369] [Medline: 32259192]
19. Atchison CJ, Bowman L, Vrinten C, Redd R, Pristera P, Eaton JW, et al. Perceptions and behavioural responses of the general public during the COVID-19 pandemic: a cross-sectional survey of UK Adults. medRxiv 2020 Apr 03:A. [doi: 10.1101/2020.04.01.20050039]
20. Varshney M, Parel JT, Raizada N, Sarin SK. Initial psychological impact of COVID-19 and its correlates in Indian Community: an online (FEEL-COVID) survey. PLoS One 2020;15(5):e0233874 [FREE Full text] [doi: 10.1371/journal.pone.0233874] [Medline: 32470088]
21. Flaxman S, Mishra S, Gandy A, Unwin H, Coupland H, Mellan T, et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London 2020 Mar 30:1-35 [FREE Full text] [doi: 10.25561/77731]
22. Instituto Nacional de Estadística. URL: https://www.ine.es/en/
23. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Lancet 2020 Aug;396(10250):535-544. [doi: 10.1016/S0140-6736(20)31483-5]
24. Datos enero 2020. Cifras PyME. 2020. URL: http://www.ipyme.org/Publicaciones/CifrasPYME-enero2020.pdf
25. Goodman LA. Snowball Sampling. Ann Math Statist 1961 Mar;32(1):148-170. [doi: 10.1214/aoms/1177705148]
26. Survey123. URL: https://survey123.arcgis.com/
27. Covid19Impact Survey. 2020. URL: https://covid19impactsurvey.org
28. Portal d'Informacio Argos. URL: http://www.argos.gva.es/va/inicio/
29. Rao JNK, Scott AJ. On simple adjustments to chi-square tests with sample survey data. Ann Statist 1987 Mar;15(1):385-397. [doi: 10.1214/aos/1176350273]
30. Agresti A. Categorical Data Analysis. Hoboken, New Jersey: John Wiley & Sons; Mar 31, 2003.
31. Lumley T, Scott A. Fitting regression models to survey data. Statist Sci 2017 May;32(2):265-278. [doi: 10.1214/16-sts605]
32. INEbase. URL: https://www.ine.es/jaxiT3/Tabla.htm?t=4128
33. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med 2020 Apr;26(4):506-510 [FREE Full text] [doi: 10.1038/s41591-020-0822-7] [Medline: 32284616]
34. Bayer C, Kuhn M. Intergenerational ties and case fatality rates: a cross-country analysis. SSRN J 2020 Apr 13:1-12.

35. Ausín B, González-Sanguino C, Castellanos Á, Muñoz M. Gender-related differences in the psychological impact of confinement as a consequence of COVID-19 in Spain. J Gender Stud 2020 Aug 04:1-10. [doi: 10.1080/09589236.2020.1799768]
36. Krantz SG, Rao ASRS. Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling. Infect Control Hosp Epidemiol 2020 Apr 09:1-3 [FREE Full text] [doi: 10.1017/ice.2020.116] [Medline: 32268929]
37. Tremlett G. How did Spain get its coronavirus response so wrong? 2020 Mar 26. URL: https://www.theguardian.com/world/2020/mar/26/spain-coronavirus-response-analysis
38. Escudero J, Méndez R. El quebradero de cabeza de las cifras oficiales del coronavirus en España. El Confidencial. 2020 Mar 19. URL: https://www.elconfidencial.com/espana/2020-03-19/coronavirus-comunicacion-datos-ministerio-sanidad_2505867/
39. Pearce E. What is anosmia and is it a symptom of COVID-19? ESM1. 2020 Mar 27. URL: https://www.ems1.com/coronavirus-covid-19/articles/what-is-anosmia-and-is-it-a-symptom-of-covid-19-Mcy5H9WmR7v6gUbk/
40. MoMo Nacional. URL: https://momo.isciii.es/public/momo/dashboard/momo_dashboard.html
41. Tenforde MW, Kim SS, Lindsell CJ, Billig Rose E, Shapiro NI, Files DC, IVY Network Investigators, CDC COVID-19 Response Team. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network - United States, March-June 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 31;69(30):993-998. [doi: 10.15585/mmwr.mm6930e1] [Medline: 32730238]
42. Orden SND/307/2020, 30 de marzo, por la que se establecen los criterios interpretativos para la aplicación del Real Decreto-ley 10/2020, de 29 de marzo, y el modelo de declaración responsable para facilitar los trayectos necesarios entre el lugar de residencia y de trabajo. BOE. 2020 Mar 30. URL: https://www.boe.es/buscar/pdf/2020/BOE-A-2020-4196-consolidado.pdf
43. Anuario de la hostelería. Hosfrinor.com. 2019. URL: https://hosfrinor.com/wp-content/uploads/2019/12/Anuario-de-la-Hosteleri%CC%81a-de-Espan%CC%83a-2019.pdf
44. Evolución anual del peso de la industria de la construcción en el PIB de España desde 2005 hasta 2018. Statista. URL: https://es.statista.com/estadisticas/549605/aportacion-del-sector-de-la-construccion-al-pib-en-espana/

Abbreviations

GDP: gross domestic product
INE: National Institute of Statistics
Q: question

Edited by A Moorhead, G Eysenbach; submitted 11.06.20; peer-reviewed by C García, A Sudaryanto, P Valente; comments to author 11.08.20; revised version received 31.08.20; accepted 01.09.20; published 10.09.20.

Please cite as:
Oliver N, Barber X, Roomp K, Roomp K
Assessing the Impact of the COVID-19 Pandemic in Spain: Large-Scale, Online, Self-Reported Population Survey
J Med Internet Res 2020;22(9):e21319
URL: http://www.jmir.org/2020/9/e21319/
doi:10.2196/21319
PMID:32870159

Original Paper

Association Between Generalized Anxiety Disorder Scores and Online Activity Among US Adults During the COVID-19 Pandemic: Cross-Sectional Analysis

Parvati Singh[1], PhD; William G Cumberland[2], PhD; Dominic Ugarte[3], MD; Tim-Allen Bruckner[4], PhD; Sean D Young[1], PhD

[1]University of California Institute for Prediction Technology, University of California, Irvine, Irvine, CA, United States

[2]Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

[3]School of Medicine, University of California, Irvine, Irvine, CA, United States

[4]Program in Public Health, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Parvati Singh, PhD
University of California Institute for Prediction Technology
University of California, Irvine
Irvine, CA
United States
Phone: 1 310 794 8530
Email: parvatis@uci.edu

Abstract

Background: Evidence from past pandemics suggests that fear, uncertainty, and loss of control during large-scale public health crises may lead to increased pandemic-related information seeking, particularly among persons predisposed to high anxiety. In such groups, a greater consumption of information pertaining to the COVID-19 pandemic may increase anxiety.

Objective: In this study, we examine the association between online activity and Generalized Anxiety Disorder 7 (GAD-7) scores in the United States.

Methods: We recruited participants for an online survey through advertisements on various platforms such as Google, Facebook, and Reddit. A total of 406 adult US participants with moderate to severe (≥10) GAD-7 scores met the inclusion criteria and completed the survey. Anxiety levels measured using the GAD-7 scale formed our primary outcome. Our key independent variables were average daily time spent online and average daily time spent online searching about COVID-19 within the past 14 days. We used as controls potential confounders of the relation between our key independent variables and GAD-7 scores, namely, sleep quality, the COVID-19 Fear Inventory scale, binge drinking, substance use, prescription drug abuse, and sociodemographic attributes.

Results: Linear multivariate regression analyses showed that GAD-7 scores were higher among those who spent >4 hours online (per day) searching for information about COVID-19 (coefficient 1.29, *P*=.002), controlling for all other covariates. The total time spent online was not statistically associated with GAD-7 scores.

Conclusions: Results from this study indicate that limiting pandemic-related online information seeking may aid anxiety management in our study population.

(J Med Internet Res 2020;22(9):e21490) doi:10.2196/21490

KEYWORDS

online activity; COVID-19; anxiety; generalized anxiety disorder; GAD; scores; stress; anxiety; internet; survey; cross-sectional

Introduction

The COVID-19 pandemic has impacted multiple aspects of everyday life across the world. Social distancing measures mandated by state governments in the United States, including stay-at-home orders and closure of nonessential businesses, have preceded record-breaking unemployment rates and an unprecedented US $3 trillion relief and economic stimulus package from Congress [1-3]. These sweeping changes have altered daily routines of the US population and may correspond

with marked changes in mental and behavioral health, exacerbated by social isolation, fear, uncertainty, and financial strain during this pandemic [4].

The convergence of social isolation, heightened ambient stress, and potential loss of resources including access to health care has elicited calls for research on correlates of mental health outcomes amid the current pandemic [5]. Understandably, COVID-19 has spurred a substantial amount of research on mental health in the present times. According to a recent meta-analytic review, populations across multiple geographies worldwide have experienced an approximately 30% increase in anxiety, depression, and stress [6]. These reports cohere with mental distress observed in the United States [7]. About one-third of the adult population in the United States has reported increased anxiety, trauma, depression, and substance use in the months following the COVID-19 pandemic [7]. Approximately 1 in 10 US adults have also reported having seriously contemplated suicide [7].

The initial stages of pandemics invariably correspond with widespread anxiety [8]. Scholars examining past pandemics report overconsumption of information (and misinformation) from 24-hour news cycles and social media as potential coping mechanisms among populations faced with fear and uncertainty [8]. Individual-level psychological responses to pandemics may vary by personal intolerance to uncertainty thresholds [8,9]. For instance, during the 2014 swine flu epidemic, people with high intolerance to uncertainty were also more likely to report fear and anxiety about getting infected with the virus, relative to those with lower intolerance [9]. Anxiety correlates positively with intolerance to uncertainty that in turn, may lead to reassurance-seeking behaviors such as frequently checking the internet for medical information or pandemic-related news updates [8]. This maladaptive coping behavior may turn into a self-perpetuating negative feedback cycle. Among persons vulnerable to emotional distress, repeat consumption of information online may thus lead to overestimation of the pandemic threat and exacerbation of anxiety [8].

Research on the mental health impacts of the COVID-19 pandemic in China, which was the first country affected by this pandemic, finds a positive correlation between anxiety and exposure to news and information about COVID-19 online [10]. In one study, a cross-sectional sample of about 4800 online survey participants reported 72% higher odds of anxiety symptoms (but not depression) with frequent exposure to news and information about COVID-19, relative to those with infrequent or low exposure [10]. Another web-based cross-sectional survey of over 7000 people in China also reported an increase in Generalized Anxiety Disorder 7 (GAD-7) scores with more than 3 hours (per day) spent discussing or searching about COVID-19 [11]. Huang and Zhao [11] observed poorer sleep quality- a strong correlate of anxiety and depression- among health care workers, in alignment with contemporaneous reports of adverse mental health outcomes in this occupational group [12]. Research examining trends in mental health–related search terms on Google across multiple countries also reported an increase in search volume for “insomnia” immediately following the current pandemic in those regions [13].

Online studies are increasingly becoming a popular means of conducting health research during the COVID-19 pandemic [14-16]. Online surveys have wide reach, as a large majority of the population has rapidly gained access to the internet over the past decade with over 80% of households in the United States reporting having an internet-connected computer or smartphone [17]. Quarantine and social distancing measures currently in effect owing to the COVID-19 pandemic make online surveys particularly lucrative for data collection and research [15,16].

This study examines the relation between GAD-7 scores and time spent online searching for information about COVID-19 in a cross-section of 406 online survey participants with moderate to severe anxiety in the United States. To our knowledge, this is the first study to examine the relation between online activity pertaining to COVID-19 and GAD-7 scores in a US sample, and may help augment our understanding of the mental health consequences of the current pandemic.

Methods

Participant Recruitment

We recruited study participants through online advertisements on Facebook, Google ads, and online forums such as Reddit. Potential participants who clicked on the online ads were routed to a Qualtrics survey where they were screened for eligibility using the GAD-7 screening tool, which we modified in relation to COVID-19 (for the detailed modified GAD-7 instrument, please see the first table in Multimedia Appendix 1). To be eligible for participation, participants needed to meet the following inclusion criteria: adults 18 years or older who were competent enough to give informed consent, English speakers, moderate to severe GAD-7 scores (≥10) in relation to COVID-19, not currently taking anxiety medication, uses social media or online communities greater than twice per week, willing and capable of understanding and assenting to an online informed consent form, has or is willing to accept a group invitation from our Facebook social media page, and has completed the survey.

Those who did not meet all items on the inclusion list were ineligible for the study. After eligible participants provided their consent online and liked or messaged our study’s Facebook page, they were invited to complete our study survey, for which they received US $15 Amazon gift cards post survey completion. Given that persons with pre-existing anxiety appear differentially exposed and vulnerable to online information seeking during pandemics, we selected persons who have moderate to severe anxiety (ie, ≥10 GAD-7 score) [18] to have a sufficiently large sample size of persons with high anxiety for our study.

The survey was conducted from March 15, 2020, to April 8, 2020. University of California, Irvine’s Institutional Review Board (IRB) deemed this study as IRB exempt. Our analytic sample comprised of 406 respondents who met the inclusion criteria and responded to all items on the survey.

Variables

For the dependent variable, we measured anxiety levels using the GAD-7 scale [18]. Scores on the GAD-7 scale are determined through participant responses to a 7-item anxiety

questionnaire. The resulting (additive) GAD-7 score was defined as the outcome of interest.

The key independent variables included time spent online (6 categories: none, 1 minute to 2 hours, 2-4 hours, 4-6 hours, 6-8 hours, >8 hours) and time spent online searching about COVID-19 (6 categories: none, 1 minute to 2 hours, 2-4 hours, 4-6 hours, 6-8 hours, >8 hours). We converted time spent online searching about COVID-19 from categorical to binary (≤4 hours per day or >4 hours per day) for clarity of interpretation (very few participants reported spending <2 hours searching about COVID-19 per day).

As described previously, pandemics correspond with a decline in quality of sleep, increased pandemic fear, and increased substance use [7,8]. These factors may drive greater online information seeking or serve as mediators of the relation between time spent online and increased anxiety. To that end, we included as our control variables binge drinking (binary: <5 or ≥5 alcoholic drinks in any day for males, ≥4 alcoholic drinks in any day for females, in the past 14 days), drug use (eg, marijuana, cocaine, heroin), and prescription drug abuse (eg, opioids, Adderall) in the past 14 days (binary: yes or no). We also derived a sleep scale using responses to items adapted from the Medical Outcomes Study (MOS) scale through Cronbach alpha and used this continuous variable as an exposure (α=.78; second table in Multimedia Appendix 1) [19]. In addition, we used a validated pandemic fear inventory scale and defined a COVID-19 Fear Inventory scale using Cronbach alpha estimates from 10 item responses (α=.76; third table in Multimedia Appendix 1) [20]. Inclusion of these covariates allows us to estimate the relation between our dependent and key independent variables "net of" plausible and measurable confounding factors. Other control variables included education (binary: <high school or ≥high school), sex (male, female), race (non-Hispanic White, non-Hispanic Black, Hispanic, Asian, others), and age (in years).

Statistical Analysis

We used ordinary least squares multivariate regression analysis to examine the relation between GAD-7 scores and the independent, control variables. We specified robust standard errors to account for heteroskedasticity. Regression coefficients with $P<.05$ were deemed statistically significant. All analyses were performed in Stata SE version 14.2 (StataCorp) [21]. There were no missing data in our analytic data set, as all 406 participants completed the survey in entirety.

Results

Table 1 shows the descriptive statistics of the variables included in this study. The average GAD-7 score among the study participants was 17.2, and mean scores on the MOS sleep scale and COVID-19 Fear Inventory scale were 2.2 and 5, respectively. About 83% (336/406) of the participants reported greater than 4 hours spent searching for information about COVID-19 online per day in the past 14 days. The mean age of study participants was approximately 39 years, with 82% (332/406) females and 18% (74/406) males. Approximately 75% (307/406) were non-Hispanic White, 7% (28/406) non-Hispanic Black, and 10% (41/406) Hispanic. Most participants reported having a high school or General Educational Development (GED), or higher educational attainment (394/406, 97%).

Table 1. Descriptive statistics of variables included in study (N=406).

Variables	Participants
GAD-7[a] score, mean (SD)	17.2 (3.2)
MOS[b] Sleep Scale score, mean (SD)	2.2 (0.9)
COVID-19 Fear Inventory score, mean (SD)	4.2 (0.5)
Age (years), mean (SD)	38.9 (12.7)
Time spent online per day in the past 14 days, n (%)	
None	0 (0.0)
1 minute to 2 hours	6 (1.5)
2-4 hours	62 (15.27)
4-6 hours	103 (25.37)
6-8 hours	107 (26.35)
>8 hours	128 (31.53)
Greater than 4 hours daily spent online searching for/reading information/watching videos about COVID-19 (eg, causes of it, symptoms, effects of COVID-19 on society/the stock market/employment issues, news reports about it) in the past 14 days, n (%)	336 (82.8)
Binge drinking on any day within the past 14 days, n (%)	98 (24.1)
Used any drugs including marijuana, cocaine or crack, heroin, methamphetamine (crystal meth), hallucinogens, ecstasy/MDMA in the past 14 days, n (%)	69 (17)
Used any prescription medications just for the feeling, more than prescribed, or that were not prescribed (eg, OxyContin, Vicodin, Percocet, Methadone, Xanax, Ativan, Klonopin, Adderall, or Ritalin), n (%)	20 (4.9)
Race/ethnicity, n (%)	
Non-Hispanic White	307 (75.6)
Non-Hispanic Black	28 (6.9)
Hispanic	41 (10.1)
Asian	14 (3.45)
Other	16 (3.94)
Greater than or equal to high school/GED[c] education, n (%)	394 (97)
Sex, n (%)	
Female	332 (81.8)
Male	74 (18.2)

[a]GAD-7: Generalized Anxiety Disorder 7.

[b]MOS: Medical Outcomes Study.

[c]GED: General Educational Development.

Table 2 shows the results from the linear regression analysis predicting GAD-7 scores as a function of covariates included in the study. The time spent in daily online activity was not statistically associated with GAD-7 scores. However, more than 4 hours of daily online activity searching for information about COVID-19 was associated with higher GAD-7 scores (coefficient 1.29, *P*=.002). This relation remained robust when using the 6-category (none, 1 minute to 2 hours, 2-4 hours, 4-6 hours, 6-8 hours, >8 hours) version of daily online activity searching for information about COVID-19 as well (results available upon request). The sleep scale scores varied inversely (coefficient –1.02, *P*<.001), and the COVID-19 Fear Inventory scale scores varied positively and significantly (coefficient 1.46, *P*<.001) with GAD-7 scores. Misuse or abuse of prescription drugs in the past 14 days was associated with higher GAD-7 scores (coefficient 1.12, *P*=.007). Participants with high school or GED, or higher education showed lower GAD-7 scores relative to those without high school education (coefficient –2.77, *P*<.001). Relative to non-Hispanic Whites, GAD-7 scores did not vary across non-Hispanic Blacks, Hispanics, Asians, and other races or ethnicities. We also observed no relation between GAD-7 scores and age, sex, alcohol use, and use of nonprescription drugs such as marijuana, cocaine, or heroin.

Table 2. Multivariate linear regression results predicting Generalized Anxiety Disorder 7 score as a function of key covariates.

Covariates	Coefficient (95% CI)	*P* value
Time spent online per day in the past 14 days (reference: 1 minute to 2 hours)		
2-4 hours	–0.59 (–2.54 to 1.36)	.55
4-6 hours	–0.05 (–1.95 to 1.85)	.96
6-8 hours	–0.55 (–2.45 to 1.35)	.57
>8 hours	–0.03 (–1.92 to 1.87)	.98
Greater than 4 hours online activity searching for/reading information/watching videos about COVID-19 per day	1.29 (0.47 to 2.11)	.002
Consumed more than 5 alcoholic drinks a day (in the past 14 days)	0.31 (–0.32 to 0.93)	.34
Used any drugs including marijuana, cocaine or crack, heroin, methamphetamine (crystal meth), hallucinogens, ecstasy/MDMA in the past 14 days	–0.18 (–0.94 to 0.57)	.64
Used prescription medications just for the feeling, more than prescribed, or that were not prescribed (eg, opiates, medications for anxiety, sleep, ADHD[a]) in the past 14 days	1.12 (0.30 to 1.93)	.007
MOS[b] Sleep Scale score	–1.02 (–1.35 to –0.69)	<.001
COVID-19 Fear Inventory Scale score	1.47 (0.76 to 2.17)	<.001
Age	0.00 (–0.03 to 0.02)	.91
Female (reference: male)	0.66 (–0.12 to 1.45)	.10
Greater than or equal to high school/GED[c] education	–2.77 (–4.10 to –1.44)	<.001
Race/ethnicity (reference: non-Hispanic White)		
Non-Hispanic Black	1.24 (–0.89 to 1.12)	.82
Hispanic	1.43 (–0.72 to 1.22)	.61
Asian	1.49 (–0.33 to 2.60)	.13
Other	–1.25 (–2.55 to 0.05)	.06
Sample size (N)	406	N/A[d]
Adjusted R^2	0.221	N/A

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]MOS: Medical Outcomes Study.

[c]GED: General Educational Development.

[d]N/A: not applicable.

Discussion

Principal Results

We found that spending an average of 4 or more hours per day searching online about COVID-19 was associated with greater anxiety (measured using GAD-7 scores) in our cross-sectional sample of 406 US adults with moderate to severe anxiety. Relative to those who spent less than 4 hours on pandemic-related online searches, those who exceeded 4 hours of COVID-19–related online activity had GAD-7 scores that were 1.3 units higher. Our analyses controlled for individual sociodemographic attributes, time spent online overall, substance use, binge drinking, sleep quality, and the COVID-19 Fear Inventory. Our findings align with those from prior research and highlight the relation between online information seeking and worsened anxiety symptoms in the context of the COVID-19 pandemic [8,10,11].

Limitations

Limitations include the cross-sectional nature of our study design that restricts causal inference and generalizability. We also note that, although GAD-7 scores may be used to assist with a psychiatric diagnosis of generalized anxiety disorder, this study uses them to approximate anxiety levels (over the past 2 weeks preceding the survey only) and not as a clinical diagnosis. Clinical diagnosis of generalized anxiety disorder, per the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition criteria, typically requires continued observation over a period of 6 months [18]. For the purposes of this study, GAD-7 scores are well suited to capture proximate responses to a sudden, exogenous, and acute stressor (ie, the COVID-19 pandemic).

Implications for Future Research

We encourage future research to examine whether (as indicated by our findings) relatively simple measures such as limiting exposure to COVID-19–related information may help in

managing anxiety during pandemics. Our study may aid future research in examining the effect of misinformation during pandemics because persons exposed to a higher volume of COVID-19–related information may also, statistically speaking, be exposed to greater misinformation (owing to the sheer volume of exposure). People with high anxiety and high intolerance to uncertainty, when exposed to quack cures and misinformation, may be more likely to act on potentially fatal "preventive" measures [8,22]. Curation of accurate information available online (eg, fact-checking by scientific organizations) may serve as an effective tool in preventing adverse outcomes [23].

It is plausible that persons with high anxiety *select into* greater pandemic-related information seeking online. In such cases, the propensity of persons with high anxiety to spend substantial amounts of time online may be used in targeted internet-based mental health service delivery. High frequency internet and social media users with pre-existing mental health conditions may serve as an accessible group for telepsychiatry, telemedicine, and social media–based interventions [5]. Such interventions and services offer the dual benefit of reaching target (ie, high anxiety) populations while adhering to pandemic prevention guidelines (eg, lower interpersonal contact, social distancing) [5].

Conclusion

In this cross-sectional study, we observed a positive relation between time spent online searching for information about COVID-19 and GAD-7 scores. Our analyses may assist in understanding correlates of mental health outcomes during the ongoing COVID-19 pandemic.

Acknowledgments

We wish to thank John Liu for feedback on the anxiety measure, GAD-7.

Conflicts of Interest

SDY is an advisor to health technology startups, has received compensation for speaking at the PriMed Conference, and has received gift funding from Facebook and Intel.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 19 KB - jmir_v22i9e21490_app1.docx]

References

1. Gostin LO, Wiley LF. Governmental public health powers during the COVID-19 pandemic: stay-at-home orders, business closures, and travel restrictions. JAMA 2020 Apr 02:A. [doi: 10.1001/jama.2020.5460] [Medline: 32239184]
2. News release: unemployment insurance weekly claims. US Department of Labor. Washington, D.C. 20210: U.S. Department of Labor Employment and Training Administration; 2020 Aug 27. URL: https://www.dol.gov/ui/data.pdf [accessed 2020-09-02]
3. Coronavirus: Congress passes $484bn economic relief bill. BBC News. 2020 Apr 24. URL: https://www.bbc.com/news/world-us-canada-52398980 [accessed 2020-09-02]
4. Holmes E, O'Connor RC, Perry V, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560. [doi: 10.1016/S2215-0366(20)30168-1]
5. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Apr 10:A. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
6. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
7. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057. [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
8. Taylor S. The Psychology of Pandemics: Preparing for the next Global Outbreak of Infectious Disease. Newcastle upon Tyne, UK: Cambridge Scholars Publishing; 2019.
9. Taha S, Matheson K, Anisman H. The 2009 H1N1 influenza pandemic: the role of threat, coping, and media trust on vaccination intentions in Canada. J Health Commun 2013;18(3):278-290. [doi: 10.1080/10810730.2012.727960] [Medline: 23301849]
10. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]

11. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
12. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
13. Misiak B, Szcześniak D, Koczanowicz L, Rymaszewska J. The COVID-19 outbreak and Google searches: is it really the time to worry about global mental health? Brain Behav Immun 2020 Jul;87:126-127 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.083] [Medline: 32360605]
14. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
15. Stanton R, To QG, Khalesi S, Williams SL, Alley SJ, Thwaite TL, et al. Depression, anxiety and stress during COVID-19: associations with changes in physical activity, sleep, tobacco and alcohol use in Australian adults. Int J Environ Res Public Health 2020 Jun 07;17(11):4065 [FREE Full text] [doi: 10.3390/ijerph17114065] [Medline: 32517294]
16. Choi EPH, Hui BPH, Wan EYF. Depression and anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740 [FREE Full text] [doi: 10.3390/ijerph17103740] [Medline: 32466251]
17. Table 702.60. Number and percentage of households with computer and internet access, by state: 2016. National Center for Education Statistics. 2016. URL: https://nces.ed.gov/programs/digest/d17/tables/dt17_702.60.asp [accessed 2020-09-02]
18. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
19. Shahid A, Wilkinson K, Marcu S, Shapiro C. Medical Outcomes Study Sleep Scale (MOS-SS). In: STOP, THAT and One Hundred Other Sleep Scales. New York Dordrecht Heidelberg London: Springer; 2011:219-222.
20. Blakey SM, Reuman L, Jacoby RJ, Abramowitz JS. Tracing "fearbola": psychological predictors of anxious responding to the threat of Ebola. Cognit Ther Res 2015;39(6):816-825 [FREE Full text] [doi: 10.1007/s10608-015-9701-9] [Medline: 32214559]
21. Stata Statistical Software: Release 14. StataCorp. 2015. URL: https://www.stata.com/company/ [accessed 2020-09-02]
22. Yip L, Bixler D, Brooks DE, Clarke KR, Datta SD, Dudley S, et al. Serious adverse health events, including death, associated with ingesting alcohol-based hand sanitizers containing methanol - Arizona and New Mexico, May-June 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1070-1073. [doi: 10.15585/mmwr.mm6932e1] [Medline: 32790662]
23. Kriplean T, Bonnar C, Borning A, Kinney B, Gill B. Integrating on-demand fact-checking with public dialogue. In: CSCW '14: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2014 Feb Presented at: CSCW'14: Computer Supported Cooperative Work; February, 2014; Baltimore Maryland USA p. 1188-1199 URL: https://dl.acm.org/doi/abs/10.1145/2531602.2531677 [doi: 10.1145/2531602.2531677]

Abbreviations

GAD-7: Generalized Anxiety Disorder 7
IRB: Institutional Review Board
MOS: Medical Outcomes Study

Edited by G Eysenbach, T Rashid Soron; submitted 24.06.20; peer-reviewed by L Lavorgna, K Aguirre, EPH Choi; comments to author 30.07.20; revised version received 18.08.20; accepted 21.08.20; published 10.09.20.

Please cite as:
Singh P, Cumberland WG, Ugarte D, Bruckner TA, Young SD
Association Between Generalized Anxiety Disorder Scores and Online Activity Among US Adults During the COVID-19 Pandemic: Cross-Sectional Analysis
J Med Internet Res 2020;22(9):e21490
URL: http://www.jmir.org/2020/9/e21490/
doi:10.2196/21490
PMID:32841152

Original Paper

Excess Patient Visits for Cough and Pulmonary Disease at a Large US Health System in the Months Prior to the COVID-19 Pandemic: Time-Series Analysis

Joann G Elmore[1], MD, MPH; Pin-Chieh Wang[1], PhD; Kathleen F Kerr[2], PhD; David L Schriger[3], MD, MPH; Douglas E Morrison[4], MS; Ron Brookmeyer[4], PhD; Michael A Pfeffer[1], MD; Thomas H Payne[5], MD; Judith S Currier[1], MD

[1]Department of Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, United States

[2]Department of Biostatistics, UW School of Public Health, Seattle, WA, United States

[3]Department of Emergency Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, United States

[4]Department of Biostatistics, Fielding School of Public Health, UCLA, Los Angeles, CA, United States

[5]Department of Medicine, UW School of Medicine, Seattle, WA, United States

Corresponding Author:
Joann G Elmore, MD, MPH
Department of Medicine
David Geffen School of Medicine
UCLA
1100 Glendon Ave
Suite 900
Los Angeles, CA, 90024
United States
Phone: 1 310 794 9505
Email: jelmore@mednet.ucla.edu

Abstract

Background: Accurately assessing the regional activity of diseases such as COVID-19 is important in guiding public health interventions. Leveraging electronic health records (EHRs) to monitor outpatient clinical encounters may lead to the identification of emerging outbreaks.

Objective: The aim of this study is to investigate whether excess visits where the word “cough” was present in the EHR reason for visit, and hospitalizations with acute respiratory failure were more frequent from December 2019 to February 2020 compared with the preceding 5 years.

Methods: A retrospective observational cohort was identified from a large US health system with 3 hospitals, over 180 clinics, and 2.5 million patient encounters annually. Data from patient encounters from July 1, 2014, to February 29, 2020, were included. Seasonal autoregressive integrated moving average (SARIMA) time-series models were used to evaluate if the observed winter 2019/2020 rates were higher than the forecast 95% prediction intervals. The estimated excess number of visits and hospitalizations in winter 2019/2020 were calculated compared to previous seasons.

Results: The percentage of patients presenting with an EHR reason for visit containing the word “cough” to clinics exceeded the 95% prediction interval the week of December 22, 2019, and was consistently above the 95% prediction interval all 10 weeks through the end of February 2020. Similar trends were noted for emergency department visits and hospitalizations starting December 22, 2019, where observed data exceeded the 95% prediction interval in 6 and 7 of the 10 weeks, respectively. The estimated excess over the 3-month 2019/2020 winter season, obtained by either subtracting the maximum or subtracting the average of the five previous seasons from the current season, was 1.6 or 2.0 excess visits for cough per 1000 outpatient visits, 11.0 or 19.2 excess visits for cough per 1000 emergency department visits, and 21.4 or 39.1 excess visits per 1000 hospitalizations with acute respiratory failure, respectively. The total numbers of excess cases above the 95% predicted forecast interval were 168 cases in the outpatient clinics, 56 cases for the emergency department, and 18 hospitalized with acute respiratory failure.

Conclusions: A significantly higher number of patients with respiratory complaints and diseases starting in late December 2019 and continuing through February 2020 suggests community spread of SARS-CoV-2 prior to established clinical awareness and

testing capabilities. This provides a case example of how health system analytics combined with EHR data can provide powerful and agile tools for identifying when future trends in patient populations are outside of the expected ranges.

(J Med Internet Res 2020;22(9):e21562) doi:10.2196/21562

KEYWORDS

COVID-19; pandemic; electronic health record; time-series analysis; prediction; forecast

Introduction

Health systems, medical providers, bioinformaticians, and researchers worldwide are working tirelessly to understand, contain, and ameliorate the COVID-19 pandemic. During this health emergency, clinicians have anecdotally noted an unusual number of patients with respiratory complaints at the end of 2019 and early 2020, well before COVID-19 was officially categorized by the World Health Organization (WHO) as a pandemic [1]. It is unclear whether such anecdotal reports are correct or the result of hindsight bias. If correct, the excess could represent typical variation in disease patterns. Alternatively, the excess, especially if it is significantly above prediction intervals based on historical data, could represent undetected and early COVID-19 cases prior to established clinical awareness and testing capabilities for the virus.

In the past decade, there has been widespread adoption of electronic health records (EHRs) in the United States. However, there have been limited efforts to date to leverage EHRs to support the delivery of high-value medical care or otherwise improve the delivery of health care services [2]. Using EHR data to model and forecast trends has the potential to improve resource management and the preparedness of health systems [3-7], which in turn could improve the quality of medical care. In particular, EHR data paired with analytical tools can potentially identify unusual trends in health care delivery that can alert clinicians and public health experts to critical changes in disease patterns.

The purpose of this paper is to use discrete raw EHR data to evaluate whether there was an excess of patients presenting with symptoms and diseases suggestive of COVID-19 in the months prior to the first known COVID-19 cases in the US health system in March 2020, using words found in chief complaint fields and International Classification of Diseases (ICD) codes from hospital discharge diagnoses. We analyzed 5 years of data from a large Los Angeles–area health system using time-series methods to address whether there was an excess number of patients presenting for complaints of cough, or hospitalizations for respiratory ailments. These methods highlight how health care analytics coupled with EHR data can be harnessed for disease surveillance. In particular, surveillance starting from the larger outpatient setting, which is often the tip of the iceberg, can provide an early warning of a public health emergency before patients fill hospital intensive care units and deaths accumulate.

Methods

The study included data from July 1, 2014, to February 29, 2020, from UCLA Health, a large health system with over 2.5 million total outpatient visits and 3 hospitals (UCLA Medical Center Santa Monica, Resnick Neuropsychiatric Hospital at UCLA, and Ronald Reagan UCLA Medical Center). Health system utilization data during the winter season, from December 1, 2019, to February 29, 2020, the months prior to increased public awareness of COVID-19 in the United States, were evaluated using the previous 5 years as the comparison period in a time-series analysis. Data were collected using SQL reports from Epic Clarity production databases supporting the EHRs used throughout UCLA Health.

Analyses included three different care settings: outpatient clinics, emergency departments, and hospital. All primary, specialty, and urgent care outpatient visits were considered and searched for the word “cough” within the reason for visit, further examining the percentages of patients presenting with cough in the current winter season with forecast predictions based on data for the preceding 5 years. All patient visits to emergency departments for cough and data on patients hospitalized with acute respiratory failure for the current winter season were separately examined with the corresponding data for recent years in the same manner (see Table 7 in Multimedia Appendix 1 for a list of ICD codes).

Seasonal autoregressive integrated moving average (SARIMA) models were applied on weekly data in SAS (SAS Institute) from July 1, 2014, through November 30, 2019, to forecast data for December 2019 through February 2020 (Tables 1-6 and Figures 1-3 in Multimedia Appendix 1). These models take into account seasonal effects [8,9] and use maximum likelihood to estimate model parameters based on historical data from July 1, 2014, to November 30, 2019. A winter season is defined as the time period from December 1 to the last day of February the subsequent year. Using the SARIMA model, a forecast of the 2019/2020 winter season was provided. Specifically, a SARIMA$(1,0,1)x(1,0,1)_{52}$ model was used to analyze the outpatient time series, a SARIMA$(1,0,3)x(1,0,1)_{52}$ model was used to analyze the emergency department data, and a SARIMA$(1,0,1)x(1,0,1)_{52}$ model was used to analyze the inpatient data for patients with acute respiratory failure. The autoregressive and moving-average orders were identified by both SCAN and extended sample autocorrelation function (ESACF) methods [10].

The 95% prediction intervals for the forecast allowed an assessment of whether the observed data at weekly intervals for the 2019/2020 winter season were outside of the time-series prediction interval.

For example, for the outpatient clinic visit data, the following SARIMA$(1,0,1)x(1,0,1)_{52}$ model was used (model parameter estimates are shown in Table 1 in Multimedia Appendix 1):

Where X_t is the percentage of visits in week for which cough was a recorded symptom; B is the backshift operator, $BX_t = X_{t-1}$;  is a white noise process of uncorrelated random variables with mean 0 and variance  ; and Z_t is an indicator variable for the months of December to February.

Excess cases per 1000 visits for the 2019/2020 winter season compared to previous seasons were estimated using three methods (Tables 8-11 in Multimedia Appendix 1). First, for a conservative estimate of excess cases, the maximum of the five previous seasons was subtracted from the current season. Second, the average of the five previous seasons was subtracted from the current season. For both methods, the excess percentage was multiplied by the total number of patient visits in the current season to estimate the excess cases for each month. For the third method, the upper limit of the time-series 95% prediction interval was subtracted from the observed rate for each of the 13 weeks in the 2019/2020 season to estimate the weekly excess percentages. The weekly excess percentage was multiplied by the weekly patient visits and aggregated to estimate the excess cases.

To visualize the data, the local regression (LOESS) technique was used to smooth daily data with a smoothing span of 20% (see Figure 4 in Multimedia Appendix 1 for scatter plots of daily data) [11], with analyses performed in R (R Foundation for Statistical Computing) [12] and figures generated in Microsoft Excel (Microsoft Corp).

Two sensitivity analyses were performed. First, patient insurance status was considered to see if variation in insurance could explain trends over time in the outpatient and emergency department data; the analysis was repeated using only those outpatient clinics that existed for all years of the study period to ensure that changing case mix did not confound our results. Second, patients hospitalized with a broader set of respiratory illnesses were investigated: patients hospitalized with ICD codes used in a study of respiratory tract illnesses associated with influenza [13], and those with any pneumonia (see Table 7 in Multimedia Appendix 1 for a list of ICD codes). A SARIMA$(2,0,2)x(1,0,1)_{52}$ model was used to analyze the inpatient data for "any respiratory tract disease," and a SARIMA$(2,0,1)x(1,0,1)_{52}$ model was used to analyze the inpatient data for "pneumonia." Institutional Review Board approval was obtained (UCLA number 20-000528).

Results

Outpatient Clinic Data

The data encompass 9,501,091 outpatient clinic visits, with the average number of clinic visits increasing over time (eg, there were 314,832 visits from December 1, 2014, to February 28, 2015, and 511,687 visits during the 2019/2020 winter season). The expected cyclical increase in patients presenting with reports of cough each winter is observed for all 6 years studied (Figure 1A; Table 1; Figure 1A in Multimedia Appendix 1). The percentage of patients presenting for complaint of a cough was within the prediction intervals in early and mid-December 2019. Starting the week of December 22, 2019, the data exceeded the 95% prediction interval and consistently exceeded the 95% prediction interval each week through the end of February 2020 (Figure 2).

The estimated number of total excess visits for cough over the three winter months of 2019-2020 was 739 (1.6/1000 visits) when compared with the highest historical monthly value and 1047 (2.0/1000 visits) when compared with the average monthly value for all 5 years of historical controls. There were 168 excess visits above the 95% prediction interval forecast according to the time-series analysis (Table 2; Table 8 in Multimedia Appendix 1).

Figure 1. Percentage of outpatient and emergency department visits for cough and hospitalizations for acute respiratory failure from July 1, 2014, to February 29, 2020. Vertical reference lines align to the first day of each month. (A) Outpatient clinic visits for complaints of cough. (B) Emergency department visits for cough. (C) Hospitalizations for acute respiratory failure.

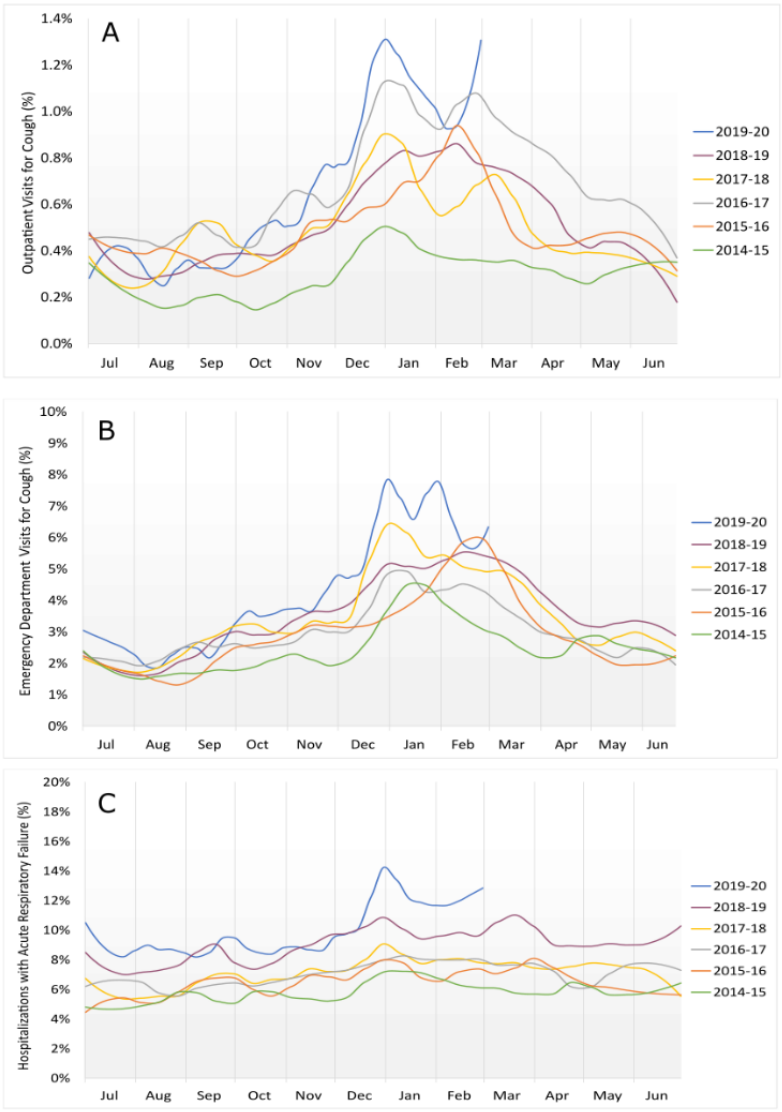


Table 1. Outpatient and emergency department visits for cough and hospitalizations for acute respiratory failure by years of winter season (December to February).

Calendar year	Outpatient			Emergency department			Hospitalization		
	Total Outpatient visits	Cough Number of patients (%)	Cough Cases per 1000	Total ED[a] visits	Cough Number of patients (%)	Cough Cases per 1000	Total Hospital-izations	Acute respiratory failure Number of patients (%)	Cases per 1000
2014-2015	314,832	929 (0.30)	3.0	24,127	853 (3.54)	35.4	11,016	680 (6.17)	61.7
2015-2016	391,089	1499 (0.38)	3.8	25,977	1134 (4.37)	43.7	10,925	760 (6.96)	69.6
2016-2017	405,620	1671 (0.41)	4.1	25,505	1072 (4.20)	42.0	10,831	827 (7.64)	76.4
2017-2018	425,686	1670 (0.39)	3.9	27,022	1429 (5.29)	52.9	10,640	830 (7.80)	78.0
2018-2019	446,673	1635 (0.37)	3.7	25,555	1263 (4.94)	49.4	10,646	996 (9.36)	93.6
2019-2020	511,687	2938 (0.57)	5.7	26,748	1708 (6.39)	63.9	9903	1138 (11.49)	114.9

[a]ED: emergency department.

Figure 2. Time series analysis of outpatient data. (A) Forecast, with 95% prediction intervals, of expected rates of outpatient presentations for cough (based on time-series analyses of the previous 5 years), with observed data for each week shown for December 1, 2019, to February 29, 2020. (B) Time-series analysis of outpatient data.

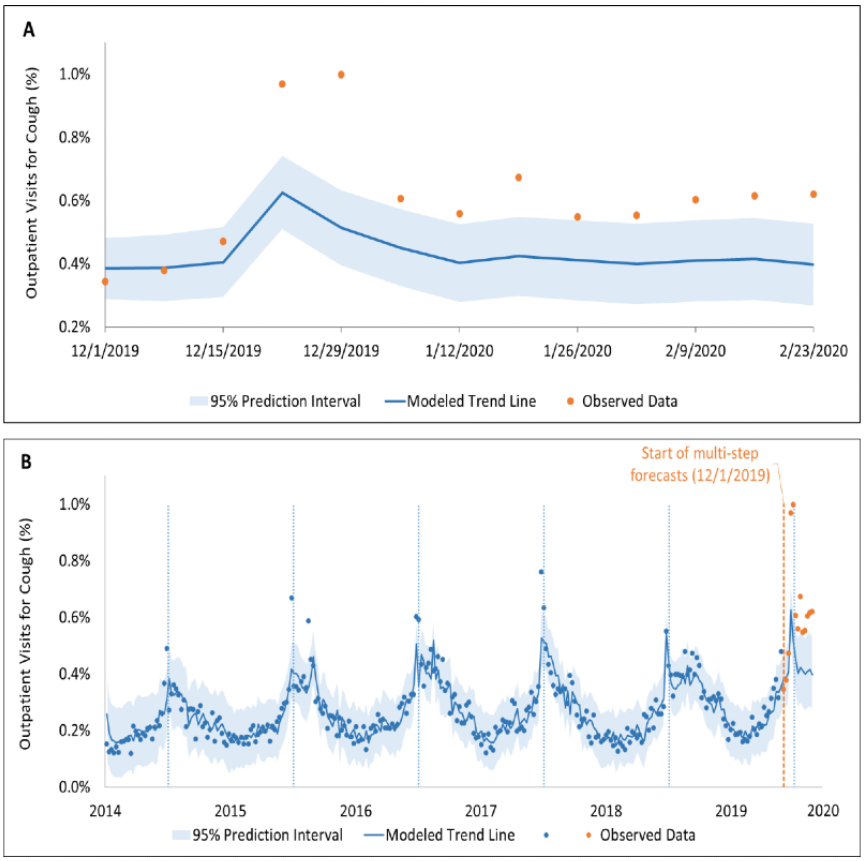


Table 2. Estimated excess cases during the 2019/2020 winter season via different methods.

Estimated excesses	Outpatient (cough)	Emergency department (cough)	Hospitalization (acute respiratory failure)
Estimated excess cases per 1000 visits[a]	1.6, 2.0	11.0, 19.2	21.4, 39.1
Estimated excess cases[a]	739, 1047	229, 514	210, 387
Total number of weeks the observed data was above the 95% prediction interval	10/13 (100% of weeks after December 22, 2019)	6/13 (60% of weeks after December 22, 2019)	7/13 (70% of weeks after December 22, 2019)
Excess cases above the 95% prediction interval[b]	168	56	18

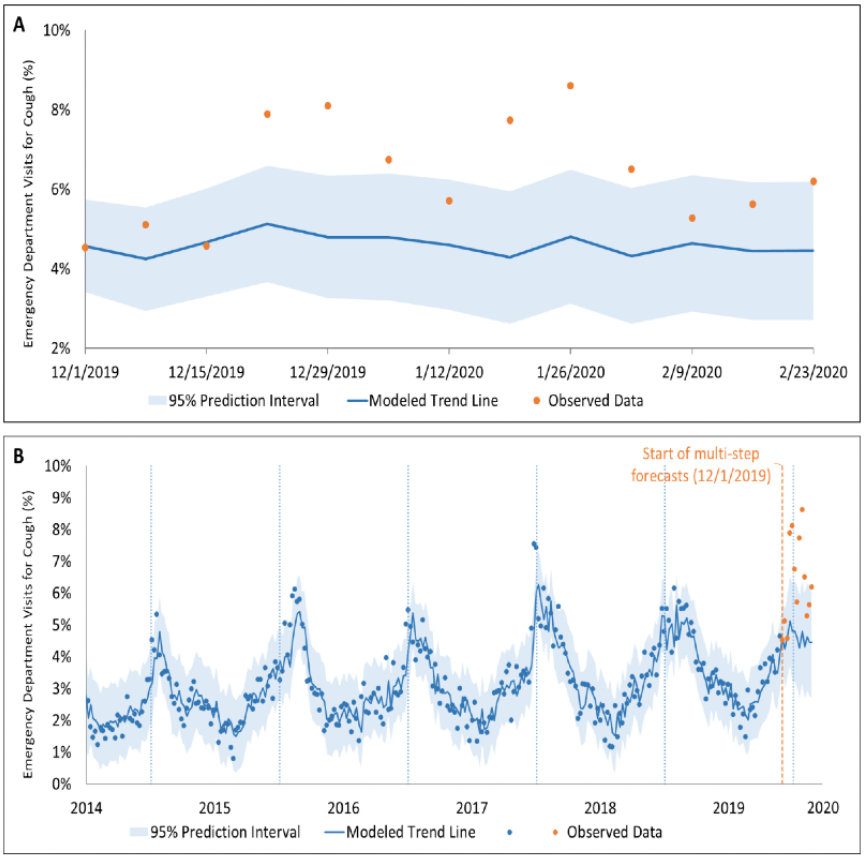
[a]Two methods were used to estimate the excess cases per 1000 visits and excess cases in the 2019/2020 winter season (December 2019, January 2020, and February 2020) compared to previous seasons. First, for a conservative estimate of excess cases per 1000 visits, using percentages we subtracted the maximum of the five previous seasons with the current season. Second, we subtracted the average of the five previous seasons from the current season. For both methods, we multiplied the excess percentage by the total number of patient visits in the current season to estimate the excess cases.

[b]For the third method, we subtracted the upper limit of the time-series 95% prediction interval from the observed rate for each of the 13 weeks in the current season to estimate the weekly excess percentages (data not shown in table). We multiplied the weekly excess percentage by the weekly patient visits and aggregated these numbers to estimate the excess cases.

Emergency Department Visits

The emergency department data encompass 574,813 visits from July 1, 2014, to February 29, 2020, with an average of 25,822 visits per winter season. Similar to outpatient visits for cough, seasonal variation in the proportion of emergency department visits for cough was observed (Figure 1B, Table 1). An excess above the time-series 95% prediction interval was noted starting December 22, 2019; in total, 6 of the 10 weeks exceeded the 95% prediction interval (Figure 3). The estimated number of total excess in patient visits to the emergency departments for cough over the 2019/2020 winter season using the three methods was 229 (11.0/1000 visits), 514 (19.2/1000 visits), and 56, respectively (Table 2; Table 9 in Multimedia Appendix 1).

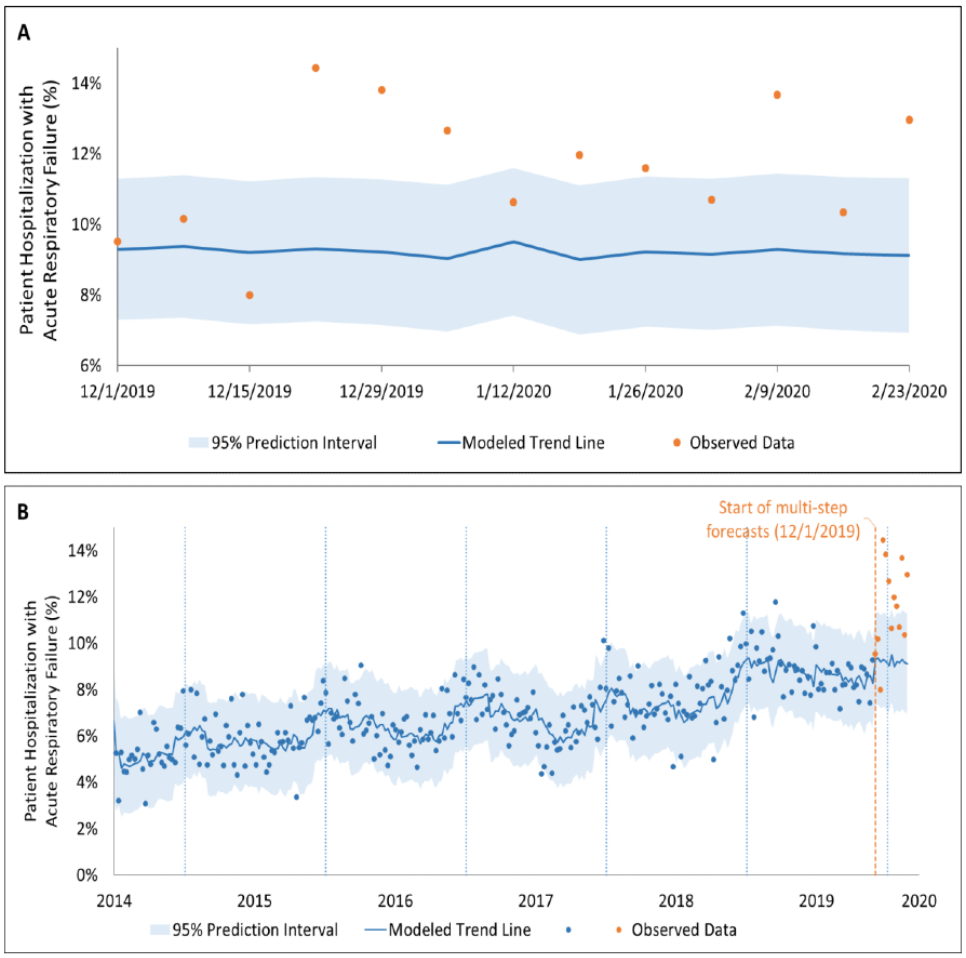
Figure 3. Time-series analysis of emergency department visits for cough. (A) Forecast, with 95% prediction intervals, of expected rates of emergency department presentations for cough (based on time-series analyses of the previous 5 years), with observed data for each week shown for December 1, 2019, to February 29, 2020. (B) Time-series analysis of emergency department data.



Hospitalized Patients

There were 247,774 patients admitted to the three hospitals included in this study from July 1, 2014, to February 29, 2020, with an average of 10,660 admissions each winter season. The percentage of patients with a discharge diagnosis of acute respiratory failure were higher in December 2019, January 2020, and February 2020 when compared with all 5 historical control years (Figure 1C, Table 1). The observed percentage of patients who had acute respiratory failure during the subsequent hospitalization exceeded the time-series 95% prediction interval for patients admitted starting the week of December 22, 2019; in total, 7 of the 10 weeks of observed data were above the 95% prediction interval (Figure 4). Using the three prediction methods, the estimated excess numbers of patients hospitalized with acute respiratory failure were 210 (21.4/1000 visits), 387 (39.1/1000 visits), and 18, respectively (Table 2; Table 10 in Multimedia Appendix 1).

Figure 4. Time-series analyses of hospitalizations for acute respiratory failure. (A) Forecast, with 95% prediction intervals, of expected rates of patients hospitalized with acute respiratory failure (based on time-series analyses of the previous 5 years), with observed data for each week shown for December 1, 2019, to February 29, 2020. (B) Time-series analysis of hospitalization data.



Sensitivity Analyses

The above findings were qualitatively unchanged when clinic and emergency department visits were analyzed by insurance class and clinic visits were restricted to clinics that treated patients in all years of the study period (Figures 5-7 in Multimedia Appendix 1). The percentage of patients with a discharge diagnosis of any respiratory tract disease or of any pneumonia were also higher in December 2019, January 2020, and February 2020 when compared with prior years (Table 11 in Multimedia Appendix 1).

Discussion

A significantly higher number of patients presented to outpatient clinics and emergency departments in this health system with a complaint of cough starting the last week of December 2019 and continuing through January and February 2020. These findings translate into hundreds of additional patients seeking outpatient medical attention in this health system for the respiratory symptom of cough during the 2019/2020 winter season. A significant excess in the number of patients hospitalized with acute respiratory failure during this same time period was also noted. It is possible that some of this excess represents early COVID-19 disease before clinical recognition and testing, information that may help epidemiologists better understand the spread of this pandemic. If only some of these excess visits are due to COVID-19, this could still represent community spread of SARS-CoV-2 during that time because a substantial proportion of individuals infected have no symptoms or mild symptoms and do not seek medical care [14,15], making cryptic spread of the disease within a community likely.

EHRs are widely adopted but have not been used to their fullest potential to deliver high-value care [16-18]. This work demonstrates the potential of using EHR data for symptom or disease surveillance. A strength of this study is the use of raw EHR data that are already collected in most health care systems to determine whether patients were presenting with reports of "cough" at excess numbers in the months before the first known case of COVID-19 in this health system. In addition, this study considers data from three separate locations of patient care contained in the EHR: the outpatient setting, the emergency department, and the hospital. For many diseases, data from the outpatient setting can provide an early warning to emergency departments and hospital intensive care units of what is to come. By leveraging time-series analysis to calculate prediction intervals of expected patients, outlier numbers of patient visits can be quickly identified.

While asymptomatic transmission and community spread of COVID-19 are possible explanations for the observed excess

patient encounters, other reasons and limitations need to be considered. The study was performed in a single health system and we only searched for the word "cough." Although the term "cough" is possibly more specific to COVID-19 than other symptoms such as "fever" or "aches," this search method has imperfect specificity and sensitivity as it does not include the full spectrum of COVID-19 symptoms [19-25]. The health system patient mix could have changed over the study period, but sensitivity analyses did not find evidence that this affected the findings. It is possible that the findings are due to lung injury from e-cigarettes (vaping), but this explanation is doubtful because the Centers for Disease Control and Prevention reported a continued decline after September 2019 [26].

Another limitation is not knowing for certain whether and what percentage of excess patient visits were due to influenza. An increase in influenza-positive test results and emergency department visits for influenza-like illnesses was noted in Los Angeles County and the United States during the 2019/2020 winter season when compared with prior years [27,28]. The incidence of influenza-like illness symptoms in the United States peaked earlier in 2019 when compared with previous years [28], with the 2019/2020 season peak at a similar level as the 2017/2018 season peak. This suggests that some of the observed incidence was due to influenza rather than COVID-19. However, the analysis shows an excess of patients above the forecast 95% prediction interval based on five previous seasons.

It is plausible that some of the excess visits might be due to SARS-CoV-2 as studies of rapid sentinel surveillance and genome sequencing suggest community transmission of SARS-CoV-2 much earlier than initially thought. Studies in the United States and France found evidence for cryptic spread of the virus as early as December 2019 or January 2020, before community surveillance was actively implemented [29-32]. Especially early in outbreaks, existing methods for case identification may not capture incident infections; thus, novel and complementary methods using EHR data such as those reported here may play an important role.

Heightened media attention regarding the coronavirus pandemic could influence patients to seek medical care for cough-related concerns and should be considered [33]. Such data can complement these EHR data approaches. Using data from Google Trends [34], the popularity of "cough" as a search term in the United States was slightly higher from late December 2019 through January 2020 when compared with the average national popularity in the previous 4 years (Figure 8 in Multimedia Appendix 1). Popularity of searches for "cough" increased substantially in mid-March, corresponding to the sharp increase in "coronavirus" Google news searches. Although there was almost no mention of a COVID-19–type illness by United States media in December 2019 and little mention in January 2020, substantial media attention was present in February 2020. Therefore, this limitation is pronounced for February 2020, but less concerning for the earlier months.

In summary, health system analytics combined with EHR data can be harnessed to quickly identify changes in underlying patient populations. This study identified a significant excess of patients with COVID-19–like presentations starting the last week of December 2019 and continuing through February 2020, a time period before the availability of testing or providers considered clinical diagnoses of COVID-19. A unique feature of this study is the evaluation of three different stages of health care settings, which expands surveillance beyond just reporting the number of patients in emergency departments or hospital intensive care unit beds to include consideration of 9.5 million outpatient clinic records. The Centers for Disease Control and Prevention Outpatient Influenza-like Illness Surveillance Network (ILINet) currently monitors patients presenting with fever and a cough and/or a sore throat in the United States [35]. Data from the outpatient clinic setting is usually a harbinger of what is to come for hospital emergency departments and intensive care units.

Harnessing larger electronic health data systems to monitor outpatient visits for the growing and diverse set of symptoms associated with COVID-19 should be considered [21-25].

This SARS-CoV-2 pandemic highlights the urgent need to support the development of agile health care analytics that enable real-time symptom and disease surveillance [36]. Lessons learned from this pandemic will hopefully lead to better preparation and the ability to quickly provide warnings and track the next pandemic.

Acknowledgments

We appreciate the University of California, Los Angeles Institutional Review Board rapid review and the care provided by the Health System personnel during the COVID-19 pandemic.

Conflicts of Interest

JGE serves as Editor-in-Chief for Adult Primary Care topics at UpToDate. All other authors declare no conflicts of interest.

Multimedia Appendix 1
All supplementary figures and tables.
[DOCX File , 1706 KB - jmir_v22i9e21562_app1.docx]

References

1. Parodi E, Aloisi S. Italian scientists investigate possible earlier emergence of coronavirus. Reuters. 2020 Mar 26. URL: https://www.reuters.com/article/us-health-coronavirus-italy-timing/italian-scientists-investigate-possible-earlier-emergence-of-coronavirus-idUSKBN21D2IG [accessed 2020-04-09]
2. Rudin RS, Friedberg MW, Shekelle P, Shah N, Bates DW. Getting Value From Electronic Health Records: Research Needed to Improve Practice. Ann Intern Med 2020 Jun 02;172(11_Supplement):S130-S136 [FREE Full text] [doi: 10.7326/M19-0878] [Medline: 32479182]
3. Parker CA, Liu N, Wu SX, Shen Y, Lam SSW, Ong MEH. Predicting hospital admission at the emergency department triage: A novel prediction model. Am J Emerg Med 2019 Aug;37(8):1498-1504 [FREE Full text] [doi: 10.1016/j.ajem.2018.10.060] [Medline: 30413365]
4. Kadri F, Harrou F, Chaabane S, Tahon C. Time series modelling and forecasting of emergency department overcrowding. J Med Syst 2014 Sep;38(9):107 [FREE Full text] [doi: 10.1007/s10916-014-0107-0] [Medline: 25053208]
5. Zhu T, Luo L, Zhang X, Shi Y, Shen W. Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients. IEEE J Biomed Health Inform 2017 Mar;21(2):515-526. [doi: 10.1109/JBHI.2015.2511820] [Medline: 28055928]
6. Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. BMC Health Serv Res 2017 Jul 10;17(1):469 [FREE Full text] [doi: 10.1186/s12913-017-2407-9] [Medline: 28693579]
7. Zhou L, Zhao P, Wu D, Cheng C, Huang H. Time series model for forecasting the number of new admission inpatients. BMC Med Inform Decis Mak 2018 Jun 15;18(1):39 [FREE Full text] [doi: 10.1186/s12911-018-0616-8] [Medline: 29907102]
8. Box GEP, Tiao GC. Intervention Analysis with Applications to Economic and Environmental Problems. Journal of the American Statistical Association 1975 Mar;70(349):70-79. [doi: 10.1080/01621459.1975.10480264]
9. Shumway R, Stoffer D. Time Series Analysis and Its Applications With R Examples. 4th ed. Pittsburgh, PA, USA: Springer; 2017. URL: https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf [accessed 2020-08-26]
10. Stadnytska T, Braun S, Werner J. Model Identification of Integrated ARMA Processes. Multivariate Behav Res 2008 Mar 19;43(1):1-28 [FREE Full text] [doi: 10.1080/00273170701836646] [Medline: 26788970]
11. Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association 1988 Sep;83(403):596-610. [doi: 10.1080/01621459.1988.10478639]
12. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio Inc; 2020. URL: http://www.rstudio.com [accessed 2020-08-26]
13. Chow EJ, Rolfes MA, O'Halloran A, Alden NB, Anderson EJ, Bennett NM, et al. Respiratory and Nonrespiratory Diagnoses Associated With Influenza in Hospitalized Adults. JAMA Netw Open 2020 Mar 02;3(3):e201323 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.1323] [Medline: 32196103]
14. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, Public Health – Seattle & King County, CDC COVID-19 Investigation Team. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility - King County, Washington, March 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 03;69(13):377-381 [FREE Full text] [doi: 10.15585/mmwr.mm6913e1] [Medline: 32240128]
15. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971. [doi: 10.1056/NEJMc2001468] [Medline: 32003551]
16. Ye Q, Zhou J, Wu H. Using Information Technology to Manage the COVID-19 Pandemic: Development of a Technical Framework Based on Practical Experience in China. JMIR Med Inform 2020 Jun 08;8(6):e19515 [FREE Full text] [doi: 10.2196/19515] [Medline: 32479411]
17. Powell J, Buchan I. Electronic health records should support clinical research. J Med Internet Res 2005;7(1):e4 [FREE Full text] [doi: 10.2196/jmir.7.1.e4] [Medline: 15829476]
18. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
19. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
20. Centers for Disease Control and Prevention. Symptoms of coronavirus. Coronavirus (COVID-19). 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html [accessed 2020-05-04]
21. Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. Am J Gastroenterol 2020 May;115(5):766-773 [FREE Full text] [doi: 10.14309/ajg.0000000000000620] [Medline: 32287140]
22. Wu P, Duan F, Luo C. Characteristics of Ocular Findings of Patients With Coronavirus Disease 2019 (COVID-19) in Hubei Province, China. JAMA Ophthalmology 2020;138(5):575-578. [Medline: 32232433]
23. Galván Casas C, Català A, Carretero Hernández G, Rodríguez-Jiménez P, Fernández-Nieto D, Rodríguez-Villa Lario A, et al. Classification of the cutaneous manifestations of COVID-19: a rapid prospective nationwide consensus study in Spain with 375 cases. Br J Dermatol 2020 Jul;183(1):71-77 [FREE Full text] [doi: 10.1111/bjd.19163] [Medline: 32348545]

24. Han C, Duan C, Zhang S, Spiegel B, Shi H, Wang W, et al. Digestive Symptoms in COVID-19 Patients With Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes. Am J Gastroenterol 2020 Jun;115(6):916-923 [FREE Full text] [doi: 10.14309/ajg.0000000000000664] [Medline: 32301761]
25. Printza A, Constantinidis J. The role of self-reported smell and taste disorders in suspected COVID-19. Eur Arch Otorhinolaryngol 2020 Sep;277(9):2625-2630 [FREE Full text] [doi: 10.1007/s00405-020-06069-6] [Medline: 32447496]
26. Centers for Disease Control and Prevention, Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion. Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products. 2020 Feb 25. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html [accessed 2020-03-27]
27. LA County Department of Public Health. Influenza Watch: Influenza and Related Disease Updates for Los Angeles County. URL: http://publichealth.lacounty.gov/acd/docs/Flu/Season19-20/Week12.pdf
28. Centers for Disease Control and Prevention. 2019-2020 Influenza Season Week 21 ending May 23, 2020. 2020. URL: https://www.cdc.gov/flu/weekly/index.htm [accessed 2020-05-31]
29. Zwald ML, Lin W, Sondermeyer Cooksey GL, Weiss C, Suarez A, Fischer M, et al. Rapid Sentinel Surveillance for COVID-19 - Santa Clara County, California, March 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 10;69(14):419-421 [FREE Full text] [doi: 10.15585/mmwr.mm6914e3] [Medline: 32271724]
30. After Retesting Samples French Hospital Discovers Covid-19 Case From December. Reuters. 2020. URL: https://www.reuters.com/article/us-health-coronavirus-france-idUSKBN22G20L [accessed 2020-05-04]
31. Spellberg B, Haddix M, Lee R, Butler-Wu S, Holtom P, Yee H, et al. Community Prevalence of SARS-CoV-2 Among Patients With Influenzalike Illnesses Presenting to a Los Angeles Medical Center in March 2020. JAMA 2020 Mar 31:e1-e2 [FREE Full text] [doi: 10.1001/jama.2020.4958] [Medline: 32232421]
32. Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang M, et al. Cryptic transmission of SARS-CoV-2 in Washington State. medRxiv 2020 Apr 06:1. [doi: 10.1101/2020.04.02.20051417] [Medline: 32511596]
33. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014 [FREE Full text] [doi: 10.1038/nature07634] [Medline: 19020500]
34. Google Trends. URL: https://trends.google.com/trends/?geo=US [accessed 2020-03-26]
35. Centers for Disease Control and Prevention. US Influenza Surveillance System. 2020. URL: https://www.cdc.gov/flu/weekly/overview.htm#Outpatient [accessed 2020-05-31]
36. Schulz WL, Durant TJS, Torre CJ, Hsiao AL, Krumholz HM. Agile Health Care Analytics: Enabling Real-Time Disease Surveillance With a Computational Health Platform. J Med Internet Res 2020 May 28;22(5):e18707 [FREE Full text] [doi: 10.2196/18707] [Medline: 32442130]

Abbreviations

EHR: electronic health record
ICD: International Classification of Diseases
SARIMA: seasonal autoregressive integrated moving average
SCAN: smallest canonical correlation
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.06.20; peer-reviewed by SP Tu, C Geller-Bernstein; comments to author 24.07.20; revised version received 31.07.20; accepted 03.08.20; published 10.09.20.

Please cite as:
Elmore JG, Wang PC, Kerr KF, Schriger DL, Morrison DE, Brookmeyer R, Pfeffer MA, Payne TH, Currier JS
Excess Patient Visits for Cough and Pulmonary Disease at a Large US Health System in the Months Prior to the COVID-19 Pandemic: Time-Series Analysis
J Med Internet Res 2020;22(9):e21562
URL: https://www.jmir.org/2020/9/e21562
doi:10.2196/21562
PMID:32791492

Original Paper

Association of Socioeconomic Changes due to the COVID-19 Pandemic With Health Outcomes in Patients With Skin Diseases: Cross-Sectional Survey Study

Yeye Guo[1,2,3], MD; Minxue Shen[1,2,3,4], MD, PhD; Xu Zhang[1,2,3], PhD; Yi Xiao[1,2,3], MD, MPH, PhD; Shuang Zhao[1,2,3], MD, PhD; Mingzhu Yin[1,2,3], MD, PhD; Wenbo Bu[5], MD, PhD; Yan Wang[5], MD, PhD; Xiang Chen[1,2,3], MD, PhD; Juan Su[1,2,3], MD, PhD

[1]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China

[2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Changsha, China

[3]Hunan Engineering Research Center of Skin Health and Disease, Hunan Key Laboratory of Skin Cancer and Psoriasis, Xiangya Hospital, Central South University, Changsha, China

[4]Department of Social Medicine and Health Management, Xiangya School of Public Health, Central South University, Changsha, China

[5]Institute of Dermatology, Chinese Academy of Medical Science and Peking Union Medical College, Nanjing, China

Corresponding Author:
Juan Su, MD, PhD
Department of Dermatology
Xiangya Hospital
Central South University
Xiangya Road #87
Changsha, 410008
China
Phone: 86 731 84327377
Email: sujuanderm@csu.edu.cn

Abstract

Background: The outbreak of COVID-19 has profoundly influenced people's lifestyles; these impacts have varied across subgroups of people. The pandemic-related impacts on the health outcomes of people with dermatological conditions are unknown.

Objective: The aim of this paper was to study the association of COVID-19 pandemic–related impacts with health-related quality of life in patients with skin diseases.

Methods: This was a cross-sectional study among Chinese patients with skin diseases. A self-administered web-based questionnaire was distributed through social media. Demographic and clinical data and pandemic-related impacts (isolation status, income changes, and employment status) were collected. The main outcomes included perceived stress (Visual Analog Scale), symptoms of anxiety (Generalized Anxiety Disorder-7) and depression (9-Item Patient Health Questionnaire), quality of life (Dermatology Life Quality Index), and health utility mapping based on the EQ-5D-3L descriptive system. Multivariable logistic regression was used to investigate the associations.

Results: A total of 506 patients with skin diseases completed the survey. The mean age of the patients was 33.5 years (SD 14.0), and 217/506 patients (42.9%) were male. Among the 506 respondents, 128 (25.3%) were quarantined, 102 (20.2%) reported unemployment, and 317 (62.6%) reported decrease or loss of income since the pandemic. The pandemic-related impacts were significantly associated with impaired mental well-being and quality of life with different effects. Unemployment and complete loss of income were associated with the highest risks of adverse outcomes, with increases of 110% to 162% in the prevalence of anxiety, depression, and impaired quality of life.

Conclusions: Isolation, income loss, and unemployment are associated with impaired health-related quality of life in patients with skin diseases during the COVID-19 pandemic.

(J Med Internet Res 2020;22(9):e22288) doi:10.2196/22288

KEYWORDS

skin diseases; coronavirus disease 2019; unemployment; quality of life; web-based; survey; dermatology; COVID-19; lifestyle; impact; outcome; isolation

Introduction

The outbreak of COVID-19 has resulted in the infection of over 4,000,000 people in 216 countries and regions worldwide as of May 17, 2020 [1]. A global response of social distancing was undertaken to control the transmission of the disease, and people started voluntarily isolating themselves at home. The Chinese government suggested that the public quarantine at home in January 2020. In the following months, many other countries, such as the United Kingdom, Italy, and the United States, called for self-isolation [2,3]. These actions effectively reduced the rapid spread of the disease but unavoidably had a substantial impact on the global economy. During this time, great socioeconomic changes occurred, such as reduced workforce and increasing unemployment [4]. According to data from the US Labor Department, the unemployment rate soared to 14.7% in April 2020 [5]. According to a survey in the United States, up to 43.4% of adults reported that their families suffered a job or income loss during the pandemic [6].

In addition to the socioeconomic implications of the pandemic, the health status of the public is a critical issue that should not be ignored. It is well-known that isolation itself is a risk factor for many mental health issues, including suicide and self-harm [2]. Therefore, long-term isolation in conjunction with loss of employment or income may produce adverse effects on individuals' mental well-being and quality of life [7], especially for people who have chronic diseases.

Skin disorders are among the most prevalent human diseases; they affect 30% to 70% of individuals [8]. Skin diseases may cause disfigurement and cutaneous symptoms, and they may result in considerable discomfort and disability. Although most skin diseases are not life-threatening, many of these diseases, such as psoriasis and eczema, are chronic and recurrent. Additionally, itch and fatigue are common symptoms reported by patients with skin diseases, and some also report experiencing pain [9]. Consequently, skin diseases cause not only physical but also psychological discomfort as well as impaired quality of life [10]. Additionally, skin diseases create financial burdens on families and society. A study in the United States suggested that skin diseases cost an estimated $39.3 billion per year [11]. As a result, isolation, unemployment, and loss of income related to the COVID-19 pandemic may place patients with skin diseases at additional health risk.

To date, no study has been published regarding the potential influence of socioeconomic changes on the health status of patients with skin diseases during the COVID-19 pandemic. In this study, we investigated the association of COVID-19–related impacts, including isolation, unemployment, and income loss, with health-related quality of life in patients with skin diseases, including perceived stress, symptoms of anxiety and depression, and quality of life during the pandemic.

Methods

Study Design and Participants

We performed a cross-sectional study among Chinese patients with skin diseases. A web-based survey link was created and posted on social media platforms (WeChat groups and teledermatology platforms) to facilitate the collection of questionnaires. To avoid repeat submissions, each single IP address was allowed to submit answers only one time. Completion of all questions was required before submission of the questionnaire. All the participants were allowed to quit at any time if any question made them feel uncomfortable. The survey was conducted from April 15 to 27, 2020. The study was reviewed and approved by the institutional research ethics board of Xiangya Hospital, Central South University (Changsha, China; approval number: 202002024). Electronic informed consent was collected from all participants.

Exposure Variables

Three exposure variables were defined. Outdoor activity restriction was determined by a single question: "During the last month, what measures did you take for isolation?" with the following four responses: "I was not isolated, and my outdoor activity was unaffected," "I was not isolated, but my outdoor activity was partly affected," "I was isolated at home and receiving medical observation," and "I was quarantined in hospital and receiving medical observation or treatment." Because only one patient reported having COVID-19 and being quarantined in hospital, patients who were quarantined at home or in the hospital were combined into one group in the analysis.

The employment status was determined by a single question: "How is your employment status since the epidemic?" with the following responses: "I am unemployed since the epidemic" and "My employment status has been unaffected since the epidemic."

Income change was measured by a single question: "Has your monthly income changed since the epidemic of COVID-19?" with the following responses: "I have completely lost my income," "My monthly income has decreased," "My monthly income has been unaffected," and "My monthly income has increased." Because only two patients reported increased income, they were categorized in the "unaffected" group in the analysis.

Patient-Reported Outcomes

The primary outcomes were perceived stress and depression and anxiety, which were measured using some short, simple scales to minimize respondent burden.

The Visual Analog Scale (VAS) was used to assess perceived stress during the past two weeks. The area under the receiver operating characteristic curve of the VAS was 0.9 to 0.93, with a cutoff of 6.8 to 7.2 according to the Perceived Stress Scale-14

(PSS-14) [12,13]. Here, we used a cutoff of ≥7 to define significant perceived stress.

The Generalized Anxiety Disorder-7 (GAD-7) and the 9-item Patient Health Questionnaire (PHQ-9) were applied to examine the symptoms of anxiety and depression during the past two weeks. The cutoff point for both scales was ≥8 [14,15]. The Cronbach α coefficients of the GAD-7 and PHQ-9 in our sample were .93 and .89, respectively.

The Dermatology Life Quality Index (DLQI) was used to assess the respondents' quality of life during the past two weeks. The cutoff point was ≥10 in our study [16]. The Cronbach α coefficient of the DLQI in our sample was .76. Health utility was mapped using the generic tool EQ-5D-3L according to the method proposed by Liu et al [17].

Covariates

Demographic and clinical data of the individuals were collected and analyzed as covariates, including gender (male or female), age, marital status (unmarried, married, divorced, or widowed), educational level (primary school and below, middle school, high school, or college and above), annual income (<¥10,000, ¥10,000 to ¥49,999, ¥50,000 to ¥99,999, or ≥¥100,000, equivalent to US $1,464.47, $1,464.4 to $7,322.21, $7,322.36 to $14,644.57, or ≥$14,644.72), type of skin diseases (infectious skin diseases, papulosquamous disorders, allergic skin diseases, disorders of appendages, pigmentary disorders, skin tumors), course of disease (<1 year, 1 to 5 years, 6 to 10 years, >10 years), adherence to treatment, and use of health care services.

Statistical Analyses

The data were exported from the web-based survey system and analyzed with R version 3.4 (R Project). Continuous variables with normal distribution were expressed as mean (SD) and compared with analysis of variance. Continuous data with skewed distributions were presented as median (IQR) and compared with the Wilcoxon rank sum test. Categorical variables were summarized as counts (percentages) and compared using the chi-square test or Fisher exact test. The interaction effects of the exposure variables were examined. The effect sizes of the associations were presented as adjusted odds ratios (aORs) and 95% CIs. A *P* value <.05 was considered statistically significant. Reporting of the results followed the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) guidelines.

Data Accessibility

Data are available upon request from JS.

Results

In total, 506 valid questionnaires were collected and analyzed. No patient reported confirmed infection with COVID-19. The average time to complete the survey was 7.2 minutes (IQR 4.1-8.2). The mean age of the patients was 33.5 years (SD 14.0), and 217/506 patients (42.9%) were men. As shown in Table 1, allergic skin diseases are the most commonly reported type (103/506, 20.4%), followed by papulosquamous disorders (58/506, 11.5%), infectious skin diseases (57/506, 11.3%), disorders of appendages (46/506, 9.1%), pigmentary disorders (38/506, 7.5%), hair disorders (38/506, 7.1%), and skin tumors (30/506, 5.9%).

Of the 506 participants, 252 (49.8%) reported unaffected outdoor activity, 126 (24.9%) reported partly restricted outdoor activity, and 128 (25.3%) reported isolation. As shown in Table 2, outdoor activity restriction was significantly associated with anxiety, depression, and impaired quality of life (all *P*<.05). A total of 102/506 patients (20.2%) reported unemployment, and approximately two-thirds of the participants (317/506, 62.6%) experienced decrease or loss of income during the pandemic. Both decreased income and unemployment were significantly associated with perceived stress, symptoms of anxiety and depression, and impaired quality of life (all *P*<.05).

Table 1. Characteristics of the participants (N=506).

Characteristic	Value
Age (years), mean (SD)	33.5 (14.0)
Gender, n (%)	
Male	217 (42.9)
Female	289 (57.1)
Marital status, n (%)	
Unmarried	189 (37.4)
Married	291 (57.5)
Widowed	22 (4.3)
Divorced	4 (0.8)
Educational level, n (%)	
Primary school and below	49 (9.7)
Middle school	69 (13.6)
High school	82 (16.2)
College and above	306 (60.5)
Income (¥) [a], n (%)	
<10,000	155 (30.6)
10,000-49,999	132 (26.1)
50,000-99,999	122 (24.1)
≥100,000	97 (19.2)
Skin disease, n (%)	
Infectious skin diseases	57 (11.3)
Papulosquamous disorders	58 (11.5)
Allergic skin diseases	103 (20.4)
Disorders of appendages	46 (9.1)
Disorders of hairs	36 (7.1)
Pigmentary disorders	38 (7.5)
Skin tumors	30 (5.9)
Other	189 (37.4)
Course of disease (years), n (%)	
<1	202 (39.9)
1-5	165 (32.6)
6-10	57 (11.3)
>10	82 (16.2)
Isolation, n (%)	
Outdoor activity unrestricted	252 (49.8)
Outdoor activity partly restricted	126 (24.9)
Isolated at home or in hospital	128 (25.3)
Loss of income, n (%)	
Unaffected	189 (37.4)
Reduced	208 (41.1)
Completely lost	109 (21.5)
Adherence to treatment, n (%)	

Characteristic	Value
Adherent to treatment	121 (23.9)
No treatment prescribed	210 (41.5)
Not adherent to treatment	175 (34.6)
Use of health care services, n (%)	
Visited a physician in hospital	126 (24.9)
Consulted a doctor remotely	92 (18.2)

[a]1 ¥ = US $0.15.

Table 2. Associations of epidemic-related impacts with patient-reported outcome scores of skin diseases (N=506).

Exposure group	Stress VAS[a]		PHQ-9[b]		GAD-7[c]		DLQI[d]		Health utility	
	Mean SE	*P* value	Mean SE	*P* value	Mean SE	*P* value	Mean SE	*P* value	Mean SE	*P* value
Isolation										
Unaffected	4.05 (0.18)	Ref [e]	4.56 (0.30)	Ref	3.31 (0.26)	Ref	5.47 (0.41)	Ref	0.96 (0.01)	Ref
Restricted	4.56 (0.26)	.10	6.33 (0.42)	<.001	4.97 (0.37)	.001	6.13 (0.58)	.35	0.93 (0.01)	.02
Isolated	4.38 (0.25)	.30	5.77 (0.42)	.02	4.20 (0.37)	.049	7.11 (0.58)	.02	0.93 (0.01)	.02
Loss of income										
Unaffected	3.23 (0.20)	Ref	4.17 (0.34)	Ref	3.19 (0.30)	Ref	4.85 (0.47)	Ref	0.96 (0.01)	Ref
Reduced	4.54 (0.19)	<.001	5.77 (0.32)	.001	4.19 (0.29)	.02	6.18 (0.45)	.04	0.94 (0.01)	.051
Completely lost	5.50 (0.26)	<.001	6.39 (0.45)	<.001	4.79 (0.40)	.002	7.87 (0.62)	<.001	0.92 (0.01)	<.001
Unemployment										
Unaffected	3.91 (0.14)	Ref	4.84 (0.23)	Ref	3.60 (0.21)	Ref	5.28 (0.32)	Ref	0.95 (0.00)	Ref
Unemployed	5.64 (0.28)	<.001	7.15 (0.46)	<.001	5.30 (0.41)	<.001	9.10 (0.63)	<.001	0.91 (0.01)	<.001
Adherence to treatment										
Adherent	4.75 (0.26)	Ref	5.64 (0.43)	Ref	3.59 (0.38)	Ref	6.24 (0.58)	Ref	0.93 (0.01)	Ref
No treatment needed	3.94 (0.20)	.01	4.41 (0.32)	.02	3.30 (0.29)	.54	4.12 (0.44)	<.001	0.97 (0.01)	<.001
Nonadherent	4.30 (0.22)	.18	6.14 (0.36)	.37	4.97 (0.31)	.005	8.23 (0.48)	.008	0.93 (0.01)	.75

[a]VAS: Visual Analog Scale.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]DLQI: Dermatology Life Quality Index.

[e] Ref: reference group.

We further categorized the scale scores by clinically relevant cutoffs and performed a series of logistic regression models with adjustments. As shown in Table 3, outdoor activity restriction was significantly associated with increased symptoms of depression (aOR 1.36-1.81) and anxiety (aOR 1.39-2.20) as well as impaired quality of life (aOR 1.22-1.78) in a dose-dependent manner (quarantined > partly restricted compared with unrestricted); however, it was not significantly associated with stress. Loss of income was correlated with stress (aOR 1.59-4.05), depression (aOR 2.56-2.56), anxiety (aOR 1.64-2.48), and impaired quality of life (aOR 1.27-2.62) in a dose-dependent manner (loss of income > reduced income compared with unaffected income). Similarly, unemployment was significantly associated with adverse outcomes, including perceived stress, depression, anxiety, and impaired quality of life.

Table 3. Associations of epidemic-related impacts with patient-reported outcomes of skin diseases (N=506).

Exposure group	Perceived stress (VAS[a] ≥7)			Depression (PHQ-9[b] ≥8)			Anxiety (GAD-7[c] ≥8)			Impaired quality of life (DLQI[d] ≥10)		
	aOR[e,f]	95% CI	*P* value	aOR	95% CI	*P* value	aOR	95% CI	*P* value	aOR	95% CI	*P* value
Isolation												
Unaffected	Ref[g]	N/A[h]	N/A	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A
Restricted	0.90	0.52-1.56	.71	1.36	0.84-2.21	.21	1.39	0.73-2.64	.31	1.22	0.70-2.12	.48
Isolated	1.02	0.59-1.77	.93	1.81	1.13-2.89	.013	2.20	1.23-3.96	.008	1.78	1.05-3.01	.03
Loss of income												
Unaffected	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A
Reduced	1.59	0.91-2.79	.10	2.22	1.38-3.57	.001	1.64	0.88-3.06	.12	1.27	0.73-2.20	.40
Complete loss	4.05	2.13-7.72	<.001	2.56	1.43-4.58	.002	2.48	1.19-5.13	.02	2.62	1.40-4.92	.003
Unemployment												
Unaffected	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A
Unemployed	2.41	1.42-4.08	.001	2.11	1.30-3.43	.003	2.60	1.45-4.65	.001	2.59	1.54-4.35	<.001
Adherence to treatment												
Adherent	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A	Ref	N/A	N/A
No treatment needed	0.81	0.461.41	.45	0.86	0.52-1.44	.26	0.66	0.33-1.29	.22	0.58	0.32-1.04	.07
Nonadherent	0.81	0.46-1.42	.46	1.39	0.84-2.31	.68	1.39	0.75-2.58	0.297	1.35	0.78-2.33	.29

[a]VAS: Visual Analog Scale.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]DLQI: Dermatology Life Quality Index.

[e]aOR: adjusted odds ratio.

[f]Adjusted for age, gender, income, and educational level.

[g]Ref: reference group.

[h]N/A: not applicable.

Among the 506 participants, 175 (34.6%) reported nonadherence to treatment. The reasons for nonadherence included limited accessibility to health care due to isolation (n=273, 53.9%), voluntarily stopping a prescription (n=106, 20.9%), forgetting to take medicine (n=71, 14.0%), and limited accessibility to medications (n=56, 11.1%). However, nonadherence was not significantly associated with adverse outcomes (Tables 2 and 3).

Discussion

Principal Findings

In the current study, we investigated the association of the impact of the COVID-19 pandemic with health-related quality of life in Chinese patients with skin diseases through a web-based survey. Our results indicated that over half of the respondents experienced quarantine and loss of income during the pandemic. These patients reported higher levels of perceived stress, increased symptoms of anxiety and depression, and impaired quality of life. To the best of our knowledge, this is the first impact analysis of the COVID-19 pandemic in patients with skin diseases. Our study reveals considerable proportions of impaired quality of life and mental well-being in these patients. Telemedicine, mental health intervention, and social support are needed for patients with skin diseases during this particular period.

Studies have indicated increases in mental health issues among medical staff [18] and patients with chronic diseases [19] since the outbreak of COVID-19. However, there is little research regarding the impact of the pandemic on patients with skin diseases. Here, we identified isolation, unemployment, and loss of income as significant risk factors for poor mental well-being and quality of life in patients with skin diseases.

Public health efforts to prevent the spread of COVID-19 are having a growing impact on the global economy. A considerable number of individuals are losing their jobs, at least temporarily [20]. It has been estimated that the worldwide unemployment rate will increase from 4.936% to 5.644% due to the COVID-19 pandemic [2]. According to our data, over 20% of participants (109/506, 21.5%) became unemployed, which is notably higher than the reported unemployment rate. Additionally, 208 of the 506 patients (41.1%) experienced reduced income even though they were not unemployed; this also resulted in impaired mental health and quality of life. In addition to the financial impacts, job loss can disrupt health insurance coverage [21]. Although many countries have provided free testing or even treatment for COVID-19, medical costs for other conditions related to the pandemic are generally not covered by these reimbursement policies.

It should be noted that 65% of the respondents' nonadherence to treatment was related to limited accessibility to health institutions or medications based on our observations. Since the pandemic, many hospitals and health care institutions have temporarily closed their outpatient services to avoid nosocomial infection and better allocate health care resources, which has created unique challenges to health care delivery at present [22]. A previous study described the possibility of using telemedicine in disasters or public health emergencies [23]. Most developed areas have been using telemedicine not only for physical health care, but also for mental health services [24]. However, it is difficult for patients in rural areas who lack internet services and smartphones to use remote health care. Current telemedicine programs in China and some countries enable consultation services but do not provide medication delivery services; this could eventually cause nonadherence to treatment due to lack of medication.

Limitations

Our study has some limitations. First, the survey was web-based, with low representativeness due to nonprobability sampling. Second, the exposure and outcome variables were self-reported, and recall bias may have been introduced. Third, some relevant outcomes such as sleep quality were not included in our study because a heavy survey burden may lead to fewer responses and lower accuracy. Last, the survey was conducted among Chinese patients and may not fully represent all patients in areas beyond China due to cultural differences as well as variations in reimbursement policies and social systems.

Conclusions

Taken together, our findings indicate that pandemic-related impacts are associated with adverse patient-reported outcomes of skin diseases. Early and timely mental health intervention, telemedicine, and health education are needed for these patients. Preferably, social support and reimbursement policies will further help patients under heavy financial burdens endure this difficult period.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62041208, 81974478), the Ministry of Science and Technology of the People's Republic of China (2016YFC0900802), the Emergency Project of Prevention and Control for COVID-19 of Central South University (502701002), and the Department of Science and Technology of Hunan Province (2018SK2086, 2018SK2092). We would like to thank all the melanoma patients who participated in our study for their support and assistance in the web-based survey. Particularly, we want to express our high regard to all the health practitioners and volunteers for their dedication in the fight against COVID-19.

Authors' Contributions

Yeye Guo and Minxue Shen contributed equally as first authors. Xiang Chen and Juan Su contributed equally as last authors.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-05-21]
2. Kawohl W, Nordt C. COVID-19, unemployment, and suicide. Lancet Psychiatry 2020 May;7(5):389-390 [FREE Full text] [doi: 10.1016/S2215-0366(20)30141-3] [Medline: 32353269]
3. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. BMJ 2020 Mar 23;368:m1165. [doi: 10.1136/bmj.m1165] [Medline: 32205334]
4. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
5. Long H, Van Dam A. U.S. unemployment rate soars to 14.7 percent, the worst since the Depression era. Washington Post. 2020 May 08. URL: https://www.washingtonpost.com/business/2020/05/08/april-2020-jobs-report/ [accessed 2020-09-01]

6. Acs G, Karpman M. Employment, Income, and Unemployment Insurance during the COVID-19 Pandemic. Urban Institute. 2020 Jun 30. URL: https://www.urban.org/research/publication/employment-income-and-unemployment-insurance-during-covid-19-pandemic [accessed 2020-09-01]
7. Alradhawi M, Shubber N, Sheppard J, Ali Y. Effects of the COVID-19 pandemic on mental well-being amongst individuals in society- A letter to the editor on "The socio-economic implications of the coronavirus and COVID-19 pandemic: A review". Int J Surg 2020 Jun;78:147-148 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.070] [Medline: 32380230]
8. Hay RJ, Johns NE, Williams HC, Bolliger IW, Dellavalle RP, Margolis DJ, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. J Invest Dermatol 2014 Jun;134(6):1527-1534 [FREE Full text] [doi: 10.1038/jid.2013.446] [Medline: 24166134]
9. Verhoeven EWM, Kraaimaat FW, van de Kerkhof PCM, van Weel C, Duller P, van der Valk PGM, et al. Prevalence of physical symptoms of itch, pain and fatigue in patients with skin diseases in general practice. Br J Dermatol 2007 Jun;156(6):1346-1349. [doi: 10.1111/j.1365-2133.2007.07916.x] [Medline: 17535233]
10. Basra MKA, Shahrukh M. Burden of skin diseases. Expert Rev Pharmacoecon Outcomes Res 2009 Jun;9(3):271-283. [doi: 10.1586/erp.09.23] [Medline: 19527100]
11. Bickers DR, Lim HW, Margolis D, Weinstock MA, Goodman C, Faulkner E, American Academy of Dermatology Association, Society for Investigative Dermatology. The burden of skin diseases: 2004 a joint project of the American Academy of Dermatology Association and the Society for Investigative Dermatology. J Am Acad Dermatol 2006 Sep;55(3):490-500. [doi: 10.1016/j.jaad.2006.05.048] [Medline: 16908356]
12. Dutheil F, Pereira B, Moustafa F, Naughton G, Lesage F, Lambert C. At-risk and intervention thresholds of occupational stress using a visual analogue scale. PLoS One 2017;12(6):e0178948 [FREE Full text] [doi: 10.1371/journal.pone.0178948] [Medline: 28586383]
13. Lesage FX, Berjot S. Validity of occupational stress assessment using a visual analogue scale. Occup Med (Lond) 2011 Sep;61(6):434-436. [doi: 10.1093/occmed/kqr037] [Medline: 21505089]
14. Liu Z, Yu Y, Hu M, Liu H, Zhou L, Xiao S. PHQ-9 and PHQ-2 for Screening Depression in Chinese Rural Elderly. PLoS One 2016;11(3):e0151042 [FREE Full text] [doi: 10.1371/journal.pone.0151042] [Medline: 26978264]
15. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
16. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol 1994 May;19(3):210-216. [doi: 10.1111/j.1365-2230.1994.tb01167.x] [Medline: 8033378]
17. Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. Value Health 2014 Jul;17(5):597-604 [FREE Full text] [doi: 10.1016/j.jval.2014.05.007] [Medline: 25128053]
18. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiat 2020 Apr;7(4):e15-e16. [doi: 10.1016/s2215-0366(20)30078-x]
19. Kuang Y, Shen M, Wang Q, Xiao Y, Lv C, Luo Y, et al. Association of outdoor activity restriction and income loss with patient-reported outcomes of psoriasis during the COVID-19 pandemic: A web-based survey. J Am Acad Dermatol 2020 Aug;83(2):670-672 [FREE Full text] [doi: 10.1016/j.jaad.2020.05.018] [Medline: 32407741]
20. Gangopadhyaya AA, Garrett AB. Unemployment, Health Insurance, and the COVID-19 Recession. SSRN Journal 2020 Apr 01:online. [doi: 10.2139/ssrn.3568489]
21. Woolhandler S, Himmelstein DU. Intersecting U.S. Epidemics: COVID-19 and Lack of Health Insurance. Ann Intern Med 2020 Jul 07;173(1):63-64 [FREE Full text] [doi: 10.7326/M20-1491] [Medline: 32259195]
22. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
23. Lurie N, Carr BG. The Role of Telehealth in the Medical Response to Disasters. JAMA Intern Med 2018 Jun 01;178(6):745-746. [doi: 10.1001/jamainternmed.2018.1314] [Medline: 29710200]
24. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

Abbreviations

aOR: adjusted odds ratio
DLQI: Dermatology Life Quality Index
GAD-7: Generalized Anxiety Disorder-7
PHQ-9: 9-Item Patient Health Questionnaire
PSS-14: Perceived Stress Scale-14
STROBE: Strengthening the Reporting of Observational studies in Epidemiology
VAS: Visual Analog Scale

Edited by G Eysenbach; submitted 07.07.20; peer-reviewed by Y Yu, Y Luo; comments to author 22.07.20; revised version received 06.08.20; accepted 10.08.20; published 11.09.20.

Please cite as:
Guo Y, Shen M, Zhang X, Xiao Y, Zhao S, Yin M, Bu W, Wang Y, Chen X, Su J
Association of Socioeconomic Changes due to the COVID-19 Pandemic With Health Outcomes in Patients With Skin Diseases: Cross-Sectional Survey Study
J Med Internet Res 2020;22(9):e22288
URL: http://www.jmir.org/2020/9/e22288/
doi:10.2196/22288
PMID:32845850

Original Paper

Excessive Media Consumption About COVID-19 is Associated With Increased State Anxiety: Outcomes of a Large Online Survey in Russia

Nikita A Nekliudov[1*], BSc; Oleg Blyuss[1,2*], PhD; Ka Yan Cheung[3], BSc; Loukia Petrou[3,4], BSc; Jon Genuneit[5], MSc, MD, PhD; Nikita Sushentsev[6], MD; Anna Levadnaya[7], MD, PhD; Pasquale Comberiati[8], MD; John O Warner[3*], OBE, MD, FRCP, FRCPCH, FMedSci, FERS; Gareth Tudor-Williams[9*], MD, MRCP, MRCPCH; Martin Teufel[10*], MD, PhD; Matthew Greenhawt[11*], MSc, MBA, MD; Audrey DunnGalvin[1,12*], PhD; Daniel Munblit[1,9,13*], MD, MSc, PhD

[1]Department of Paediatrics and Paediatric Infectious Diseases, Institute of Child's Health, Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russian Federation

[2]School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield, United Kingdom

[3]Inflammation, Repair and Development Section, National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, United Kingdom

[4]Department of Bioengineering, Imperial College London, London, United Kingdom

[5]Pediatric Epidemiology, Department of Pediatrics, Medical Faculty, Leipzig University, Leipzig, Germany

[6]Department of Radiology, Addenbrooke's Hospital and University of Cambridge, Cambridge, United Kingdom

[7]National Medical Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russian Federation

[8]Department of Clinical and Experimental Medicine, Section of Pediatrics, University of Pisa, Pisa, Italy

[9]Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, United Kingdom

[10]Clinic for Psychosomatic Medicine and Psychotherapy, LVR University Hospital, University of Duisburg-Essen, Essen, Germany

[11]Department of Pediatrics, Section of Allergy/Immunology, Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO, United States

[12]Applied Psychology and Paediatrics and Child Health, University College Cork, Cork, Ireland

[13]Solov'ev Research and Clinical Center for Neuropsychiatry, Moscow, Russian Federation

[*]these authors contributed equally

Corresponding Author:
Daniel Munblit, MD, MSc, PhD
Department of Infectious Disease
Faculty of Medicine
Imperial College London
Wright Fleming Building
Norfolk Place, Paddington
London, W2 1PG
United Kingdom
Phone: 44 (0) 7898257151
Email: daniel.munblit08@imperial.ac.uk

Abstract

Background: The COVID-19 pandemic has potentially had a negative impact on the mental health and well-being of individuals and families. Anxiety levels and risk factors within particular populations are poorly described.

Objective: This study aims to evaluate confidence, understanding, trust, concerns, and levels of anxiety during the COVID-19 pandemic in the general population and assess risk factors for increased anxiety.

Methods: We launched a cross-sectional online survey of a large Russian population between April 6 and 15, 2020, using multiple social media platforms. A set of questions targeted confidence, understanding, trust, and concerns in respondents. The State-Trait Anxiety Inventory was used to measure anxiety. Multiple linear regressions were used to model predictors of COVID-19–related anxiety.

Results: The survey was completed by 23,756 out of 53,966 (44.0% response rate) unique visitors; of which, 21,364 were residing in 62 areas of Russia. State Anxiety Scale (S-Anxiety) scores were higher than Trait Anxiety Scale scores across all regions of Russia (median S-Anxiety score 52, IQR 44-60), exceeding published norms. Time spent following news on COVID-19 was strongly associated with an increased S-Anxiety adjusted for baseline anxiety level. One to two hours spent reading COVID-19 news was associated with a 5.46 (95% CI 5.03-5.90) point difference, 2-3 hours with a 7.06 (95% CI 6.37-7.74) point difference, and more than three hours with an 8.65 (95% CI 7.82-9.47) point difference, all compared to less than 30 minutes per day. Job loss during the pandemic was another important factor associated with higher S-Anxiety scores (3.95, 95% CI 3.31-4.58). Despite survey respondents reporting high confidence in information regarding COVID-19 as well as an understanding of health care guidance, they reported low overall trust in state and local authorities, and perception of country readiness.

Conclusions: Among Russian respondents from multiple social media platforms, there was evidence of higher levels of state anxiety associated with recent job loss and increased news consumption, as well as lower than expected trust in government agencies. These findings can help inform the development of key public health messages to help reduce anxiety and raise perceived trust in governmental response to this current national emergency. Using a similar methodology, comparative surveys are ongoing in other national populations.

(J Med Internet Res 2020;22(9):e20955) doi:10.2196/20955

KEYWORDS

anxiety; COVID-19; media consumption; SARS-CoV-2; STAI; state anxiety; trait anxiety; trust to government; trust; mental health; social media; survey

Introduction

In December 2019, the first patients with pneumonia of unknown cause were linked to a seafood wholesale market in Wuhan, China [1]. This is generally recognized as the beginning of a previously unknown beta-coronavirus pandemic related to the SARS-CoV-2 virus, which has subsequently spread worldwide. This COVID-19 pandemic has resulted in dramatic changes to normal life in many countries, leading to disruptions of social and economic functioning comparable to the impact of the Spanish flu pandemic of 1918.

COVID-19 illness has rapidly spread, producing high numbers of fatalities in the absence of proven pharmacologic treatments and vaccines, [2] and lag time in application of testing, contact tracing, and mass quarantine measures. Concomitantly, there has been a rise in reported mental health problems such as fear, anxiety, depression, and sleep problems among different subgroups worldwide [3,4]. A recent UK survey and Ipsos MORI poll showed public concerns about the effect of social isolation or social distancing on well-being; increased anxiety, depression, stress, and other negative feelings; and concerns related to current and potential future financial difficulties [5].

A recently published Lancet Psychiatry position paper highlighted the need to collect high-quality data on the mental health effects of the COVID-19 pandemic across the whole population as a top priority [5] requiring global action. The World Health Organization (WHO) has outlined research priorities for containing COVID-19 and supporting those affected. A March 26, 2020, press briefing specifically addressed the pandemic's effect on mental health: *"With the disruptive effects of COVID-19 – including social distancing – currently dominating our daily lives, it is important that we...are mindful of and sensitive to the unique mental health needs of those we care for. Our anxiety and fears should be...better understood and addressed"* [6]. Strategies advocating physical or social distancing have become central to pandemic control in many countries, but to be effective, these require universal adoption within society, which has been variable internationally. Quality of communication can be quite impactful on personal psychology and behaviors during health emergencies. The WHO Director-General has cautioned about an *infodemic* of misinformation in online platforms, which may negatively impact how society is perceiving and responding to the COVID-19 outbreak.

The aim of this study is to try to measure associations between COVID-19 perception and anxiety from an international perspective to address unmet needs in the field. We hypothesized that there would be increased COVID-19–related state anxiety (driven by pandemic-specific events affecting individuals) compared with pre-existing levels of trait anxiety, which likely varies depending on the country surveyed. Herein we present the initial findings on levels of state and trait anxiety (and their determinants) in a large sample of people residing in Russia during the COVID-19 pandemic, a previously poorly described region with respect to global mental health concerns.

Methods

Study Design and Population

A cross-sectional open 160-item online survey was conducted between April 6 and 15, 2020, timed to follow the Russian government's announcement of a "stay-at-home" order through April 30 [7]. The survey was promoted on three social media platforms (VK, Facebook, and Instagram) via influencers, a popular Russian search engine (Yandex), and the Russian internet media portal (Meduza). We used nonprobability sampling [8] through social media to allow for rapid data collection, which was particularly important at the peak of the lockdown period. The survey was pretested with members of the public that had no role in the questionnaire design to ensure good understandability and identify discrepancies in wording as well as missing facets that may have been previously overlooked. No registration was required to access the survey.

This paper is compliant with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [9].

Ethics Statement

The study was reviewed and approved by Sechenov University Ethics Committee on April 2, 2020. Participants were informed at the outset that by completing and submitting their responses to the online survey, they were consenting to voluntary participation in a research study of attitudes and behaviors surrounding COVID-19. The survey settings and web analytics were accessible only to authors DM and NAN. At the end of the survey, participants were asked to provide a limited amount of personal data (email) that participants agreed could be held on the research database. This request was optional and did not have any impact on survey completion.

Survey Questionnaire

For the purpose of this survey, we modified and further adapted a questionnaire developed by author MT at the University of Duisburg-Essen, Clinic for Psychosomatic Medicine and Psychotherapy, LVR University Hospital Essen [10]. A single master survey was developed for adaptation, translation according to WHO protocol, and similar dissemination in other countries to provide a comparative analyses of COVID-19 risk factors for mental health across countries, with particular attention to the impact of information.

The survey consisted of several modules, assessing basic demographic information; socioeconomic status; employment; living conditions; health status; medications intake; time following news on COVID-19; confidence in and understanding of information; trust of state authorities; trust of local authorities; worry, concern, or adverse expectations; perception of risk; personal protection measures; and behavioral aspects.

The health status of each participant was categorized, taking into account previous data on chronic conditions' impact on mental health and up-to-date evidence on the risk factors for mortality from COVID-19 infection. People with depression or cardiological and respiratory conditions were among the main potential determinants of anxiety; cardiological and respiratory conditions are known risk factors for mortality from COVID-19 [11], whereas depression is known to have a detrimental association with anxiety [12]. If more than one of these factors were present in an individual, we considered this a higher risk and subcategorized respondents into another subgroup. Participants reporting oncological conditions or HIV and diabetes, renal, or hepatic problems were combined into subcategories due to a limited number of people reporting these chronic diseases.

Four psychometric measures, previously validated in the Russian population, were included: the State-Trait Anxiety Inventory (STAI), Big Five Inventory-10, General Self-Efficacy Scale, and the Patient Health Questionnaire. We applied branching logic where it was justified to spare the users' time while providing their responses. The survey was partitioned into 7 pages, including the introduction page containing general information about the survey. The estimated time to complete the survey did not exceed 20 minutes.

State and Trait Anxiety Inventory

The STAI was used to measure self-reported presence and severity of current symptoms of anxiety and a generalized propensity to be anxious [13]. The STAI has been previously validated for use in the Russian population and is freely available in the Russian language [14]. The index consists of two subscales. The State Anxiety Scale (S-Anxiety) measures current anxiety in the moment, assessing subjective feelings of apprehension, tension, nervousness, worry, and activation or arousal of the autonomic nervous system. The Trait Anxiety Scale (T-Anxiety) evaluates relatively stable aspects of *anxiety proneness*, including general states of calmness, confidence, and security [13]. The range of scores for each subtest is 20-80, the higher score indicating greater anxiety.

Confidence, Understanding, Trust, and Concerns

Fifteen ad-hoc designed questions assessed self-perceived confidence in information and understanding (feeling informed about COVID-19 and understanding the guidance from health care authorities); trust in state and local authorities, and country readiness for the pandemic; governmental measures (whether respondents consider measures excessive or not); worry, concern, or adverse outcome expectations, including potential consequences to the individual and the country. Respondents were provided with a 9-point Likert scale, where 1 represented *complete disagreement* and 9 *complete agreement* with a given statement.

Statistical Analysis

Descriptive statistics were used for baseline characteristics of responders including sociodemographics and scores of psychometric tools. Multiple linear regressions were used to model the association of potential variables of interest with the results of psychometric tools. Additional multiple logistic regression was also performed with the dichotomized S-Anxiety score as the outcome using the cutoff of 45. Multiple analyses were adjusted for the available baseline characteristics. In line with suggested recommendations [15], the results were adjusted for the multiple comparisons using a Bonferroni correction, which resulted in an alpha value threshold of P=.001 being used for statistical significance.

Distributions of the tests' results were assessed with density plots and box and whisker plots. All the analyses were performed using R version 3.6.3 (The R Foundation for Statistical Computing). Maps of state and trait anxiety across regions of Russia were plotted using the "ssplot" library. Two-sided P values were reported for all statistical tests.

Results

Participants

The survey link was accessed 57,877 times. The number of unique visitors was 53,966, of which 42,643 (79.0% participation rate) gave their consent to participate and accessed the first survey page; 23,756 out of 42,643 (55.7% completion rate) users completed the questionnaire. The response rate was 44.0% with 23,756 responses from 53,966 unique visitors. Unique users were identified using cookies set at the top-level

domain, the expiration date was preset to 1 year. The view ratio could not be calculated due to the variety of traffic sources used for survey distribution. The presets did not allow users to send incomplete questionnaires, so those were not present in the exported tabulated data set. There was no specific cutoff that would exclude a questionnaire from the analysis.

The demographic characteristics of the study participants are summarized in Table 1. Out of 23,756 respondents, 21,364 were residing in Russia at the time of survey completion; data from all of the latter were included in the analyses (median age 32 years, IQR 28-36; range 18-82 years). Out of 21,364 respondents, 18,609 (87.1%) were female, 14,752 (68.2%) were married or in a relationship, 14,371 (67.3%) had children younger than 18 years, and 4.4% (n=933) were expecting a child at the time of survey completion. Of the 21,364 participants, 53% (n=11,450) reported various chronic conditions and 3789 (17.7%) were current smokers.

The 21,364 respondents were equally distributed between the capital (n=7468, 35.0%), large cities (n=6348, 29.7%), and smaller localities (n=7548, 35.2%). A total 62 areas of Russia were represented with at least 40 respondents from each location. The population was highly educated with 17,688 (82.8%) having a higher degree and a further 1635 (7.7%) studying at university. There were 1648 (7.7%) respondents that lost their jobs due to the COVID-19 pandemic. There were 19,589 (91.7%) respondents who were not related to the health care profession and did not study medicine.

Table 1. Sociodemographic characteristics of survey respondents residing in the Russian Federation at the time of the COVID-19 pandemic (N=21,364).

Characteristics	Participants
Sex (female), n (%)	18,609 (87.1)
Age (years), median (IQR)	32 (28-36)
Age (years), range	18-82
18-25, n (%)	2991 (14)
26-35, n (%)	12,893 (60.3)
36-45, n (%)	4418 (20.7)
≥46, n (%)	1062 (5)
Marital status, married or in relationship, n (%)	14,752 (68.2)
Have children younger than 18 years, n (%)	14,371 (67.3)
Expecting a child, n (%)	933 (4.4)
City of residence, n (%)	
Capital	7468 (35.0)
Large city (over 500,000 inhabitants)	6348 (29.7)
Smaller cities/towns	7548 (35.2)
Education status, n (%)	
PhD	547 (2.6)
More than one degree	1458 (6.8)
Master's degree	4362 (20.4)
Bachelor's degree	11,321 (53.0)
Higher education in progress	1635 (7.7)
Vocational school	1507 (7.0)
School	412 (1.9)
Other	122 (0.6)
Income (R), n (%)	
Decline to answer	928 (4.3)
<20,000	4377 (20.5)
20,000-34,999	5308 (24.8)
35,000-69,999	6012 (28.1)
70,000-99,999	2412 (11.3)
100,000-149,999	1351 (6.3)
≥150,000	976 (4.6)
Chronic medical conditions, n (%)	
No	9603 (44.9)
Decline to answer	311 (1.5)
Depression and cardiological or respiratory	247 (1.2)
Depression or neurological	523 (2.4)
Allergies (food allergy/allergic rhinitis) or dermatological (eczema/psoriasis)	2283 (10.7)
Cardiological	742 (3.5)
Cardiological and respiratory	65 (0.3)
Renal/hepatic/diabetes	705 (3.3)
Oncology/HIV	269 (1.3)
Other	6438 (30.1)

Characteristics	Participants
Respiratory	178 (0.8)
Neuroleptics/antidepressant use, n (%)	832 (3.9)
Time spent on reading COVID-19 news, n (%)	
Decline to answer	39 (0.2)
Do not follow	335 (1.6)
Do not follow but they find me	2808 (13.1)
<30 min	6641 (31.1)
30 min-1 hour	6922 (32.4)
1 hour-2 hours	3019 (14.1)
2 hours-3 hours	964 (4.5)
>3 hours	636 (3)
Smoking status, n (%)	
Nonsmoker	13,546 (63.4)
Former smoker	4029 (18.9)
Current smoker	3789 (17.7)
Job status, n (%)	
Decline to answer	430 (2.0)
Do not work	8294 (38.8)
Lost job due to COVID-19 and out of job now	1648 (7.7)
Work from home	8366 (39.2)
Commute to work	2626 (12.3)
Health care–related job, n (%)	
No	19,589 (91.7)
Medical student	222 (1.0)
Volunteer/hospital management	283 (1.3)
Nurse	305 (1.4)
Physician	965 (4.5)
S-Anxiety[a], median (IQR)	52 (44-60)
T-Anxiety[b], median (IQR)	44 (39-51)

[a]S-Anxiety: State Anxiety Scale.

[b]T-Anxiety: Trait Anxiety Scale.

Levels of Trait and State Anxiety

Scores for both T-Anxiety and S-Anxiety of survey respondents were high, with medians of 44 (IQR 39-51) and 52 (IQR 44-60), respectively.

Median scores for S-Anxiety were higher than T-Anxiety across all areas of Russia (Figures 1-3 and Multimedia Appendix 1) with four areas (Belgorod, Kostroma, Mordovia, and Orel) having a score difference of 10 points or greater. In 32 out of 62 (52%) areas, the difference between S-Anxiety and T-Anxiety reached 8-9.5 points and, in 24 out of 62 (39%) areas, between 5 and 7.5 points. In two areas (Sakhalin and Karelia), the difference was less than 5 points. The difference between S-Anxiety and T-Anxiety in the largest Russian cities, Moscow and St. Petersburg, was approximately 8 points. A reference map providing detail on geographical locations of the main areas of Russia is provided in Multimedia Appendix 2.

Figure 1. Density plot showing the difference between the state and trait anxiety based on the responses of all participants (N=21,364). STAI: State-Trait Anxiety Inventory.

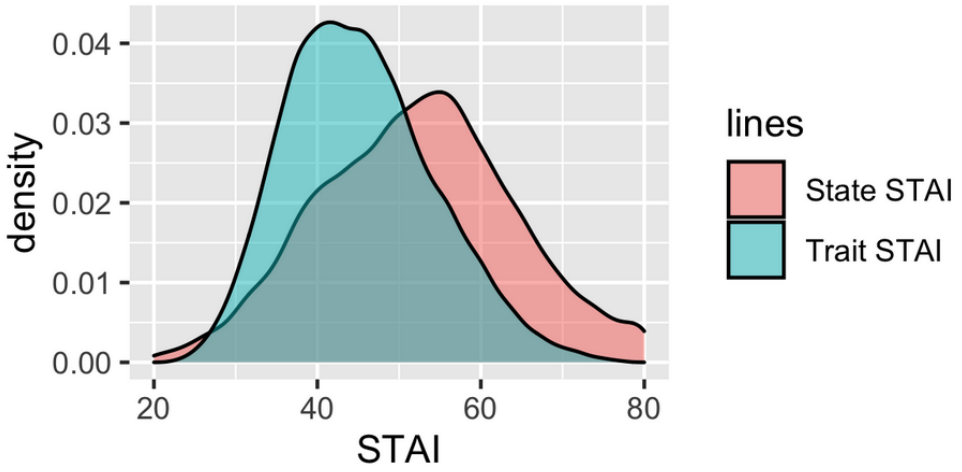


Figure 2. Map of Russia showing the levels of respondents' trait anxiety (T-Anxiety). Areas with data from less than 40 respondents are not shown on the map. T-Anxiety: Trait Anxiety Scale.

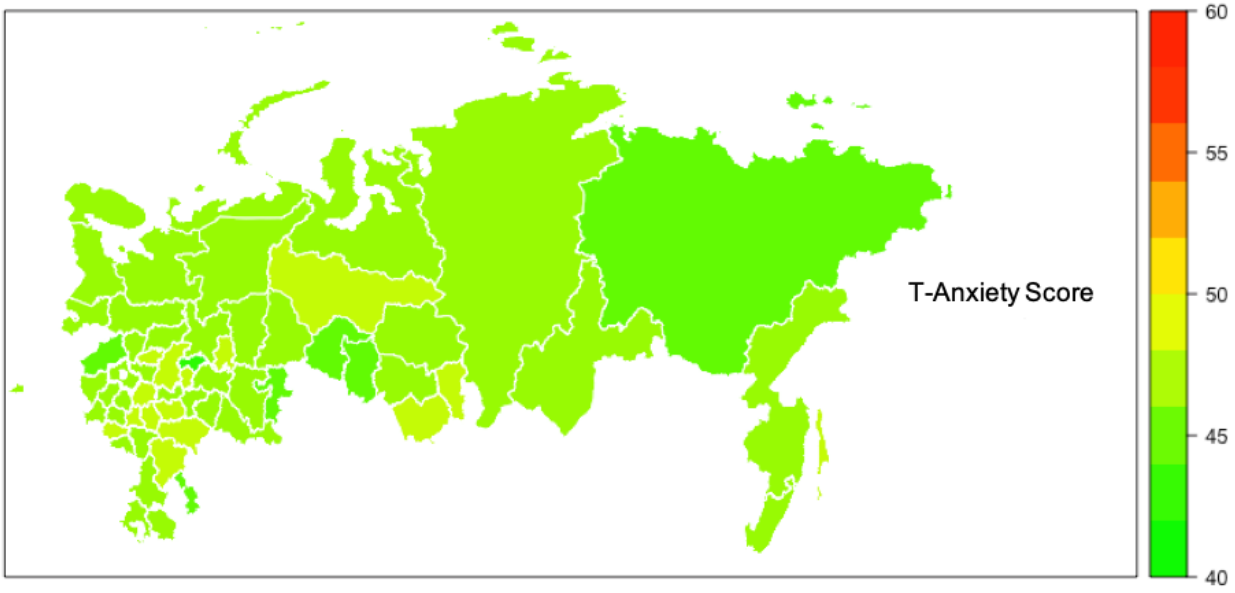
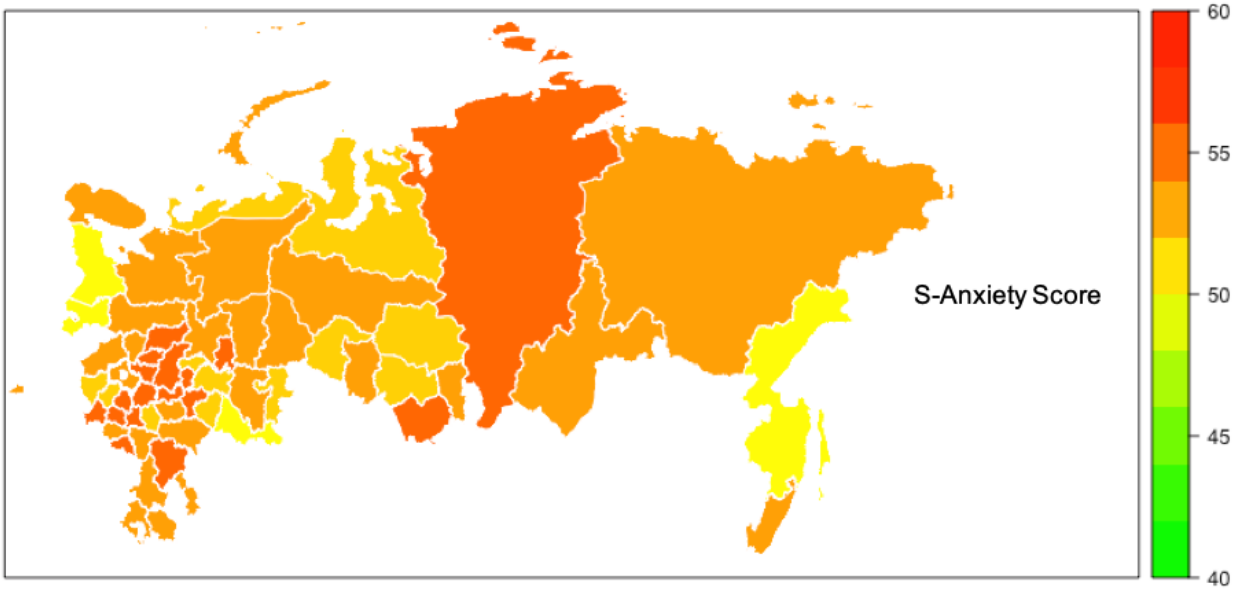


Figure 3. Map of Russia showing the levels of respondents' state anxiety (S-Anxiety). Areas with data from less than 40 respondents are not shown on the map. S-Anxiety: State Anxiety Scale.



Determinants of State Anxiety

To assess factors associated with S-Anxiety (related to COVID-19), we developed a multiple regression model. T-Anxiety was included to adjust for pre-existing or typical state of anxiety for each individual (Table 2). The multiple logistic regression model with a cut-off value of S-Anxiety at 45 points yielded similar results (Multimedia Appendix 3).

Time spent following news on COVID-19 was significantly associated with higher scores of S-Anxiety, with one to two hours resulting in a 5.46 (95% CI 5.03 to 5.90) point difference, two to three hours in a 7.06 (95% CI 6.37 to 7.74) point difference, and more than three hours in a 8.65 (95% CI 7.82 to 9.47) point difference, all compared to up to 30 minutes per day. In addition, *job loss due to the pandemic* resulted in a 3.95 (95% CI 3.31 to 4.58) point difference, a combination of depression with either cardiovascular or respiratory conditions in a 3.19 (95% CI 1.89 to 4.49) point difference, taking neuroleptics or antidepressants in a 1.32 (95% CI 0.59 to 2.06) point difference, and smokers in a 1.16 (95% CI 0.74 to 1.50) point difference, which all had significantly higher S-Anxiety scores.

Males had a lower level of S-Anxiety than females (–4.01, 95% CI –4.45 to –3.57). Parents of children younger than 18 years and people expecting a child also had slightly lower scores of S-Anxiety, –1.44 (95% CI –1.84 to –1.04) and –1.12 (95% CI –1.79 to –0.45), respectively.

Table 2. Regression model assessing associations between characteristics and state anxiety scores, adjusted for the trait anxiety.

Model and variable	Coefficient	SE	*P* value	95% CI
Sex				
Male vs female	–4.011	0.224	*<.001* [a]	–4.45 to –3.572
Age	–0.021	0.011	.046	–0.042 to 0
Marital status				
In relationship vs single	0.287	0.248	.25	–0.2 to 0.773
Married vs single	0.243	0.236	.30	–0.22 to 0.706
Have children younger than 18 years				
No vs yes	–1.441	0.203	*<.001*	–1.839 to –1.044
Expecting a child				
No vs yes	–1.117	0.341	*.001*	–1.785 to –0.448
Living in a capital				
No vs yes	–0.404	0.156	.009	–0.709 to –0.099
Education				
BSc vs vocational school	0.177	0.293	.55	–0.397 to 0.751
MSc vs vocational school	0.475	0.319	.14	–0.151 to 1.101
Other vs vocational school	–2.016	0.951	.03	–3.88 to –0.153
More than one degree vs vocational school	0.568	0.386	.14	–0.189 to 1.325
Higher education in progress vs vocational school	–0.45	0.387	.25	–1.209 to 0.309
PhD vs vocational school	–0.174	0.523	.74	–1.2 to 0.852
School vs vocational school	–1.836	0.567	*.001*	–2.946 to –0.725
Income (R)				
Decline to answer vs <20,000	–0.205	0.368	.58	–0.927 to 0.517
20,000-35,000 vs <20,000	–0.28	0.208	.18	–0.687 to 0.128
35,000-70,000 vs <20,000	–0.487	0.211	.02	–0.9 to –0.074
70,000-100,000 vs <20,000	–0.311	0.274	.26	–0.847 to 0.226
100,000-150,000 vs <20,000	–0.313	0.334	.35	–0.968 to 0.342
>150,000 vs <20,000	–0.529	0.381	.17	–1.276 to 0.219
Chronic medical conditions				
Any vs no	1.072	0.142	*<.001*	0.793 to 1.351
Decline to answer vs no	2.774	0.58	*<.001*	1.636 to 3.911
Depression and cardiological or respiratory vs no	3.187	0.664	*<.001*	1.885 to 4.489
Depression or neurological vs no	0.109	0.467	.82	–0.805 to 1.024
Food allergy/rhinitis/eczema/psoriasis vs no	0.633	0.234	.007	0.174 to 1.091
Cardiological vs no	0.765	0.387	.048	0.007 to 1.522
Cardiological and respiratory vs no	2.123	1.252	.09	–0.33 to 4.577
Renal/hepatic/diabetes vs no	1.221	0.392	.002	0.452 to 1.99
Oncology/HIV vs no	0.857	0.625	.17	–0.368 to 2.083
Other vs no	1.281	0.165	*<.001*	0.958 to 1.605
Respiratory vs no	0.533	0.761	.48	–0.959 to 2.025
Medications				
Neuroleptics/antidepressant vs no	1.324	0.376	*<.001*	0.586 to 2.061
Time spent on reading COVID-19 news				

Model and variable	Coefficient	SE	*P* value	95% CI
Decline to answer vs <30 mins	2.955	1.616	.07	–0.213 to 6.122
Do not follow vs <30 mins	–4.767	0.563	*<.001*	–5.87 to –3.663
Do not follow but they find me vs <30 mins	1.16	0.226	*<.001*	0.716 to 1.603
30 mins-1 hour vs <30 mins	3.083	0.173	*<.001*	2.743 to 3.423
1-2 hours vs <30 mins	5.463	0.222	*<.001*	5.027 to 5.899
2-3 hours vs <30 mins	7.059	0.349	*<.001*	6.374 to 7.743
>3 hours vs <30 mins	8.645	0.421	*<.001*	7.819 to 9.471
Smoking				
Former smoker vs nonsmoker	0.265	0.181	.14	–0.09 to 0.621
Current smoker vs nonsmoker	1.115	0.193	*<.001*	0.736 to 1.494
Job status				
Decline to answer vs commute to work	0.189	0.529	.72	–0.847 to 1.225
Do not work vs commute to work	–0.443	0.242	.07	–0.916 to 0.031
Work from home vs commute to work	–0.86	0.24	*<.001*	–1.331 to –0.39
Lost due to COVID-19 and out of job vs commute to work	3.948	0.323	*<.001*	3.314 to 4.581
Health care–related job				
Medical student vs no	–1.185	0.701	.09	–2.559 to 0.189
Volunteer/hospital management vs no	–0.637	0.604	.29	–1.822 to 0.547
Nurse vs no	–0.876	0.607	.15	–2.066 to 0.314
Physician vs no	–0.886	0.349	.01	–1.57 to –0.201
T-Anxiety[b]	0.543	0.008	*<.001*	0.528 to 0.559

[a]Italics indicate significant results.

[b]T-Anxiety: Trait Anxiety Scale.

Subgroup Analysis

An additional regression analysis in a subgroup of participants with low T-Anxiety scores was performed to see if determinants of S-Anxiety would remain the same even in people with a generally good state of calmness, confidence, and security (Multimedia Appendix 4). The effect sizes were even stronger in this group. Time spent following news on COVID-19 was found to be significantly associated with higher scores of S-Anxiety, with one to two hours resulting in an 8.06 (95% CI 4.75 to 11.38) point difference, two to three hours in a 13.77 (95% CI 7.86 to 19.67) point difference, and more than three hours in a 21.61 (95% CI 13.97 to 29.25) point difference. Job loss due to the pandemic was also strongly associated with higher S-Anxiety scores (10.95, 95% CI 6.62 to 15.28). The level of S-Anxiety for males was lower than in females (–4.011, 95% CI –4.45 to –3.57).

As most of the survey respondents were female, we ran an additional analysis to assess S-Anxiety in a subgroup of male participants (Multimedia Appendix 5), which yielded similar results to the previous regression models with respect to time spent following news on COVID-19 and job loss due to the pandemic. We investigated the influence of the survey completion date on the outcome, using S-Anxiety scores in all participants as well as the subgroup with low T-Anxiety scores (Multimedia Appendix 6) but noted no statistically significant differences.

Confidence, Understanding, Trust, and Concerns

Regarding questions on confidence, understanding, trust, and concerns related to COVID-19 (Figure 4 and Table 3), most of the respondents felt well-informed (scores between 7 and 9) on COVID-19 (n=11,129, 52.1%), measures to prevent infection (n=14,149, 66.2%), and understanding of the guidance from health care authorities (n=16,670, 78.0%).

Figure 4. Respondents' answers to the questions addressing confidence, understanding, trust, and concerns. CC: The COVD-19 situation concerns me significantly; CE: I believe the crisis caused by COVID-19 will eventually resolve with little consequence for my country's economy; CJ: I believe the crisis caused by COVID-19 will eventually resolve with little consequence for my job/business; CP: I think the country I am responding from is well prepared for COVID-19; EMLA: I think the measures taken by the local authorities in the city/town/village/etc against COVID-19 are excessive; EMSG: I think the measures taken by the country government against COVID-19 are excessive; GHA: I understand the guidance from health care authorities related to COVID-19; IC: I feel informed about COVID-19; IMPC: I feel informed about measures to prevent infection with COVID-19; LAM: I think that all possible local authority measures to fight COVID-19 are being taken in my city/town/village/etc; NIDL: The COVID-19 situation is negatively impacting my day-to-day life; Q: question; SGM: I think all possible government measures to fight COVID-19 are being taken in my country; TLA: I trust the local authorities in the city/town/village/etc I am responding from; TSG: I trust the government in the country I am responding from; WBI: I am worried about becoming infected with COVID-19 no matter how much I take care of myself.

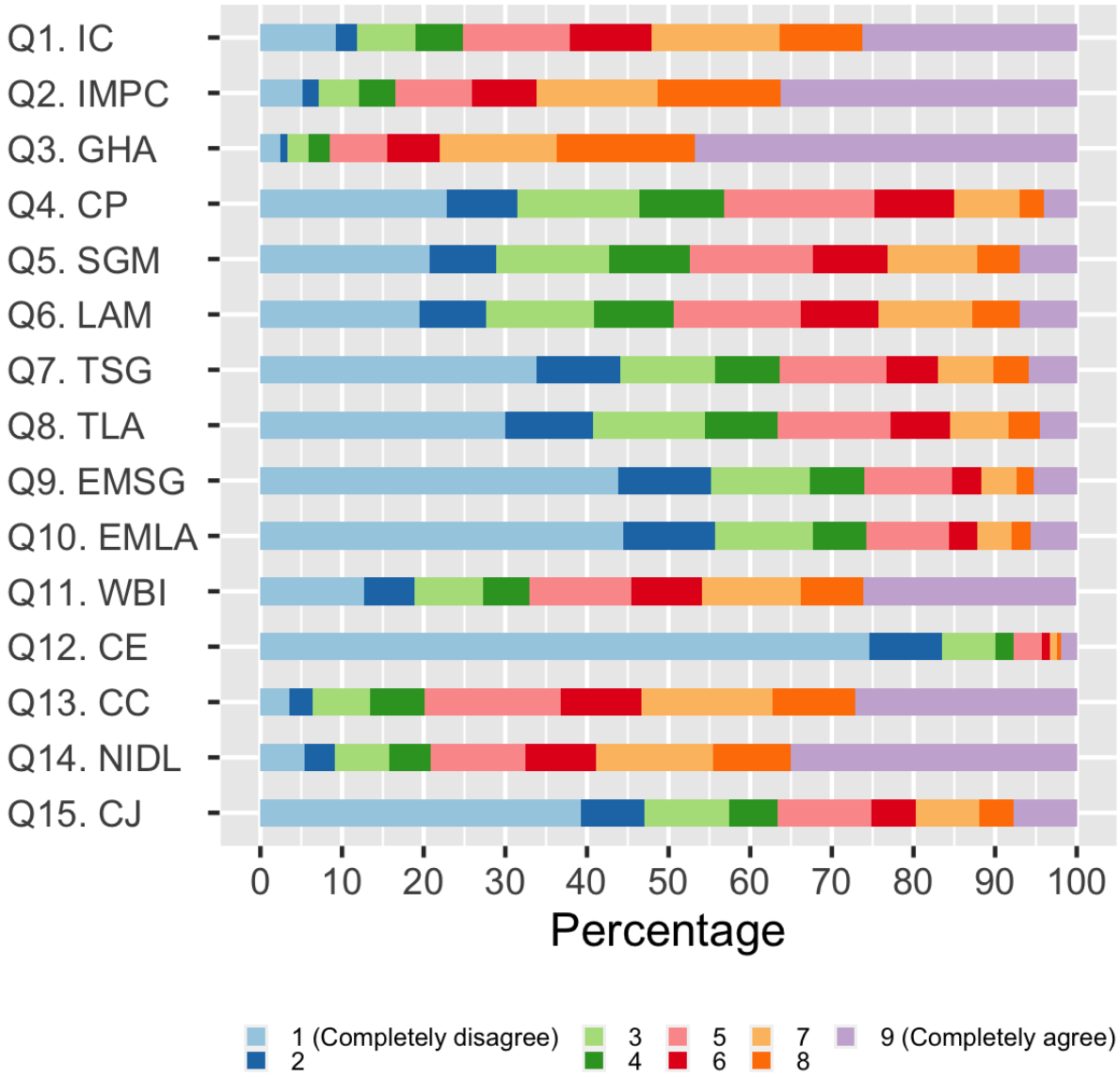


Table 3. Respondents answers to the questions addressing confidence, understanding, trust, and concerns.[a]

Views on COVID-19	Response, median (IQR)
Confidence in information and understanding	
Q[b]1. I feel informed about COVID-19.	7 (5-9)
Q2. I feel informed about measures to prevent infection with COVID-19.	8 (5-9)
Q3. I understand the guidance from health care authorities related to COVID-19.	8 (7-9)
Trust to state and local authorities, and country readiness for pandemic	
Q4. I think the country I am responding from is well prepared for COVID-19.	4 (2-5)
Q5. I think all possible government measures to fight COVID-19 are being taken in my country.	4 (2-6)
Q6. I think that all possible local authority measures to fight COVID-19 are being taken in my city/town/village/etc.	4 (2-6)
Q7. I trust the government in the country I am responding from.	3 (1-5)
Q8. I trust the local authorities in the city/town/village/etc I am responding from.	3 (1-5)
Governmental measures evaluation	
Q9. I think the measures taken by the country government against COVID-19 are excessive.	2 (1-5)
Q10. I think the measures taken by the local authorities in the city/town/village/etc against COVID-19 are excessive.	2 (1-5)
Worry/concern/adverse expectation	
Q11. I am worried about becoming infected with COVID-19 no matter how much I take care of myself.	6 (3-9)
Q12. I believe the crisis caused by COVID-19 will eventually resolve with little consequence for my country's economy.	1 (1-2)
Q13. The COVD-19 situation concerns me significantly.	7 (5-9)
Q14. The COVID-19 situation is negatively impacting my day-to-day life.	7 (5-9)
Q15. I believe the crisis caused by COVID-19 will eventually resolve with little consequence for my job/business.	3 (1-6)

[a]Answers were provided with a 9-point Likert scale, where 1 is *completely disagree* and 9 is *completely agree*.

[b]Q: question.

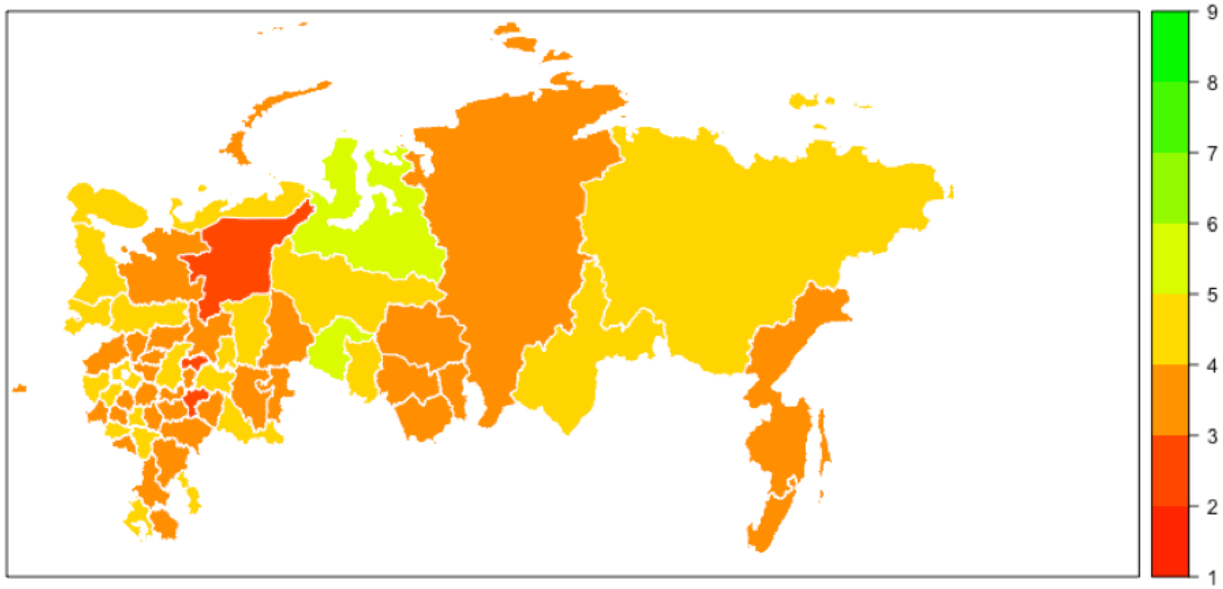
Out of the 21,364 participants, very few people considered that the country was well-prepared for the pandemic (n=3202, 15.0%) and that all possible state and local government measures to fight COVID-19 were being taken (n=4950, 23% and n=5189, 24%, respectively), and more than half of the respondents reported low trust in the state government and local authorities (n=11,890, 56% and n=11,624, 54%, respectively). Interestingly, most participants did not consider the measures taken by the state and local government to be excessive (n=14,382, 67% and n=14,459, 68%, respectively). Regarding the economic ramifications, 19,240 (90%) respondents did not believe that the crisis caused by COVID-19 will eventually resolve with little consequence for the country's economy, and 12,269 (57%) were worried about the consequences to their business or employment. More than half of participants (n=12,570, 58.8%) reported the pandemic was "negatively impacting my day to day life" and were extremely concerned with the situation (11,389, 53.3%), and 9812 (46%) were worried about becoming infected "no matter how much I take care of myself."

Respondents demonstrated good confidence in information and in understanding of COVID-19 across most areas of Russia (Multimedia Appendix 7 and Figure 5), with 51 out of 62 (82%) areas having combined median scores of seven or more. The rest of the country had a median score between 6 and 7.

Figure 5. Map of Russia showing the levels of respondents' confidence in information and understanding. Areas with data from less than 40 respondents are not shown on the map. The combined median score on confidence in information and understanding was used (Question [Q]1. I feel informed about COVID-19; Q2. I feel informed about measures to prevent infection with COVID-19; Q3. I understand the guidance from health care authorities related to COVID-19). Respondents were provided with a 9-point Likert scale, where 1 is completely disagree and 9 is completely agree.

Combined median scores on trust of state and local authorities, and country readiness for the pandemic did not exceed 5 on a nine-point scale across Russia, reaching 5 in two areas only (Tyumen' and Yamal-Nenetsk). In 23 out of 62 (37%) areas, the median score varied between 4 and 4.8; in 33 out of 62 (53%) areas, the median score ranged between 3 and 3.9. A median score below 3 was recorded in three areas of Russia (Komi, Mari El, and Ul'yanovsk). The median score in Moscow was 4 points, with an even lower median (3.1) seen in respondents from St. Petersburg. Figure 6 shows median scores in areas across Russia.

Figure 6. Map of Russia showing the levels of respondents' trust to state and local authorities, and country readiness for the pandemic. Areas with data from less than 40 respondents are not shown on the map. The combined median score on trust to state and local authorities, and country readiness for the pandemic (Question [Q]4. I think the country I am responding from is well prepared for COVID-19; Q5. I think all possible government measures to fight COVID-19 are being taken in my country; Q6. I think that all possible local authority measures to fight COVID-19 are being taken in my city/town/village/etc; Q7. I trust the government in the country I am responding from; Q8. I trust the local authorities in the city/town/village/etc I am responding from). Respondents were provided with a 9-point Likert scale, where 1 is completely disagree and 9 is completely agree.



Discussion

Principal Findings

The findings herein demonstrate higher state versus trait anxiety among a large sample of people residing in Russia during the COVID-19 pandemic. State anxiety was strongly associated with the amount of time spent following news of COVID-19, as well as job loss during the pandemic. Although our study design and analysis did not allow us to assess causal inference, the amount of time spent following news may be driven in part by low levels of trust in state and local authorities.

Widely accepted reference norms for STAI in the Russian population suggest that scores up to 30 are equivalent to a low level of anxiety, while scores above 45 indicate a high or *clinical* level of anxiety. The median scores for S-Anxiety in our online survey respondents were exceedingly high (52, IQR 44-60), which was expected, but this level of anxiety in a large sample size of young adults is concerning.

Some of the findings may be mediated by the timing of survey administration in conjunction with recent governmental actions taken less than a week prior, when the Russian President Vladimir Putin announced the prolongation of "the official non-work period" until the end of April and alluded to regional decision making regarding public health measures [7]. By that time, several (but not all) regions including Moscow had introduced social or physical distancing measures, the stringency of which were further adjusted over the course of this study while remaining variable across the country. Therefore, the results must be viewed through a lens of a fluid and evolving situation in a geographically large and culturally diverse country with variable infection rates and public health needs when analyzing interregional variation of the anxiety state. Furthermore, no data regarding anxiety, trust, or other data reported herein were available to help inform potential decisions. Such data are now available and support interventions for a psychologically burdened people [16]. Electronic mental health interventions such as *CoPE It* could be made available as evidence-based psychotherapeutic and psychological support to overcome psychological distress [17].

Moreover, similarly to other countries, additional support measures for the private sector and guarantees of wage retention for the workers of nongovernmental organizations have evolved since the launch of this survey. A window existed where respondents were surveyed between announcements, which may have influenced perception of the adequacy of the response, in particular with oil price fluctuations occurring just prior to the fielding of this survey [18]. This lack of hindsight clarity at the beginning stages could explain the observed anxiety related to job loss and the lack of belief in the crisis caused by COVID-19 resolving with little consequence for the respondents' jobs and for the country's economy. This is why longitudinal assessment of this population is important, to show if there is resolution of some of the trends.

Previous research suggested that high frequency of risk-elevating messages in the news may contribute to increased concerns of the public in relation to infectious diseases, as it was witnessed with the Ebola virus disease in the US population [19]. Our findings are in agreement with previous data both during the current pandemic [20,21] as well as the severe acute respiratory syndrome outbreak in 2006 [22]; the time of media consumption was a major factor associated with higher S-Anxiety in our respondents. The effect was particularly strong among participants with low T-Anxiety, which could indicate ceiling effects among those with high T-Anxiety (ie, individuals with high T-Anxiety already have increased anxiety, which does not leave enough space for S-Anxiety to rise to a large extent). Another interpretation is that subjects who usually having low anxiety in normal life show a stronger response during the stressful time of a pandemic. Individuals with higher T-Anxiety might be more prone to increased duration or frequency of media consumption; however, a strong association among those with low T-Anxiety may point toward the directionality of media consumption causing S-Anxiety rather than the other way around, which is supported by previously published data [23]. The design of this study does not allow establishment of causality, but this hypothesis may merit further exploration in future research.

Our data support the findings by Ni et al [20], suggesting that spending ≥2 hours a day following COVID-19–related news was associated with probable depression and anxiety in adults. Social media sources may impose a danger to mental health during lockdown due to the high volume of contradictory information they may deliver [22,24,25], which is particularly evident in light of quarantine measures imposed. A recent report by Tangcharoensathien et al [26] highlighted the importance of interaction with social media platforms to provide reliable information and keep the infodemic under control, underlining the necessity of keeping closer attention to the coherence of information in the media and to take specific action to alleviate their impact on mental health.

The impact of the COVID-19 pandemic on the Russian population remains largely unknown with a limited number of studies on a small sample size available [27-29]. Sorokin et al [27] reported an overall moderate level of anxiety and found similar associations between unemployment, female gender, and lower education with the level of distress. Similarly, high levels of anxiety were found in the Russian student population [28,29], but due to the difference in scales used, the results cannot be compared with our data.

Strengths and Limitations

The size of the data set, assessing mental aspects of a sizeable portion of the Russian population during the pandemic, is a defined strength of the study. The survey resulted in a reputable completion rate compared with an average response rate in online surveys (33%) [30]. This may be a reflection of social or physical isolation and policies urging people to shelter-in-place, who could then dedicate more time to responding.

There were several limitations to the study. Several items were not standardized and validated, although we were careful to include others that were. However, a broad international multidisciplinary team of researchers, with expertise in psychology, epidemiology, and clinical medicine including

infectious diseases were involved in this ad hoc survey development to measure the pandemic response. The survey was distributed online, with the help of influencers via social networks, media, and search engines. This may have introduced selection bias due to the probability that people using media websites and persons who follow online influencers and take a questionnaire may possess higher health literacy and be more informed on a wide range of topics, which reduces generalizability of the findings to the average Russian citizen. However, Russia is among the top-10 countries in the world with the highest number of internet users [31], and most of the people in the studied age group use internet on a regular basis all over Russia. Another concern was age bias because many of the respondents were young adults, and the views of people 65 years and older, the most vulnerable population during the COVID-19 pandemic, remained uncaptured. Finally, another limitation of this study is the disproportionately large number of female respondents, which is, however, a common observation in online and paper-based surveys [32,33]. The gender difference is likely to be secondary to the means of the questionnaire's distribution, that is, via social media and influencers.

Conclusions

The results of this survey suggest a higher rate of S-Anxiety when compared with T-Anxiety diffusely among the Russian population. Media consumption, job loss, and associated uncertainty around future employment prospects due to the pandemic were strongly associated with increased S-Anxiety. Given the evolving pandemic situation, further research is needed to track the trajectory of perceptions regarding trust and the perception of the adequacy of the governmental response. This also provides time for some of the information contained herein to help inform policy and direct intervention at a segment of the population at high risk for mental health issues related to the pandemic and what influences the anxiety. It is important to address the COVID-19 pandemic and the *infodemic* in tandem by understanding if mass communication impacts state anxiety. Our findings will increase our understanding of the risks and consequences of social isolation on the population and how these are informed by rapidly changing data, an endless news cycle, and social media.

In their position paper, Holmes et al [5], as well as the WHO experts [6], called for immediate action to assess "the effect of repeated media consumption about COVID-19 in traditional and social media on mental health" and "the role of repeated media consumption in amplifying distress and anxiety" [5]. Our data confirm the association between an excessive reception of information on COVID-19 through the media and S-Anxiety. Lack of trust in the state and local authorities is also worrying. Governments must take note and consider how presentation might be adjusted to avoid excess anxiety in the population.

Acknowledgments

We are grateful to our participants who dedicated their time to completing the survey during the COVID-19 pandemic. We would like to thank the search engine Yandex, internet media portal Meduza, and all the influencers who kindly helped with the survey distribution and promotion.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Median scores of trait (State-Trait Anxiety Inventory–trait) and state (State-Trait Anxiety Inventory–state) anxiety reported by the respondents, residing in different regions (with 40 respondents or more) of the Russian Federation.
[DOCX File , 21 KB - jmir_v22i9e20955_app1.docx]

Multimedia Appendix 2
Reference map of the Russian Federation.
[PNG File , 499 KB - jmir_v22i9e20955_app2.png]

Multimedia Appendix 3
Multiple logistic regression model assessing associations between characteristics and state anxiety scores, adjusted for the trait anxiety. Statistically significant results presented in bold.
[DOCX File , 25 KB - jmir_v22i9e20955_app3.docx]

Multimedia Appendix 4
Regression model assessing associations between characteristics and state anxiety scores in a subset of respondents with low trait anxiety scores (n=683). Statistically significant results presented in bold.
[DOCX File , 22 KB - jmir_v22i9e20955_app4.docx]

Multimedia Appendix 5

Regression model assessing associations between characteristics and state anxiety scores in a subset of respondents with low trait anxiety scores (n=683). Statistically significant results presented in bold.
[DOCX File , 23 KB - jmir_v22i9e20955_app5.docx]

Multimedia Appendix 6
Changes in State Anxiety Scale scores over time in (a) all respondents and (b) a subset of respondents with low trait anxiety scores.
[PNG File , 112 KB - jmir_v22i9e20955_app6.png]

Multimedia Appendix 7
Median scores reported by the respondents residing in different regions (with 40 respondents or more) of Russian Federation with regards to (a) confidence in information and understanding, combined median score on confidence in information and understanding was used (Question [Q]1. I feel informed about COVID-19; Q2. I feel informed about measures to prevent infection with COVID-19; Q3. I understand the guidance from health care authorities related to COVID-19) and (b) combined median score on trust to state and local authorities, and country readiness for pandemic (Q4. I think the country I am responding from is well prepared for COVID-19; Q5. I think all possible government measures to fight COVID-19 are being taken in my country; Q6. I think that all possible local authority measures to fight COVID-19 are being taken in my city/town/village/etc; Q7. I trust the government in the country I am responding from; Q8. I trust the local authorities in the city/town/village/etc I am responding from). Respondents were provided with a 9-point Likert scale, where 1 is completely disagree and 9 is completely agree.
[DOCX File , 18 KB - jmir_v22i9e20955_app7.docx]

Multimedia Appendix 8
Regression model assessing associations between characteristics and state anxiety, including trust to state and local authorities, country readiness for the pandemic, and confidence in information and understanding. Statistically significant results presented in bold.
[DOCX File , 23 KB - jmir_v22i9e20955_app8.docx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. JAMA 2020 May 13:A. [doi: 10.1001/jama.2020.6019] [Medline: 32282022]
3. Yang Y, Li W, Zhang Q, Zhang L, Cheung T, Xiang Y. Mental health services for older adults in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e19 [FREE Full text] [doi: 10.1016/S2215-0366(20)30079-1] [Medline: 32085843]
4. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiatry 2020 Mar;7(3):228-229 [FREE Full text] [doi: 10.1016/S2215-0366(20)30046-8] [Medline: 32032543]
5. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
6. Kluge H, Malik A, Nitzan D. Mental health and psychological resilience during the COVID-19 pandemic. World Health Organization. 2020 Mar 27. URL: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/mental-health-and-psychological-resilience-during-the-covid-19-pandemic [accessed 2020-05-06]
7. Putin V. Address to the nation. President of Russia. Moscow: Kremlin; 2020 Apr 02. URL: http://en.kremlin.ru/events/president/news/63133 [accessed 2020-05-05]
8. Wolf C, Joye D, Smith TW, Fu YC. The SAGE Handbook of Survey Methodology. Newbury Park, CA: SAGE Publications Ltd; 2016.
9. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
10. Teufel M. Study on handling COVID-19 / SARS-CoV-2 (corona virus). University of Duisburg-Essen, Clinic for Psychosomatic Medicine and Psychotherapy, LVR Clinic Essen2020. 2020. URL: https://ww2.unipark.de/uc/2020/ [accessed 2020-05-05]
11. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, ISARIC4C investigators. Features of 20 133 UK patients in hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020 May 22;369:m1985 [FREE Full text] [doi: 10.1136/bmj.m1985] [Medline: 32444460]

12. Kessler RC, Birnbaum HG, Shahly V, Bromet E, Hwang I, McLaughlin KA, et al. Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: results from the WHO World Mental Health Survey Initiative. Depress Anxiety 2010 May;27(4):351-364 [FREE Full text] [doi: 10.1002/da.20634] [Medline: 20037917]
13. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken) 2011 Dec;63 Suppl 11:S467-S472. [doi: 10.1002/acr.20561] [Medline: 22588767]
14. Spielberger C, Hanin Y. State-Trait Anxiety Inventory (STAI) in Russian. Russia: Psylab URL: https://tinyurl.com/y5fkote5 [accessed 2020-05-08]
15. Mundfrom DJ, Perrett JJ, Schaffer J, Piccone A, Roozeboom M. Bonferroni adjustments in tests for regression coefficients. Multiple Linear Regression Viewpoints 2006;32(1):1-6 [FREE Full text]
16. Bäuerle A, Skoda E, Dörrie N, Böttcher J, Teufel M. Psychological support in times of COVID-19: the Essen community-based CoPE concept. J Public Health (Oxf) 2020 Aug 18;42(3):649-650 [FREE Full text] [doi: 10.1093/pubmed/fdaa053] [Medline: 32307516]
17. Bäuerle A, Graf J, Jansen C, Dörrie N, Junne F, Teufel M, et al. An e-mental health intervention to support burdened people in times of the COVID-19 pandemic: CoPE It. J Public Health (Oxf) 2020 Aug 18;42(3):647-648 [FREE Full text] [doi: 10.1093/pubmed/fdaa058] [Medline: 32364242]
18. Gershkovich E, Cordell J. In first address to nation on the coronavirus, Putin holds back from stringent measures. The Moscow Times. Moscow, Russia; 2020 Mar 27. URL: https://www.themoscowtimes.com/2020/03/25/in-first-address-to-nation-on-the-coronavirus-putin-holds-back-from-stringent-measures-a69748 [accessed 2020-05-05]
19. Sell TK, Boddie C, McGinty EE, Pollack K, Smith KC, Burke TA, et al. Media messages and perception of risk for Ebola virus infection, United States. Emerg Infect Dis 2017 Jan;23(1):108-111. [doi: 10.3201/eid2301.160589] [Medline: 27983495]
20. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: cross-sectional survey. JMIR Ment Health 2020 May 12;7(5):e19009 [FREE Full text] [doi: 10.2196/19009] [Medline: 32365044]
21. Li J, Yang Z, Qiu H, Wang Y, Jian L, Ji J, et al. Anxiety and depression among general population in China at the peak of the COVID-19 epidemic. World Psychiatry 2020 Jul;19(2):249-250. [doi: 10.1002/wps.20758] [Medline: 32394560]
22. Smith RD. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. Soc Sci Med 2006 Dec;63(12):3113-3123 [FREE Full text] [doi: 10.1016/j.socscimed.2006.08.004] [Medline: 16978751]
23. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357. [doi: 10.1037/hea0000875] [Medline: 32202824]
24. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. Int J Soc Psychiatry 2020 Jun;66(4):317-320. [doi: 10.1177/0020764020915212] [Medline: 32233719]
25. Temsah M, Al-Sohime F, Alamro N, Al-Eyadhy A, Al-Hasan K, Jamal A, et al. The psychological impact of COVID-19 pandemic on health care workers in a MERS-CoV endemic country. J Infect Public Health 2020 Jul;13(6):877-882 [FREE Full text] [doi: 10.1016/j.jiph.2020.05.021] [Medline: 32505461]
26. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
27. Sorokin M, Kasyanov E, Rukavishnikov G, Makarevich O, Neznanov N, Lutova N, et al. Structure of anxiety associated with COVID-19 pandemic: the online survey results. BRSMU 2020 Jun 2((3)2020):70-76. [doi: 10.24075/brsmu.2020.030]
28. Gritsenko V, Skugarevsky O, Konstantinov V, Khamenka N, Marinova T, Reznik A, et al. COVID 19 fear, stress, anxiety, and substance use among Russian and Belarusian university students. Int J Ment Health Addict 2020 May 21:1-7 [FREE Full text] [doi: 10.1007/s11469-020-00330-z] [Medline: 32837418]
29. Yehudai M, Bender S, Gritsenko V, Konstantinov V, Reznik A, Isralowitz R. COVID-19 fear, mental health, and substance misuse conditions among university social work students in Israel and Russia. Int J Ment Health Addict 2020 Jul 06:1-8 [FREE Full text] [doi: 10.1007/s11469-020-00360-7] [Medline: 32837438]
30. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? Assess Eval Higher Education 2008 Jun;33(3):301-314. [doi: 10.1080/02602930701293231]
31. Top 20 countries with the highest number of internet users. Internet World Stats. 2019. URL: https://www.internetworldstats.com/top20.htm [accessed 2020-05-06]
32. Sax LJ, Gilmartin SK, Bryant AN. Assessing response rates and nonresponse bias in web and paper surveys. Res Higher Education 2003;44(4):409-432. [doi: 10.1023/A:1024232915870]
33. Smith W. Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior. ERIC Document Reproduction Service No. ED 501717 2008;1:1-21 [FREE Full text]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys

STAI: State-Trait Anxiety Inventory
S-Anxiety: State Anxiety Scale
T-Anxiety: Trait Anxiety Scale
WHO: World Health Organization

Edited by G Eysenbach; submitted 02.06.20; peer-reviewed by K Li, L Hassan; comments to author 15.07.20; revised version received 05.08.20; accepted 06.08.20; published 11.09.20.

Please cite as:
Nekliudov NA, Blyuss O, Cheung KY, Petrou L, Genuneit J, Sushentsev N, Levadnaya A, Comberiati P, Warner JO, Tudor-Williams G, Teufel M, Greenhawt M, DunnGalvin A, Munblit D
Excessive Media Consumption About COVID-19 is Associated With Increased State Anxiety: Outcomes of a Large Online Survey in Russia
J Med Internet Res 2020;22(9):e20955
URL: https://www.jmir.org/2020/9/e20955
doi:10.2196/20955
PMID:32788143

Original Paper

An Index for Lifting Social Distancing During the COVID-19 Pandemic: Algorithm Recommendation for Lifting Social Distancing

Sam Li-Sheng Chen[1], PhD; Amy Ming-Fang Yen[1], PhD; Chao-Chih Lai[2], MD; Chen-Yang Hsu[2], ScD; Chang-Chuan Chan[3], ScD; Tony Hsiu-Hsi Chen[2], PhD

[1]College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

[2]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

[3]Institute of Environmental and Occupational Health Science, College of Public Health, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Tony Hsiu-Hsi Chen, PhD
Institute of Epidemiology and Preventive Medicine
College of Public Health
National Taiwan University
Room 533, No 17, Xu-Zhou Road
Taipei, 100
Taiwan
Phone: 886 233668033
Email: chenlin@ntu.edu.tw

Abstract

Background: Implementing and lifting social distancing (LSD) is an urgent global issue during the COVID-19 pandemic, particularly when the travel ban is lifted to revive international businesses and economies. However, when and whether LSD can be considered is subject to the spread of SARS-CoV-2, the recovery rate, and the case-fatality rate. It is imperative to provide real-time assessment of three factors to guide LSD.

Objective: A simple LSD index was developed for health decision makers to do real-time assessment of COVID-19 at the global, country, region, and community level.

Methods: Data on the retrospective cohort of 186 countries with three factors were retrieved from a publicly available repository from January to early July. A simple index for guiding LSD was measured by the cumulative number of COVID-19 cases and recoveries, and the case-fatality rate was envisaged. If the LSD index was less than 1, LSD can be considered. The dynamic changes of the COVID-19 pandemic were evaluated to assess whether and when health decision makers allowed for LSD and when to reimplement social distancing after resurgences of the epidemic.

Results: After large-scale outbreaks in a few countries before mid-March (prepandemic phase), the global weekly LSD index peaked at 4.27 in March and lasted until mid-June (pandemic phase), during which most countries were affected and needed to take various social distancing measures. Since, the value of LSD has gradually declined to 0.99 on July 5 (postpandemic phase), at which 64.7% (120/186) of countries and regions had an LSD<1 with the decile between 0 and 1 to refine risk stratification by countries. The LSD index decreased to 1 in about 115 days. In addition, we present the results of dynamic changes of the LSD index for the world and for each country and region with different time windows from January to July 5. The results of the LSD index on the resurgence of the COVID-19 epidemic in certain regions and validation by other emerging infectious diseases are presented.

Conclusions: This simple LSD index provides a quantitative assessment of whether and when to ease or implement social distancing to provide advice for health decision makers and travelers.

(J Med Internet Res 2020;22(9):e22469) doi:10.2196/22469

KEYWORDS

COVID-19; pandemic; social distancing; index; algorithm; data analysis; decision making; global health; public health

Introduction

Although border controls and social distancing have been executed since the beginning of the COVID-19 pandemic [1-4], we must consider how to lift social distancing in the postpandemic period with real-time assessment, possibly weekly [5], as reviving economic business and normal social activities are needed.

Although there are six criteria for countries to consider when de-escalating by reversing restrictions or lockdowns [6], it is still unclear whether and when to implement the reopening policy. Doing so is highly dependent on three determinants. The first is to consider the transmission rate of COVID-19 that is often captured by the basic reproductive number (R_0; the expected number of secondary cases produced by an index case in a susceptible population) using the susceptible- infected-recovered or susceptible- exposed- infected- recovered models to evaluate the effectiveness of social distancing [2-4,7-11]. The second is pertaining to the optimal management of patients with COVID-19 that determines the rate of recovery from hospitalization or self-isolation. The third is strongly related to the second determinant and critical care capacity, and concerns preventing mild or moderate patients from deteriorating into severe patients, potentially causing further death from COVID-19. A simple lifting social distancing (LSD) index is required to quantify the impacts of these three factors on LSD. The implications for developing an LSD index are two-fold. From a global perspective, providing this LSD index can aid policy makers worldwide in evaluating and deciding when to reopen the border if the spread of SARS-CoV-2 can be contained even with small cluster infections. From an individual viewpoint, if information on this index can be evaluated periodically and incorporated into traveling apps or websites, it would be helpful for the traveler to be aware of information on these three determinants in each country or region to decide whether it is safe for them to travel to their destinations and, if they must travel in the coming weeks, how they can be prepared to protect themselves from being infected with COVID-19 in high LSD areas.

The aim of this study was to develop a simple index for guiding LSD with assessment on the global, country, region, and community level from January until early July.

Methods

Data

The data for the following analysis were derived from the web-based real-time GitHub repository created by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [12,13]. CSSE operates daily updates of publicly available data, including confirmed cases, recovered cases, and deaths from multiple sources. A total of 186 countries have reported confirmed COVID-19 cases (including presumptive positive cases and probable cases) at the country and region level, which are aligned with the World Health Organization (WHO) situation reports [14]. A total of 186 affected countries and regions were available from January 21 to July 5, 2020.

These three factors (cases, recoveries, and deaths), available from open data, were used as proxy variables for the corresponding three metrics, including transmissibility that is often captured by the R_0, number of beds for hospitalization, and number of beds for the intensive care unit. However, the two latter metrics may not be available from open data in the three periods (prepandemic, pandemic, and postpandemic), described in the next section, from a global perspective.

Social Distancing and Three Factors Related to COVID-19

At the beginning of the outbreak, each country focused on the evaluation of the spread of COVID-19, which often refers to the evolution of the R_0. With time, those accumulated COVID-19 cases have invoked demand for hospitalization and critical care provided for moderate and severe patients, respectively. To reduce the burden of medical resources and disease burden of death, various containment measures would be adopted to stamp out large-scale community-acquired outbreaks. Although different terminologies have been used worldwide, the broad term “social distancing” is used here to represent containment measures and is defined as border control across countries, lockdown across cities, physical distancing between individuals, and use of face masks. Nonetheless, if social distancing lasts for a long time, economic and social activities will be restricted. In the postpandemic period, the majority of countries and regions may still have small outbreaks after large-scale outbreaks. To balance social distancing and economic revival, each country or region has to consider whether the medical capacity can afford newly increasing confirmed COVID-19 cases arising from small outbreaks due to LSD. Assessing three factors (cases, recoveries, and deaths) simultaneously that are reciprocal and correlated with each other provides an insight into the balance between LSD and disease burden.

Statistical Analysis

Index for LSD

We developed the simple index for LSD, which originated from the susceptible-exposed-infected-recovered-death (SEIRD) compartment model for infectious disease [15]. The semantics of SEIRD is denoted by five symbols, including susceptible (S), exposed (E), infected (I), recovered (R), and death (D), as illustrated in the figure in Multimedia Appendix 1. It has been already applied to modeling the dynamic transmission of SARS-CoV-2 [16]. In a mathematical way, let the number of the compartments S, E, I, R, and D at time t be denoted by s(t), e(t), i(t), r(t), and m(t), respectively. Given a time point, participants of a country or region belong to one of the five states. The instantaneous change for the compartments thus constitutes a system of differential equations:

where β represents the transmission coefficient, and α, γ, and τ denote the progression rates to infectious status, recovery rate, and case-fatality rate, respectively.

Based on the system of differential equations in equation 1, the cumulative frequencies of cases, recoveries, and deaths with the consideration of the COVID-19 evolution among a population, given period *t*, can be derived as follows:

Note that the transmission of COVID-9 captured by the transmission coefficient, β, is altered by social distancing measures that would reduce the frequency of contact and the probability of transmission [17,18], the recovery rate and case-fatality rate that are determined by the optimal management of medical care, the capacity of hospitalization and intensive care, and the quality of care [19]. Without a proper triage and diversion of patients with COVID-19, a compromised recovery rate and an increasing case fatality will be expected [20].

Rooted from the mathematical modeling in dynamics of infectious disease from contact and transmission until recovery or death, we developed an index for LSD, taking into account three elements encrypted in the SEIRD model, namely, the force of disease transmission (β), the optimal provision of health care service reflected by the recovery rate (γ), and the critical care capacity captured by the case-fatality rate (τ). Specifically, this can be derived by:

P (Recovery | Infected COVID-19 cases) **(5)**

which can be decomposed into:

P(Recovery| Infected COVID-19 cases and still survive) × **(6)**

P(Survive | Infected COVID-19 cases)

where the first part is the recovery rate and the second part corresponds to the complement of the case-fatality rate. By using the cumulative frequencies specified by equations 2-4, equation 6 can be written as:

which is equivalent to:

An index for LSD is extended by taking the inverse of equation 8 to have:

The reason for subtracting 1 is that, in an ideal scenario, the aforementioned ratio would reach 1 when all confirmed cases have been recovered without death (case-fatality rate ), and therefore, LSD would approach zero, suggesting the region has a full recovery after the outbreak of COVID-19 and may return to the normal status.

Derivation of LSD From Empirical Data

To fit the formula of LSD(t) in equation 9 with the empirical data, the numerator would be estimated by the cumulative number of cases up to time *t* in empirical data, and the denominator would be derived by the cumulative number of recoveries and the case-fatality rate based on the corresponding date. An index for LSD for time *t* following equation 9 can be explicitly described as follows:

LSD is, thus, the ratio of cumulative confirmed cases to cumulative recovered patients without dying from COVID-19 that is captured by (1 – case fatality) – 1 during a fixed time period.

However, it is impracticable to lift social distancing until the value of LSD reaches 0. One has to consider the balance between the spread of COVID-19, the rate of recovery, and the capacity of critical care. The first element is to capture the information on the rate of the COVID-19 spread after the implementation of social distancing that is often modeled by the R_0. The second element is dependent on whether health care systems have the capacity to offer a number of beds for hospitalization. The third one is determined by the capacity of critical care that can stop the progression from acute respiratory distress syndrome to death.

It should be noted that three factors are affected by each other and may also allow for other related factors beyond the previously mentioned determinants. For example, if there is an increase in COVID-19 cases due to LSD and if hospital beds have been occupied for other reasons (trauma, elective surgeries, etc), the rate of recovery would be slow, and the case-fatality rate may be higher. This can be captured by the proposed index. Here, we also made the recommendation for LSD using this index. If the value of the LSD index is greater than 1, it is still necessary to maintain social distancing because the rate of the COVID-19 spread still outweighs the affordable capacity of hospitalization and critical care. If it is less than 1, LSD can be suggested. The degree of an LSD less than 1 was assessed by the inverse of the decile of the LSD index from the lowest decile (0.1) to the highest decile (1). The multinomial distribution using the reported number of confirmed cases, recoveries, and deaths in conjunction with the Bayesian Markov chain Monte Carlo method was used for the derivation of 95% credible interval (CI) for the LSD index.

External Applications of LSD Index to Emerging Infectious Disease

Other emerging infectious disease data including severe acute respiratory syndrome (SARS), Middle East respiratory syndrome–related coronavirus (MERS-CoV), and Ebola were used for the validation of the developed LSD index. The daily reported cases, recoveries, and deaths for the SARS outbreak in 2003 were collected from the Taiwan Centers for Disease Control. Data for the MERS-CoV outbreak in South Korea in 2015 were retrieved from published literature [21]. Data for the Ebola outbreak in the Democratic Republic of Congo in 2018 were derived from the WHO report [22].

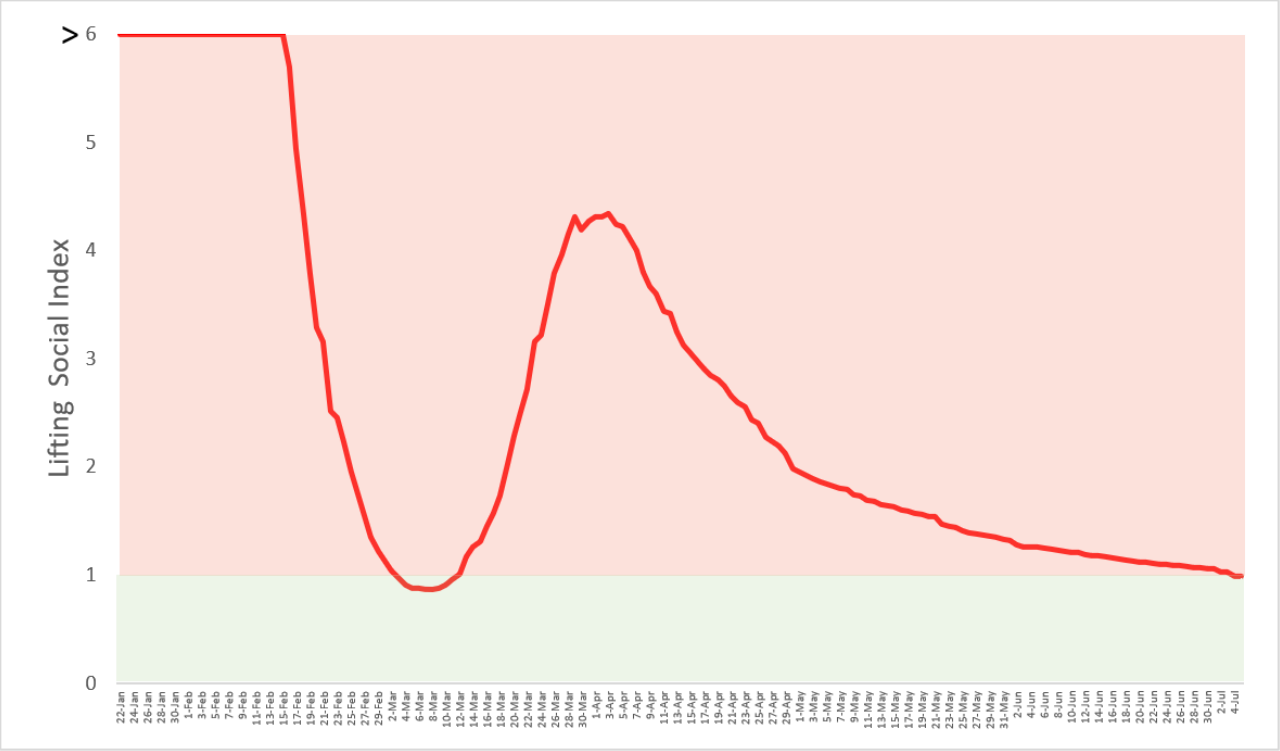
Results

The Daily LSD Index for Countries and Regions Worldwide

As of July 5, 2020, 11,388,537 confirmed cases, 6,445,646 recoveries, and 533,638 deaths from COVID-19 were reported worldwide, which gave a case-fatality rate of 4.7% and a recovery rate of 57%.

Figure 1 shows the daily temporal trend of the global LSD index. The global daily LSD index began with 40.84-11.63 between the last week of January and the second week of February, mainly from Mainland China. In the third week of February, it declined but was still large, up to 5.69 (95% CI 5.65-5.73), mainly due to the spread from China to other hot spots including South Korea, Iran, and Italy. The period from January to mid-March, just before the WHO categorized the COVID-19 pandemic as being in the "pre-pandemic phase." After that, the value of LSD peaked again on March 29 (4.27, 95% CI 4.26-4.28) due to a large-scale spread from continental regions including Australia, the United States and Canada (North America), the European Union except Italy, South Africa, and South Asia. The period from mid-March to June was called the "pandemic phase." During the pandemic phase, every country had adopted various containment measures, which rendered the value of LSD to gradually decline and fall to 1.11 on June 21 and further decrease to 0.99 by July 5. The phase after June is called the "post-pandemic phase." The LSD index took 115 days to go down to 1 at the global level.

Figure 1. Temporal trend of global lifting social distancing (LSD) index up to July 5, 2020. Overall: The LSD index ranged between 40.1 and 0.96 in the prepandemic period from January to mid-March. In the pandemic period from mid-March to June, the peak of the LSD index reached 4.27 on March 29. As of July 5, the LSD index declined to less than 1.



It should be taken with great caution that the value of LSD was below 1 on July 5, 2020, as the overall curve consisted of most countries with an LSD value smaller than 1 (65%), and the remaining countries (35%) were larger than 1. It is, therefore, necessary to stratify the overall curve into two types according to the LSD value greater than or less than 1 in the previous week of July 5. For countries with LSD values less than 1, it took 55 days, on average, to enable LSD. At the global level, Figure 2 shows that the spread of COVID-19 still exceeded the capacity of hospitalization and critical care for COVID-19 cases, whereas Figure 3 shows the opposite. These heterogeneous findings across countries and regions also suggests that each country and region may experience three phases (pre-epidemic or cluster infection, epidemic, and postepidemic phases) commensurate with the previously mentioned three corresponding phases of the COVID-19 pandemic in different time periods.

According to the decile of the LSD index for each of the two categories (≥1 and <1), Figure 4 shows the frequencies on the index of LSD until July 5, 2020, for the countries and regions worldwide, aggregated by three groups for the LSD<1, including <0.1 (n=23), 0.1-0.4 (n=67), and 0.5-1 (n=30), and by five groups for the LSD ≥1, including, 1.1-1.4 (n=22), 1.5-1.9 (n=14), 2-2.4 (n=10), 2.5-2.9 (n=7), and ≥3 (n=13). Of the 186 countries and region, 64.5% (n=120) of the LSD indices were less than 1.

Figure 2. Temporal trend of global lifting social distancing (LSD) index up to July 5, 2020. A total of 66 countries with LSD≥1 on June 28 in the pandemic phase period. The social distancing was relaxed over 4 months.

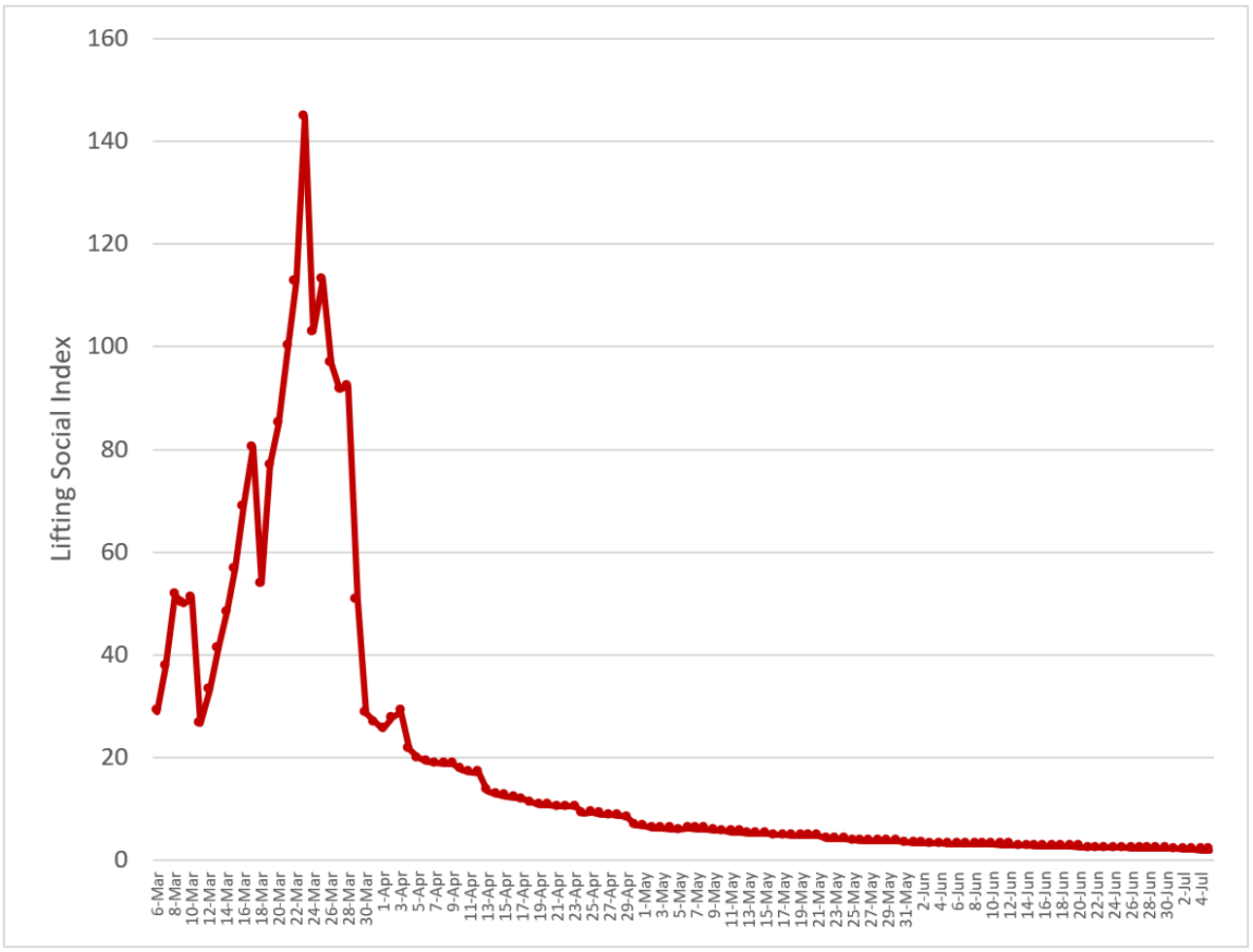


Figure 3. Temporal trend of global lifting social distancing (LSD) index up to July 5, 2020. A total of 120 countries with LSD<1 on June 28 in the pandemic phase. Between March 13 and May 6, the LSD index was greater than 1. The time required for lifting social distancing was 55 days on average.

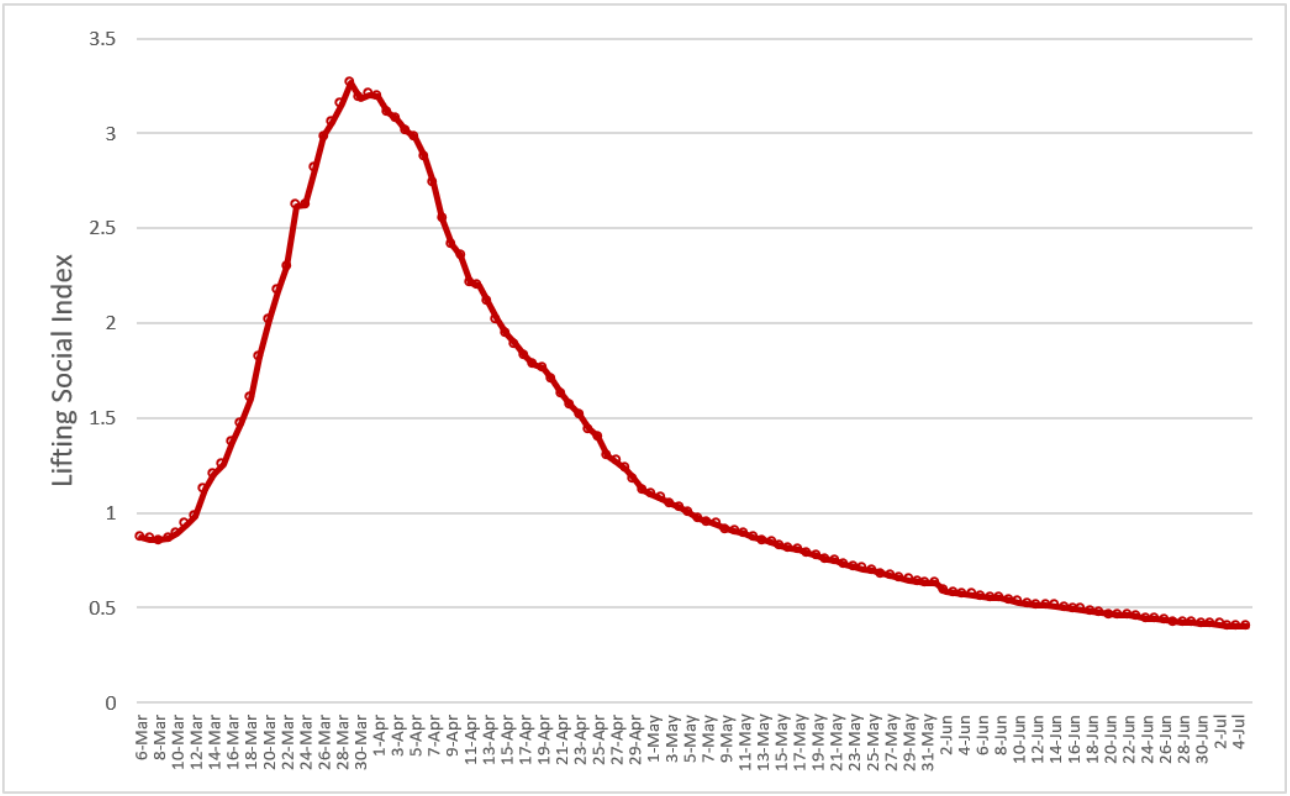
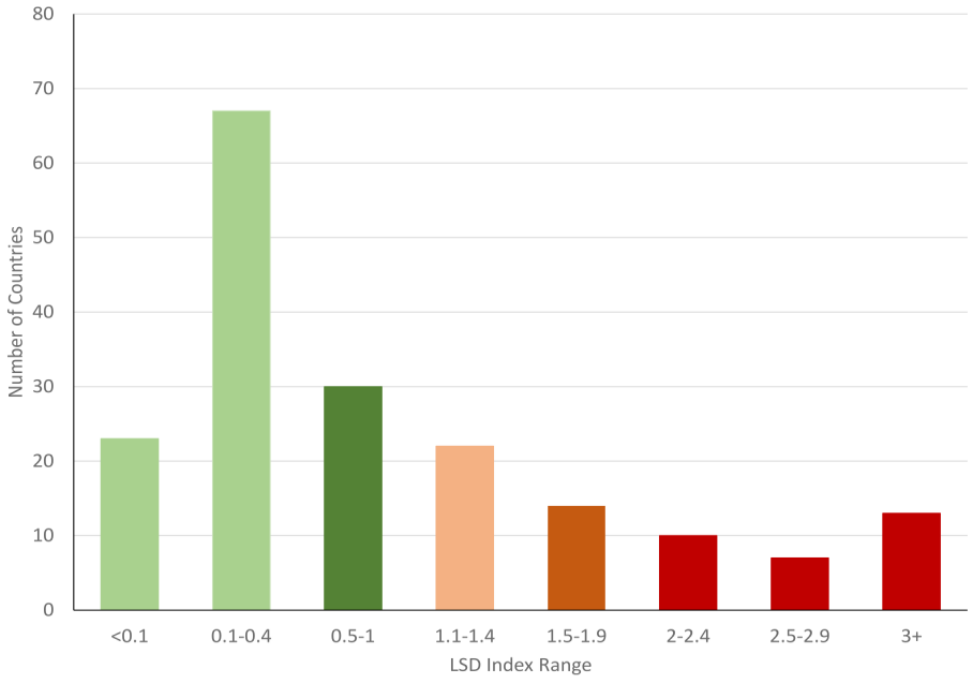


Figure 4. The number of countries and regions by ranges of the LSD index. As of July 5, 2020, the LSD index in 120 countries and regions was less than 1. The LSD index was greater than 1 in 66 countries and regions. LSD: lifting social distancing.



Dynamic Change of Weekly LSD Index Worldwide

Multimedia Appendix 2 shows the dynamic change of the global LSD index (in weeks) for 186 countries and regions from January 26 (fifth week) to July 5, 2020 (28th week), on the map worldwide. Countries that are more red are less likely to lift social distancing, whereas those that are more green are more likely to lift social distancing. Different countries had different times to transit from high (greater than 1) to low (less than 1) LSD values, representing different times required for LSD as indicated in three phases of the COVID-19 epidemic in each country or region. For example, South Korea's pre-epidemic phase with cluster infections due to religious gatherings was between early February and mid-February. The epidemic phase with large-scale community-acquired outbreaks was between mid-February and March. The postepidemic phase was after March. Various regions had experienced these three phases with different time windows as detailed in Multimedia Appendix 2. The details of the global dynamic change consisting of each country and region have been delineated as follows.

The LSD value in Western European regions had changed from greater than 1 (red) on March 31, 2020, to less than 1 (green) on July 5. However, the evolution and three factors' contribution were heterogeneous before the decline of the LSD value to being less than 1 between the first week and the fifth week in May. As of the end of May, Germany's (a lower LSD index of 0.163, 95% CI 0.160-0.166) moderate case-fatality rate (4.7%) was compensated by the high recovery rate (90%), possibly through early detection followed by the high capacity for hospitalization. Countries' high LSD indices, such as Hungary (LSD: 1.088, 95% CI 1.063-1.114), were attributed to high case-fatality rates (13.6%), mainly resulting from an insufficient capacity for critical care. The high LSD index was, to a greater extent, due to the case-fatality rate (around 2.4 times the average worldwide) and, to a lesser extent, due to a modest recovery rate (55.4%, around 1.3 times the average worldwide). However, the overall value of LSD has declined to 1 since mid-June. Similar findings have been noted in other regions of Europe with different weekly times to see the drop of LSD, including around the fourth week for Northern and Southern Europe, and the fifth week for Eastern Europe. However, the LSD index for a few countries in Southeast Europe was still greater than 1 until mid-June.

Even when there were large-scale outbreaks from mid-March to mid-April that led to the value of LSD being greater than 1 in two Oceania countries including New Zealand and Australia, their values of LSD decreased to less than 1 after both had controlled the outbreak due to multiple containment measures including restricted border control, quarantine, and planned patient care. In addition, the better recovery rate and lower case fatality were also achieved by the optimal management of patients with COVID-19 and the sufficient capacity for critical care.

In the early phase of the COVID-19 pandemic, the spread of SARS-CoV-2 had started from Northeast Asia including South Korea and Japan, and then had subsequent outbreaks in most of the countries in Southeast Asia and South Asia, with the value of LSD greater than 1 in Southeast Asia since mid-March. Some countries including Vietnam, Laos, Cambodia, Thailand, and Myanmar had an LSD value less than 1 after April 19. However, the values of LSD for a few countries such as Indonesia, Philippine, Nepal, India, and Bangladesh were still greater than 1 until July 5.

In Africa, the spread of COVID-19 started in mid-March. After 3 months, the value of LSD was still greater than 1 in most of the countries in Africa until July 5. Similar findings were observed in South Africa, probably resulting from insufficient health care personnel and capacity for critical care.

The LSD index in the United States has not yet reached less than 1 because of 5 states where the impact of high transmission from SARS-CoV-2 and a low recovery rate resulted in higher LSD values in the United States, which has been demonstrated in several states with high long-standing LSD values including New York state, even though the case-fatality rate (6.4%) was comparable to the average worldwide.

Time for LSD in selected countries with an LSD index less than 1 are summarized in Table 1. New Zealand took 22 days as the shortest time to reach an LSD index less than 1. The time for the LSD index to be less than 1 was estimated as 46-51 days for South Korea, Germany, and Australia. Japan and Hungary took longer (around 90 days). Social distancing was lifted while the LSD index was less than 1 in most countries except Finland (1.03), Hungary (1.21), Italy (1.07), Russia (1.03), and Canada (1.03). The Philippines and the United States were too early to lift social distancing. Both of the LSD indices in these two countries were greater than 4.

Table 1. Time required and LSD index for LSD in selected countries.

Country	Outbreak date	Date when LSD[a]<1	Time required for LSD (days)	On the date of LSD<1				LSD	
				COVID-19 cases	Recovered cases	Death cases	LSD index	Date	LSD index
New Zealand	March 24	April 15	22	1366	628	9	0.92	April 28	0.22
Vietnam	March 11	April 8	28	251	126	0	0.99	April 24	0.23
Iceland	March 10	April 12	34	1701	889	8	0.92	May 5	0.05
Austria	March 9	April 13	36	131	2	0	64.5	April 28	0.27
Netherlands	March 19	April 27	40	38,440	22,176	4534	0.97	June 2	0.40
Turkey	March 25	May 4	40	127,659	68,166	3461	0.92	June 1	0.31
China	January 22	March 1	41	79,932	42,162	2872	0.97	April 4	0.12
Switzerland	March 3	April 13	41	25,688	13,700	1138	0.96	June 22	0.15
Denmark	March 6	April 18	44	7437	4031	346	0.94	June 8	0.16
Ireland	March 17	April 29	44	20,253	13,386	1190	0.61	May 18	0.33
Australia	March 1	April 17	46	6522	3808	66	0.73	April 28	0.21
Germany	February 27	April 14	47	131,359	68,200	3294	0.98	June 2	0.16
South Korea	February 7	March 29	51	9583	5033	152	0.95	May 6	0.18
Malaysia	February 27	April 15	53	5072	2647	83	0.95	June 9	0.21
Belgium	March 1	May 2	57	49,517	29,418	7765	0.99	May 4	0.86
Finland	February 26	April 29	57	4395	2500	177	0.83	April 15	1.03
Thailand	January 26	April 14	61	2613	1405	41	0.89	May 17	0.08
Israel	February 27	April 29	62	15,834	8233	215	0.95	May 4	0.58
France	February 6	April 26	69	162,220	98,853	22,859	0.91	May 11	0.50
Spain	February 25	May 6	71	220,325	126,002	25,857	0.98	June 20	0.85
Sweden	February 26	May 9	73	25,921	14,957	3220	0.98	__[b]	—
Hungary	March 14	June 10	88	198,811	69,957	23,633	0.98	May 28	1.21
Italy	February 22	May 20	89	228,006	134,560	32,486	0.93	May 18	1.07
Japan	January 26	May 11	90	450,249	94,167	13,001	0.99	May 25	0.28
Singapore	February 8	May 26	90	32,343	16,444	23	0.97	June 6	0.55
Russia	March 5	June 10	97	493,023	6350	252,295	0.98	June 9	1.03
Canada	February 12	June 2	112	94,641	51,506	7579	0.99	June 1	1.03
Philippines	—	—	—	—	—	—	—	May 15	4.27
United States	—	—	—	—	—	—	—	May 1	6.18

[a]LSD: lifting social distancing.

[b]Not available.

Resurgence of the COVID-19 Epidemic in Local Communities and Regions

The LSD index can also be applied to evaluating whether social distancing has to be re-executed in local cities and regions because of COVID-19 outbreak resurgences due to small cluster infections. In the United States, states like Florida and Texas saw a resurgence of the COVID-19 epidemic after LSD. Using these two states as examples, although the overall LSD until the end of April was 0.83 (Table 2), the LSD indices from June 27 to July 4 were estimated as 0.94 and 1.15, respectively, indicating that the outbreak was re-emerging and might call for a restrengthening of social distancing measures. Similar resurgence can be noted in Texas.

Table 2. Resurgence of COVID-19 epidemic in Florida and Texas.

Date	Florida				Texas			
	COVID-19 cases	Recovered cases	Death cases	LSD[a] index	COVID-19 cases	Recovered cases	Death cases	LSD index
April 11	18,494	438	3325	4.70	12,561	254	1617	6.93
April 18	25,269	754	10,357	1.51	18,260	453	4806	2.90
April 25	30,839	1075	17,419	0.83	23,773	623	9986	1.44
May 02	35,463	1388	23,881	0.55	30,522	847	14,891	1.11
May 09	40,001	1785	29,054	0.44	37,860	1049	20,141	0.93
May 16	44,811	2040	33,423	0.40	46,999	1305	26,601	0.82
May 23	50,127	2312	37,689	0.39	54,509	1506	32,277	0.74
May 30	55,424	2530	42,281	0.37	62,338	1626	40,068	0.60
June 06	62,758	2773	47,354	0.39	73,553	1819	48,895	0.54
June 13	73,552	3016	52,408	0.46	86,011	1957	56,535	0.56
June 20	93,797	3237	59,521	0.63	103,305	2140	65,329	0.61
June 27	132,545	3489	70,063	0.94	143,371	2366	78,248	0.86
July 04	190,052	3803	89,994	1.15	191,790	2608	97,430	1.00

[a]LSD: lifting social distancing.

The LSD Index for Other Emerging Infectious Diseases

The LSD index was also applied in a case of a SARS outbreak. During the 2003 outbreak of SARS in Taiwan, a total of 664 probable cases with 81 deaths were reported in the epidemic period between February 25 and June 15, 2003 (Figure 5). The majority of infected cases were imported as a minor outbreak in the initial period before April. The LSD index was as high as 2.5 on March 22 then down to 0.3 on April 15, 2003. Since April 22, 2003, the probable cases increased, and the LSD index rose to be greater than 1. The source of the outbreak was associated with health care settings. Several nosocomial SARS clusters were reported April 22-May 22, 2003. By a fever screen at emergency departments, self-quarantine, isolation, and containment of health care transmission, the spread of SARS was finally stopped. A low LSD index of 0.3 indicated that lifting of social distancing could have been considered at the end of May 2003.

Figure 5. LSD index for severe acute respiratory syndrome outbreak in Taiwan, 2003. The 2.5 LSD index was high on March 22, 2003, at the initial epidemic stage then decreased to 0.3 on April 15, 2003. After April 22, the LSD index started increasing from 1 by nosocomial infection. The lower LSD index (<0.5) indicated that social distancing could have been lifted at the end of May. LSD: lifting social distancing.

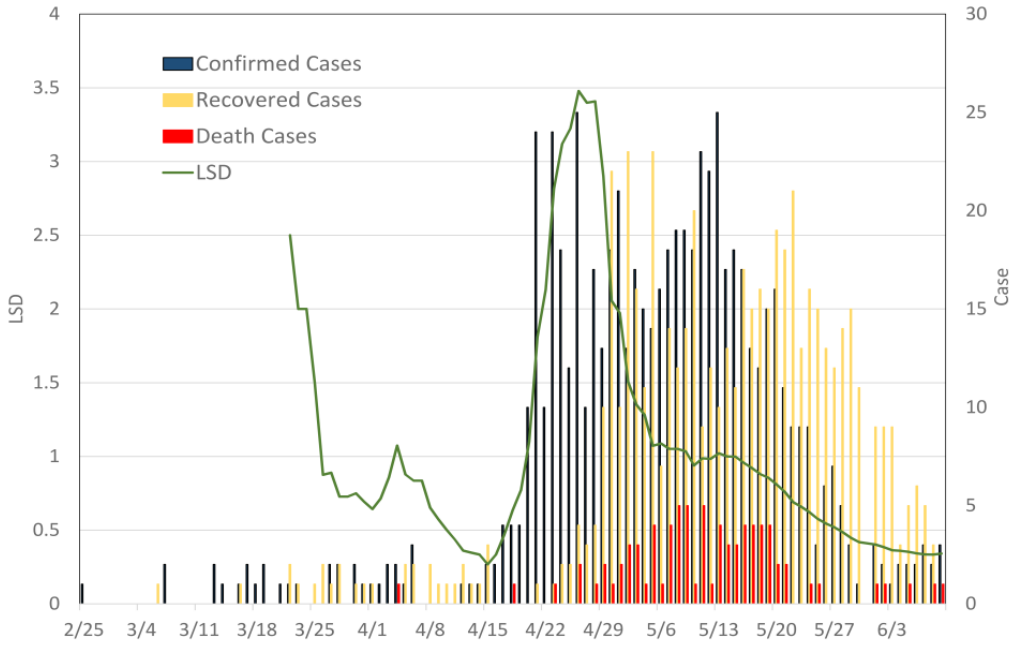


Figure 6 shows the estimated results of the LSD index applied to the 2015 MERS-CoV outbreak in South Korea. The outbreak started on May 20, 2015, with an imported case of MERS-CoV. Following the index case, the MERS-CoV outbreak in South Korea lasted until August 5, 2015, yielding 186 cases and 36 deaths. The LSD index for this outbreak was estimated since June 16, 2015, at the initial stage. The LSD index became less than 1 on July 1, 2015, around 2 weeks after the peak of 6.8 on June 17, 2015, and was kept at a low value until the end of the outbreak on August 7, 2015. This decreasing trend in LSD was attributable to the increasing recovery rate, which reached 80% after July 19, 2015, and the controlled case-fatality rate of 19% during the same period.

Figure 6. LSD index for the Middle East respiratory syndrome–related coronavirus outbreak in South Korea, 2015. The LSD index greater than 6 was estimated on June 16, 2015. After about 2 weeks, the LSD index was less than 1 on July 1, 2015, and kept at a low value until the end of the outbreak. LSD: lifted social distancing.

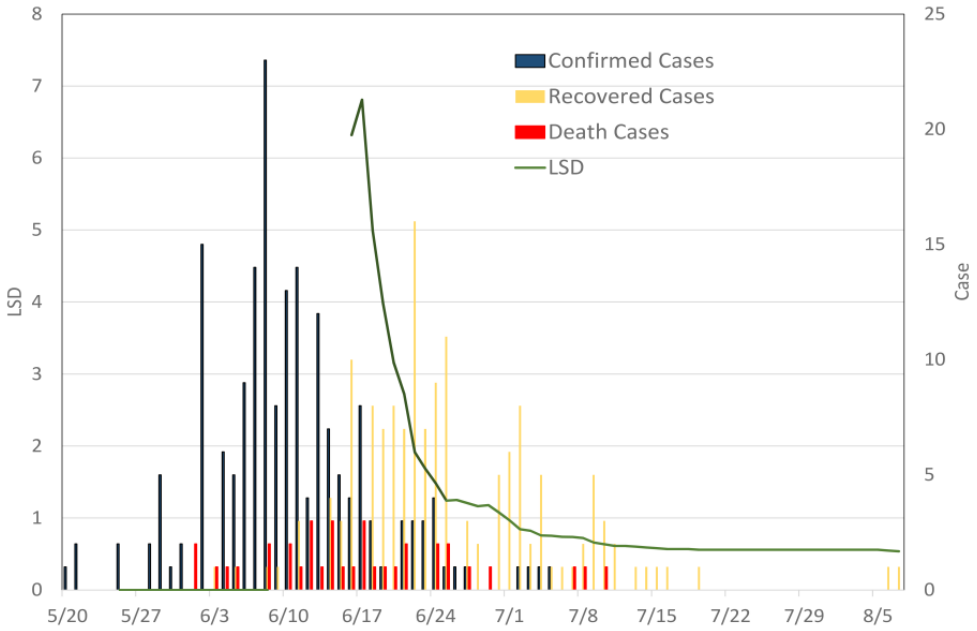
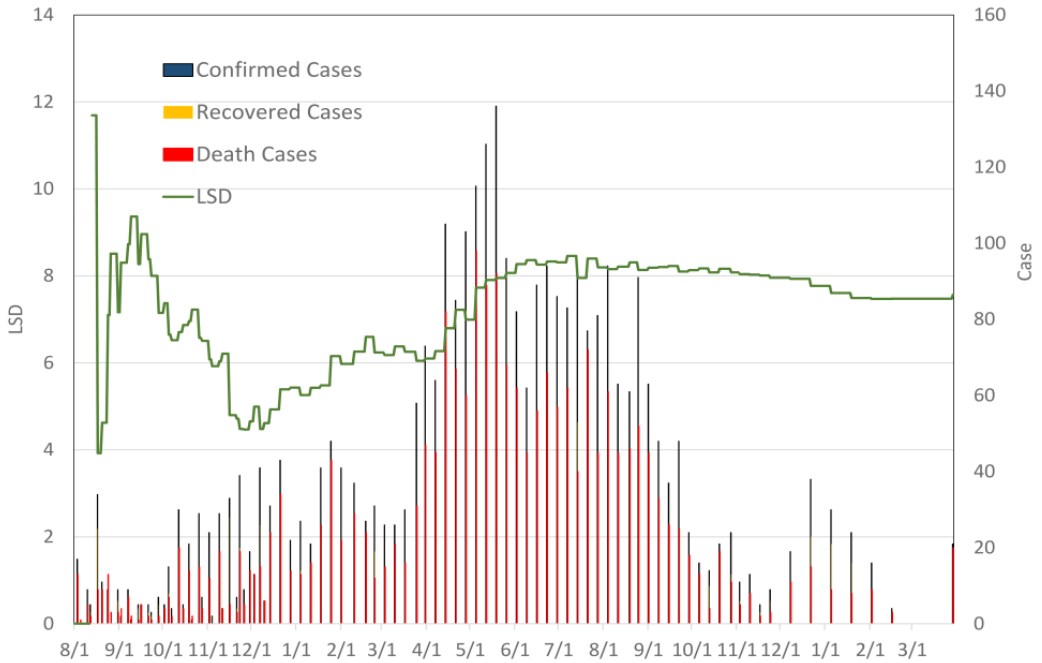


Figure 7 shows the estimated results of the LSD index along with the frequencies of cases, deaths, and recoveries for the Ebola outbreak in the Democratic Republic of Congo since mid-2018. The end of the outbreak was declared on June 25, 2020, when the 42-day lapse between the last case was reached. The LSD index estimated on August 12, 2018, was 11.7 at the initial stage. As the Ebola epidemic evolved, the LSD index never became less than 1. This was not only due to its high case-fatality rate (55-79%) but also its lower recovery rate (<45%) through the whole epidemic. This high level of LSD revealed less effectiveness to depend on nonpharmaceutical approaches for containing a disease as lethal as Ebola.

Figure 7. LSD index for the Ebola outbreak in the Democratic Republic of the Congo, 2018. The LSD index was estimated as 11.7 at the initial stage on August 12, 2018. The LSD index never became less than one. LSD: lifting social distancing.



Discussion

The COVID-19 pandemic has evolved from the dawn of large-scale outbreaks followed by large-scale social distancing plans to contain the COVID-19 epidemic. Incorporating information on the effective social distancing interventions to reduce the contacts of adults and avert the hospitalizations and deaths into mathematical models for guidance on containment measures has been suggested [17,18]. However, to revive economic and social activity, LSD is still necessary and requires a simple index for guiding LSD particularly when a travel ban has been lifted across borders and regions. We propose a simple index for LSD for global, country, region, and community levels to elucidate the global dynamic change of three factors in relation to COVID-19 through three phases (the large-scale outbreak period [between January and early March], the pandemic period [between mid-March and June], and the postpandemic period [starting in June]) and the corresponding dynamic changes at the country and region level through three epidemic phases (pre-epidemic, epidemic, and postepidemic) of COVID-19 but with different time windows in each specific country and region. The LSD index has been applied to monitoring the resurgence of small to moderate-small cluster infections after LSD. There are three main merits of using this LSD index. First, the LSD index is informative in providing a quantitative assessment for whether and when to lift social distancing. Second, the LSD index identifies which factor of the three determinants may contribute more to a higher value of LSD than others, which provides information for health decision makers to reinforce this fragile determinant for shortening the time until the lifting of social distancing. Third, the LSD index can be flexibly applicable to various scenarios from local, country, and region to the global community during the postpandemic period and accommodate the regions that have already lifted social distancing but have seen resurgences of outbreaks due to cluster infections.

The dynamic change of weekly LSD indices on a global level is informative to get a better understanding of the time required for the change from LSD≥1 to LSD<1. The longer time required for LSD, the less likely the countries or regions will be resilient enough to recover from the COVID-19 pandemic. On the other hand, the shorter the time to reach an LSD less than 1, the more prosperous traveling business can be. The time required for countries or regions to change the value of LSD from greater than 1 to less than 1 ranged from 3 weeks to more than 4 months. Moreover, the LSD cutoffs for lifting can be rectified according to the fragile degree of public health in each country and region. The more fragile the public health system, the lower the suggested LSD cutoff.

Few countries took only 3 weeks for LSD, such as New Zealand. The countries that required 1-2 months for LSD were China, South Korea, and Vietnam in Asia; Iceland, Denmark, Austria, Ireland, Netherlands, Switzerland, Germany, and Turkey in Europe; and Australia in Oceania. By contrast, Thailand and Malaysia in South Asia; Israel in Western Asia; and Belgium, Finland, France, Sweden, and Spain in Europe required 2-3 months for LSD. Japan and Singapore in Asia and Italy, Hungary, and Russia in Europe required 3-4 months for LSD. Canada required 4 months. The time required for LSD is heterogeneous and is highly dependent on three determinants. In addition to maintaining containment measures, the reinforcement of the medical resource capacity is also essential to respond to the outbreak [20,23] when LSD.

For resource-limited countries, the LSD index did not decline much due to the insufficient medical resources. Following the WHO's guidance [19] for patient triage and referral during community transmission will help health facilities cope with the COVID-19 pandemic to shorten the time for LSD. It is interesting to note that the LSD index for Ebola was always greater than 1, indicating that the outbreak was not able to lift social distancing until most of the infected died. This strongly suggested the necessity of discovering the vaccine for containing the epidemic of Ebola.

There is one limitation of using the LSD index. Complete and accurate information on confirmed cases, recoveries, and deaths is required for calculating the LSD index. There are some countries lacking the information on recoveries, which limits the use of the LSD index.

In summary, a simple index for LSD was developed to evaluate the evolution of this index through three phases, prepandemic, pandemic, and postpandemic, on the global level. The surveillance of LSD indices on the country and region level aids health policy makers in being aware of which epidemic phase they are in and assessing whether and when to lift social distancing in the postpandemic period. Decomposition of this index into three components also gives a clue to facilitating the process of LSD by improving the key determinant identified from the LSD index. The proposed LSD index for LSD can be applied to not only the current COVID-19 outbreaks but also other emerging infectious diseases in the future.

Acknowledgments

This work was funded by the Ministry of Science and Technology, Taiwan (MOST 108-2118-M-002-002-MY3; MOST 108-2118-M-038-001-MY3; MOST 108-2118-M-038-002-MY3; MOST 109-2327-B-002-009).

Authors' Contributions

SLSC, AMFY, and THHC conceptualized and designed the framework of analysis. AMFY, CCL, and CYH abstracted and prepared the data for analysis. AMFY, CCL, and CYH were responsible for the statistical analysis. SLSC, AMFY, CYH, and THHC developed the methods. SLSC, CCC, and THHC interpreted the results. All authors agreed to the findings and provided input on the revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 74 KB - jmir_v22i9e22469_app1.docx]

Multimedia Appendix 2
Dynamic change of global weekly lifting social distancing (LSD) index. Temporal change of global LSD index with 9 categories for 186 countries and regions from January 26 (fifth week) to July 5, 2020 (28th week). NA: not available.
[PNG File , 4936 KB - jmir_v22i9e22469_app2.png]

References

1. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID‐19: a rapid review. Cochrane Database Syst Rev 2020 Apr 08:CD013574. [doi: 10.1002/14651858.cd013574]
2. Wu J, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/s0140-6736(20)30260-9]
3. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020 May 22;368(6493):860-868 [FREE Full text] [doi: 10.1126/science.abb5793] [Medline: 32291278]
4. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
5. Angulo FJ, Finelli L, Swerdlow DL. Reopening society and the need for real-time assessment of COVID-19 at the community level. JAMA 2020 Jun 09;323(22):2247-2248. [doi: 10.1001/jama.2020.7872] [Medline: 32412582]
6. WHO Director-General's opening remarks at the media briefing on COVID-19 - 13 April 2020. World Health Organization. 2020 Apr 13. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--13-april-2020 [accessed 2020-07-12]
7. Tang B, Scarabel F, Bragazzi NL, McCarthy Z, Glazer M, Xiao Y, et al. De-escalation by reversing the escalation with a stronger synergistic package of contact tracing, quarantine, isolation and personal protection: feasibility of preventing a COVID-19 rebound in Ontario, Canada, as a case study. Biology (Basel) 2020 May 16;9(5):100 [FREE Full text] [doi: 10.3390/biology9050100] [Medline: 32429450]
8. Barton CM, Alberti M, Ames D, Atkinson J, Bales J, Burke E, et al. Call for transparency of COVID-19 models. Science 2020 May 01;368(6490):482-483. [doi: 10.1126/science.abb8637] [Medline: 32355024]
9. Vespignani A, Tian H, Dye C, Lloyd-Smith JO, Eggo RM, Shrestha M, et al. Modelling COVID-19. Nat Rev Phys 2020 May 6;2(6):279-281. [doi: 10.1038/s42254-020-0178-4]
10. Viceconte G, Petrosillo N. COVID-19 R0: magic number or conundrum? Infect Dis Rep 2020 Feb 25;12(1):8516. [doi: 10.4081/idr.2020.8516] [Medline: 32201554]
11. Lai C, Jiang D, Wu H, Chen H. A dynamic model for the outbreaks of hand, foot, and mouth disease in Taiwan. Epidemiol Infect 2015 Nov 16;144(7):1500-1511. [doi: 10.1017/s0950268815002630]
12. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
13. COVID19Tracking / covid-tracking-data. GitHub. URL: https://github.com/COVID19Tracking/covid-tracking-data/blob/master/data/states_daily_4pm_et.csv [accessed 2020-07-12]
14. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-07-12]
15. Daley DJ, Gani J. Epidemic Modelling: an Introduction. Cambridge, England: Cambridge University Press; 1999.
16. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 2020 May 04:A. [doi: 10.1038/s41586-020-2293-x] [Medline: 32365354]
17. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science 2020 Jul 10;369(6500):eabb9789 [FREE Full text] [doi: 10.1126/science.abb9789] [Medline: 32414780]
18. Matrajt L, Leung T. Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. Emerg Infect Dis 2020 Aug;26(8):1740-1748. [doi: 10.3201/eid2608.201093] [Medline: 32343222]
19. Algorithm for COVID-19 triage and referral : patient triage and referral for resource-limited settings during community transmission. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/331915 [accessed 2020-08-16]

20. Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA 2020 Apr 28;323(16):1545-1546. [doi: 10.1001/jama.2020.4031] [Medline: 32167538]
21. Yang Y, Hsu C, Lai C, Yen M, Wikramaratna PS, Chen H, et al. Impact of comorbidity on fatality rate of patients with Middle East respiratory syndrome. Sci Rep 2017 Sep 12;7(1):11307. [doi: 10.1038/s41598-017-10402-1] [Medline: 28900101]
22. Weekly bulletin on outbreaks and other emergencies: week 32: 3 - 9 August 2020. World Health Organization. 2020 Aug 09. URL: https://apps.who.int/iris/bitstream/handle/10665/333769/OEW32-0309082020.pdf [accessed 2020-08-16]
23. Faccincani R, Pascucci F, Lennquist S. How to surge to face the SARS-CoV-2 outbreak: lessons learned from Lombardy, Italy. Disaster Med Public Health Prep 2020 Apr 01:1-3 [FREE Full text] [doi: 10.1017/dmp.2020.64] [Medline: 32234108]

Abbreviations

CI: credible interval
CSSE: Center for Systems Science and Engineering
LSD: lifting social distancing
MERS-CoV: Middle East respiratory syndrome–related coronavirus
R_0: basic reproductive number
SARS: severe acute respiratory syndrome
SEIRD: susceptible-exposed-infected-recovered-death
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.07.20; peer-reviewed by D Hu, LF Soualmia; comments to author 04.08.20; revised version received 25.08.20; accepted 04.09.20; published 17.09.20.

Please cite as:
Chen SLS, Yen AMF, Lai CC, Hsu CY, Chan CC, Chen THH
An Index for Lifting Social Distancing During the COVID-19 Pandemic: Algorithm Recommendation for Lifting Social Distancing
J Med Internet Res 2020;22(9):e22469
URL: http://www.jmir.org/2020/9/e22469/
doi:10.2196/22469
PMID:32886622

Original Paper

Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study

Xiaomei Wang[1], BS, PhD; Sudeep Hegde[1], BEng, MS, PhD; Changwon Son[1], BS, MS; Bruce Keller[1], BA; Alec Smith[1], BS; Farzan Sasangohar[1,2], BA, BCS, MASc, SM, PhD

[1]Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[2]Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, United States

Corresponding Author:
Farzan Sasangohar, BA, BCS, MASc, SM, PhD
Industrial and Systems Engineering
Texas A&M University
Emerging Technologies Building, 3131 TAMU
101 Bizzell Street
College Station, TX, 77843
United States
Phone: 1 9794582337
Email: sasangohar@tamu.edu

Abstract

Background: Evidence suggests that the COVID-19 pandemic has generally increased levels of stress and depression among the public. However, the impact on college students in the United States has not been well-documented.

Objective: This paper surveys the mental health status and severity of depression and anxiety of college students in a large university system in the United States during the COVID-19 pandemic.

Methods: An online survey was conducted among undergraduate and graduate students recruited from Texas A&M University via email. The survey consisted of two standardized scales—the Patient Health Questionnaire-9 and the General Anxiety Disorder-7—for depression and anxiety, and additional multiple-choice and open-ended questions regarding stressors and coping mechanisms specific to COVID-19.

Results: Among the 2031 participants, 48.14% (n=960) showed a moderate-to-severe level of depression, 38.48% (n=775) showed a moderate-to-severe level of anxiety, and 18.04% (n=366) had suicidal thoughts. A majority of participants (n=1443, 71.26%) indicated that their stress/anxiety levels had increased during the pandemic. Less than half of the participants (n=882, 43.25%) indicated that they were able to cope adequately with the stress related to the current situation.

Conclusions: The proportion of respondents showing depression, anxiety, and/or suicidal thoughts is alarming. Respondents reported academic-, health-, and lifestyle-related concerns caused by the pandemic. Given the unexpected length and severity of the outbreak, these concerns need to be further understood and addressed.

(J Med Internet Res 2020;22(9):e22817) doi:10.2196/22817

KEYWORDS

mental health; online survey; COVID-19; coronavirus; college student; student; stress; depression; university

Introduction

The United States has seen a surge in the number of COVID-19 cases since March 2020, with initial peaks in April 2020 [1]. Recent assessments of mental health in the general populace of China and Iran, countries that had major outbreaks, show increased levels of stress due to the pandemic [2,3]. A key concern during the pandemic relates to the mental health of vulnerable populations, including college students. The 2019 Annual Report of the Center for Collegiate Mental Health [4] reported that anxiety continues to be the most common problem (62.7% of 82,685 respondents) among students who completed the Counseling Center Assessment of Psychological Symptoms. Consistent with the national trend, Texas A&M University has seen an increase in the number of students seeking services for anxiety disorders over the last few years. Given the vulnerability of this population during the pandemic, there is a critical need

to assess the mental health of college students in order to address concerns in a timely manner [5-8].

Recent assessments of college student mental health in China have shown an increased level of anxiety and depression in the wake of the pandemic [6,9]. These studies used standardized scales for depression and anxiety, such as the depression and anxiety stress scale, the Patient Health Questionnaire-9 (PHQ-9), and the Generalized Anxiety Disorder-7 (GAD-7) questionnaire. General panic related to the outbreak and risk of exposure were found to be contributors to increased level of depression. Similarly, Cao et al [9] assessed anxiety using the GAD-7 scale and found that risk of infection, including for family members, was a major contributor to college students' increased anxiety. Both studies also identified important protective factors such as income stability and the availability of information related to preventative measures. However, these studies do not include an assessment of strategies used by students themselves for coping with and managing their stress. Additionally, these studies have focused on the student population in China. Given the differences in cultural, geographic, economic, and other factors, an assessment of college students' mental health in the United States is needed.

The aim of this study was to conduct a survey-based assessment of mental health among college students at Texas A&M University, a large university in the United States, during the COVID-19 pandemic. We sought to identify severity levels of depression and anxiety symptoms (primary outcome), as well as stressors related to the pandemic, coping mechanisms used, and barriers experienced by students in handling pandemic-related stress.

Methods

Recruitment

An online cross-sectional survey was designed and conducted during the initial peak of the COVID-19 pandemic in the late Spring 2020 semester at Texas A&M. The research received approval from the university's institutional review board. Guidelines provided by Kelley et al [10] were used for designing, conducting, and reporting this survey research.

Participants were recruited from the student population. This university closed all campuses on March 23, 2020, and held all classes virtually in response to the COVID-19 pandemic. In addition, the State of Texas issued a stay-at-home order on April 2, 2020. The survey was published using the online survey platform Qualtrics on May 4, 2020, and data collection remained open until no additional completion was reported for two days (May 19, 2020). During this period, the survey was announced to the entire population of over 60,000 students at the Texas A&M College Station Campus, through email.

Survey Design

The survey was designed in a semistructured format comprising multiple-choice questions and free-text fields for elaboration. The survey consisted of the following five sections:

Demographics

This section included questions related to participants' age, gender, college classification (ie, undergraduate [freshman, sophomore, junior, senior] or graduate [master's, doctorate]), and program of study.

Patient Health Questionnaire

The PHQ-9 is a validated and widely used measure of depression severity in primary and mental health care, consisting of 9 items based on depression symptoms. Respondents report the frequency of symptoms experienced within the last 2 weeks. The categories of severity range are minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27) [11].

Generalized Anxiety Disorder Screener

The GAD-7 is a validated questionnaire used in most mental health care settings as a screening tool for major anxiety disorders such as generalized anxiety disorder or panic disorder [12], consisting of 7 items based on GAD symptoms. Respondents rate the frequency of experiencing these symptoms within the last 2 weeks. The categories of severity range are minimal (0-4), mild (5-9), moderate (10-14), and severe (15-21).

Questions Related to COVID-19–Related Stress

This section aimed at identifying various stressors resulting from the pandemic and comprised the following items:

- Did your overall stress increase/decrease or remain the same during the ongoing pandemic? (response: increase/decrease/remain the same)
  - If the participant chose increase/decrease, a follow-up open-answer question was asked: Can you describe the main reason for such increase/decrease of stress and anxiety?
- Do you think other students are experiencing stress/anxiety because of the pandemic? (response: yes/no)
- In the past month, what level of fear, worry, and/or changes have you experienced related to any of the following academic/health/lifestyle-related concerns? (response: none/mild/moderate/severe)
- Did you have any other academic/health/lifestyle-related concerns? (response: yes/no)
  - If participant chose "yes," a follow-up open-answer question was asked: Please specify any other academic-related concerns you have.

Coping Mechanisms and Barriers

This section consisted of multiple-choice questions and open-ended follow-up questions related to ways in which the students coped with stress during the pandemic. Follow-up questions were included to identify specific resources and technological apps that students may have been using as they coped:

- Do you feel you are able to cope adequately with the stress related to the current situation? (response: yes/no/maybe)
- What coping methods/tools/techniques have you used to mitigate your elevated stress/anxiety? (response: none;

university services; health services outside the university; support from community, family and friends; technologies; other)
    - If participant chose "university services," a follow-up question was asked: What university services did you use because of the pandemic? (response: student health services; counseling and psychological services; other)
- Where are you sourcing your information [about the pandemic] from? (response: university emails; your medical provider; newspaper and periodicals; medical websites; posts on social media; other)
- Have you been using any mobile apps or features on existing apps for managing stress, anxiety, or depression related to the ongoing COVID-19 pandemic? (response: yes/no)
    - If participant chose "yes," a follow-up open-answer question was asked: Please specify what apps/features you have been using.
- In your opinion, what are the barriers to mental health care? (response: none; lack of information about resources available; financial concerns; limited access to the services; social stigma; other)

Data Analysis

For PHQ-9 and GAD-7, mean scores were calculated for different gender and classification groups. The percentages of participants who fall in each severity category were computed. Inspection of the data and residual plots for mean PHQ-9 and GAD-7 scores did not indicate any violation of assumptions of normality, independence, and homogeneity of variance. Therefore, a two-way analysis of variance (ANOVA) was conducted to identify significant main effects and interactions between gender and classification groups. For the questions regarding concerns about COVID-19, the percentage of participants who chose each severity level was computed. For the multiple-choice questions regarding coping mechanisms and barriers, the percentage of participants who chose each item was computed. Quantitative analyses were performed using Microsoft Excel (Microsoft Corp) and R 4.0.2 (The R Foundation).

Open-ended questions were coded using thematic analysis [13]. Initial codes were created based on a previous coding scheme used for an interview study [14]. Initial coding consisted of placing all responses to the questions into the initial codes; however, responses that did not fit in the initial codes were placed in new codes generated inductively. Focused coding, following initial coding, consisted of recategorizing codes and creating additional codes as needed. For the final phase of thematic coding, common themes were identified among the codes and numbers of appearance were counted. The analysis of the open-ended questions was split among four coders: one coder analyzed the questions related to increased or decreased stress (BK); one coder analyzed the questions related to academic concerns (CS); one coder analyzed questions related to health and lifestyle (XW); and one coder analyzed questions related to coping mechanisms and barriers to treatment (AS). Between each phase of coding, the coders and other authors (SH and FS) met and discussed their process to ensure a uniform analysis method. The final coding structure and themes were decided upon in consensus meetings among all authors. Qualitative analyses were performed using Microsoft Excel (Microsoft Corp) and MAXQDA (VERBI Software) [15].

Results

Sample Demographics

A total of 2031 responses were collected, including 1252 (61.64%) from female respondents. Age of the participants ranged from 18 to 75 years (mean 22.88, SD 5.52). The sample included both undergraduate (n=1405, 69.18%) and graduate students (n=620, 30.53%), further classified as freshman (n=265, 13.05%), sophomore (n=274, 13.49%), junior (n=354, 17.43%), senior (n=512, 25.21%), master's (n=294, 14.48%), and doctorate (n=326, 16.05%). Study program was reported by 1900 participants representing all 15 colleges in the Texas A&M campus. The top represented colleges were Engineering (n=565, 29.74%), Liberal Arts (n=261, 13.74%), and Agriculture and Life Sciences (n=189, 9.95%). Table 1 shows the gender, classification, and program (college) proportions of the sample compared to the Texas A&M population.

Table 1. Demographics.

Characteristic	Sample, n (%)	Population[a], n (%)
Gender		
Female	1252 (61.64)	28,956 (46.60)
Male	757 (37.27)	33,197 (53.40)
Classification		
Undergraduate	1405 (69.18)	50,454 (81.18)
Master's	294 (14.48)	6259 (10.07)
Doctorate	326 (16.05)	4864 (7.83)
Age (years)		
<18	0 (0)	327 (0.53)
18-21	1065 (52.44)	31,839 (51.23)
22-25	532 (26.19)	22,946 (36.92)
26-30	207 (10.19)	3947 (6.35)
31-39	96 (4.73)	2176 (3.50)
>39	46 (2.26)	918 (1.48)
College		
College of Engineering	565 (29.74)	18,784 (30.22)
College of Liberal Arts	261 (13.74)	8526 (13.72)
College of Agriculture and Life Sciences	189 (9.95)	7473 (12.02)
College of Education & Human Development	176 (9.26)	6630 (10.67)
Mays Business School	154 (8.11)	6041 (9.72)
College of Science	134 (7.05)	3828 (6.16)
College of Veterinary Medicine & Biomedical Sciences	93 (4.89)	3575 (5.75)
College of Architecture	78 (4.11)	3142 (5.06)
School of Public Health	50 (2.63)	292 (0.47)
College of Geosciences	49 (2.58)	1321 (2.13)
Bush School of Government & Public Service	36 (1.89)	516 (0.83)

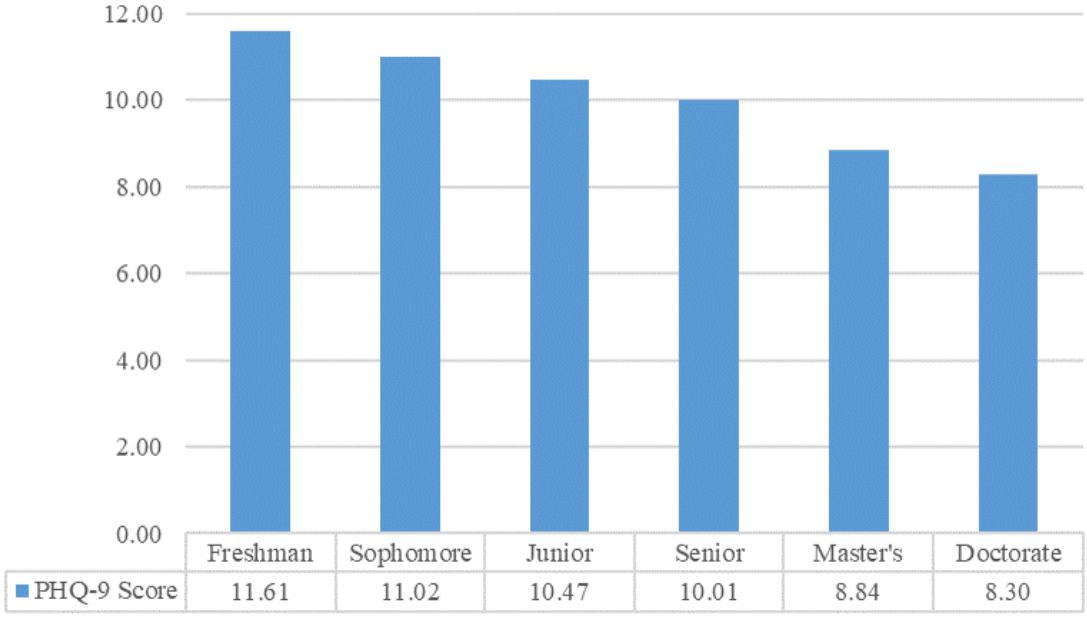
[a]Population based on Fall 2019 student demographics data at Texas A&M, College Station Campus.

Severity of Depression and Anxiety

Depression

A total of 37 responses were excluded from the PHQ-9 questionnaire analysis because of missing values. Among the 1994 complete responses, 1607 (80.57%) participants reported some (any) level of depression, as follows: mild (n=647, 32.45%), moderate (n=496, 24.87%), moderately severe (n=316, 15.85%), and severe (n=148, 7.42%). The two-way ANOVA showed that gender (P<.001, η^2=0.03) and classification (P<.001, η^2=0.03) significantly impacted PHQ-9 scores. Females had a mean score of 1.76 points higher than males (mean 10.61 and 8.84, respectively). Participants with a higher classification had lower PHQ-9 scores (Figure 1). Tukey's honest significant difference (HSD) showed a significant difference between doctoral and all undergraduate classifications (P<.001, P<.001, P<.001, and P=.001, respectively), between master's and freshman/sophomore/junior (P<.001, P<.001, and P=.04, respectively), and between senior and freshman (P=.004).

Figure 1. Mean Patient Health Questionnaire-9 (PHQ-9) score by classification.

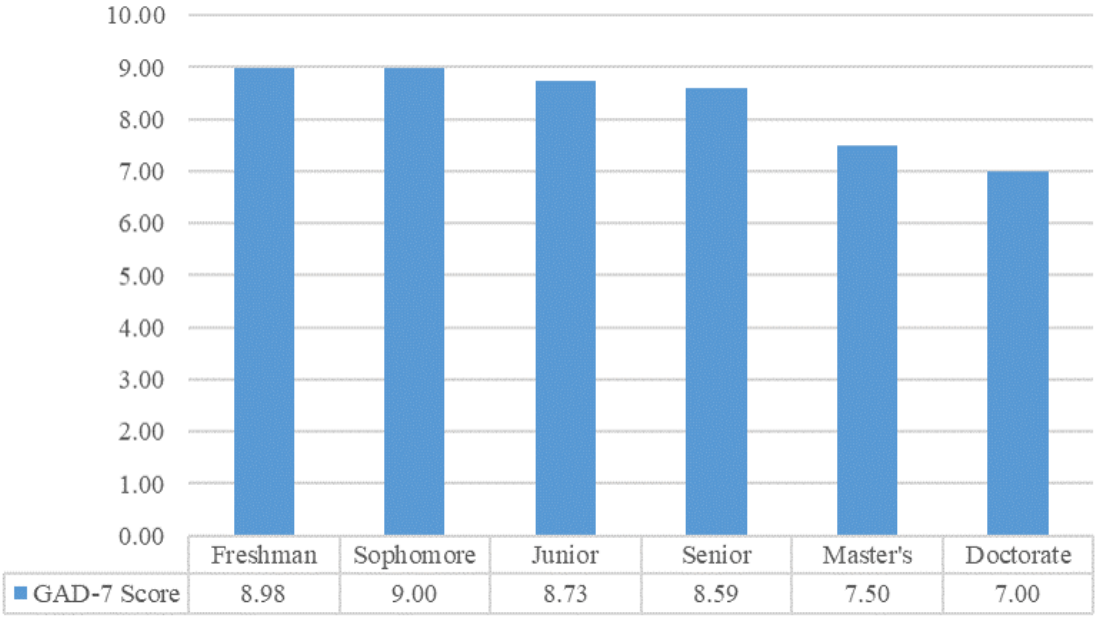


Responses to item 9 of the PHQ-9 ("over the last two weeks, how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?") showed that 366 (18.04%) participants had thoughts related to self-harm or suicide (250 responded "several days," 74 "more than half the days," and 42 "nearly every day").

Anxiety

In total, 17 responses were excluded from the GAD-7 questionnaire analysis because of missing values. Among the 2014 complete responses, 569 (28.25%) participants reported minimal anxiety, while 71.75% (n=1445) showed anxiety, with severity levels varying as mild (n=670, 33.27%), moderate (n=477, 23.68%), or severe (n=298, 14.80%). The two-way ANOVA showed a significant main effect of gender (P<.001, η^2=0.05) and classification (P<.001, η^2=0.02) on GAD-7 score. Females had a mean score of 2.22 points higher than males (means scores were 9.12 and 6.89, respectively). Participants with a higher classification had lower GAD-7 scores (Figure 2). Tukey's HSD test was conducted to test for differences between the classifications, and showed a significant difference between doctoral and all undergraduate classifications (P<.001, P<.001, P=.002, and P=.002, respectively), and between master's and sophomore (P=.03).

Figure 2. Mean Generalized Anxiety Disorder-7 (GAD-7) score by classification.



COVID-19–Related Stress

A majority of participants (n=1443, 71.26%) reported that their stress/anxiety levels had increased during the pandemic, while 111 (5.48%) indicated it had decreased and 471 (23.26%) indicated it remained the same as before. A vast majority (n=1982, 97.83%) of students thought other students were experiencing stress/anxiety because of the pandemic. Participants who indicated a change in their stress/anxiety levels were asked to specify reasons for such an increase or decrease in a follow-up question. Apart from the general comments on the stressors, participants also rated severity of the effect by specific academic-, health-, and lifestyle-related concerns.

Reasons for Increase in Stress

Among the participants who indicated increased stress/anxiety (n=1443) during the pandemic, 1360 participants elaborated on the reasons for such an increase. The biggest contributor was stress related to academics (532/1360, 39.12%), with the majority stemming from increased difficulty (n=278) due to the precipitous transition and maintenance of online classes, distantly followed by increased concerns over grades (n=58), and delayed graduation (n=53). The second most frequent contributor was general uncertainty regarding the pandemic (473/1360, 34.78%). This was followed by health concerns (472/1360, 34.71%) relating to personal mental health (n=205), health of friends and family (n=162), and fear of personally contracting COVID-19 (n=83). The fourth biggest concern related to finances (279/1360, 20.51%), primarily stemming from unemployment or uncertainty of future employment (n=183). Living/work environment (276/1360, 20.29%) was the next biggest contributor, consisting of concerns related to working from home (n=79), cabin fever (n=59), returning home (n=49), and confinement with others (n=47). Impacted social life (252/1360, 18.53%) was the last major category, primarily consisting of concerns related to isolation (n=186).

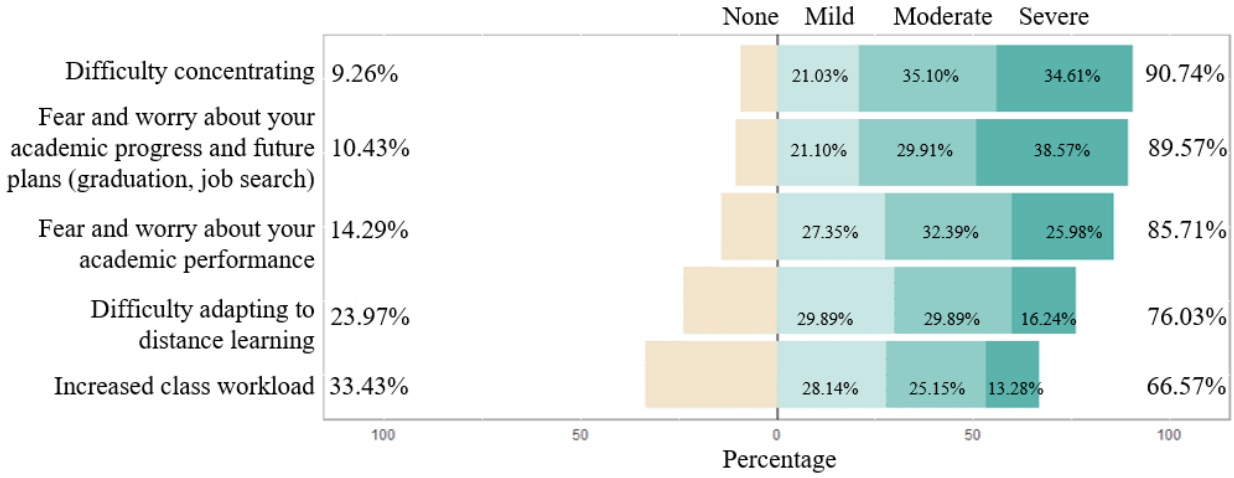
Reasons for Decrease in Stress

A few respondents (n=109) elaborated on the reasons that they were experiencing decreased stress and anxiety during the pandemic. The majority of these respondents mentioned that this was due to time saved (47/109, 43.1%) as a result of not having to commute to school, reduced schoolwork, and not having to engage in extracurricular and organizational activities, which are otherwise a part of campus life. A key benefit of the transition to distance learning was the schedule flexibility (n=20), particularly when lecture recordings could be viewed on students' own time. Several students (n=13) also mentioned using the additional time available to pursue hobbies and other interests, as well as to amplify proactive health behaviors such as meditation and exercise. Interestingly, some students (n=8) reported that they were experiencing reduced social anxiety from not having to interact with other students.

Academic-Related Concerns

In terms of academic-related concerns (Figure 3), 1851 (90.74%) participants had difficulty in concentrating, with 716 (35.10%) rating their difficulty as moderate and 706 (34.61%) as severe. Similarly, most participants had concerns regarding their academic progress and future plans (n=1830, 89.57%), as well as academic performance (n=1752, 85.71%). A majority of the participants also had difficulty adapting to distance learning (n=1554, 76.03%) or had an increased class workload (n=1358, 66.57%).

Figure 3. Academic-related concerns.

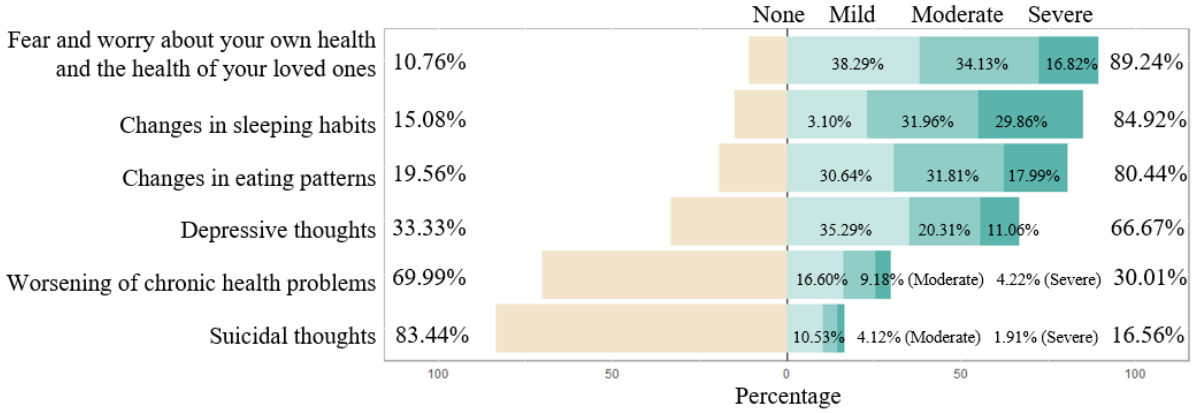


A small portion of participants (n=389) provided additional free-response reasons behind the increased level of stress related to academics. Nearly one fifth of those who indicated such reasons (73/389, 18.8%) reported financial concerns related to their academic situations such as reduced current and prospective job opportunities (n=42), increased burden to pay tuition and fees (n=16), and impacts on scholarship and funding (n=15). Some of the participants (27/389, 6.9%) presented their worry about future semesters, such as continuing online classes in following semesters (n=12), choosing a major in the middle of the pandemic situation (n=10), and resuming in-person classes with persistent risks of virus infection (n=5).

Health-Related Concerns

As Figure 4 shows, the top health-related concern was fear and worry about personal health and the health of loved ones (n=1825, 89.24%), followed by changes in sleeping habits (n=1735, 84.92%), eating patterns (n=1641, 80.44%), and depressive thoughts (n=1362, 66.67%).

Figure 4. Health-related concerns.

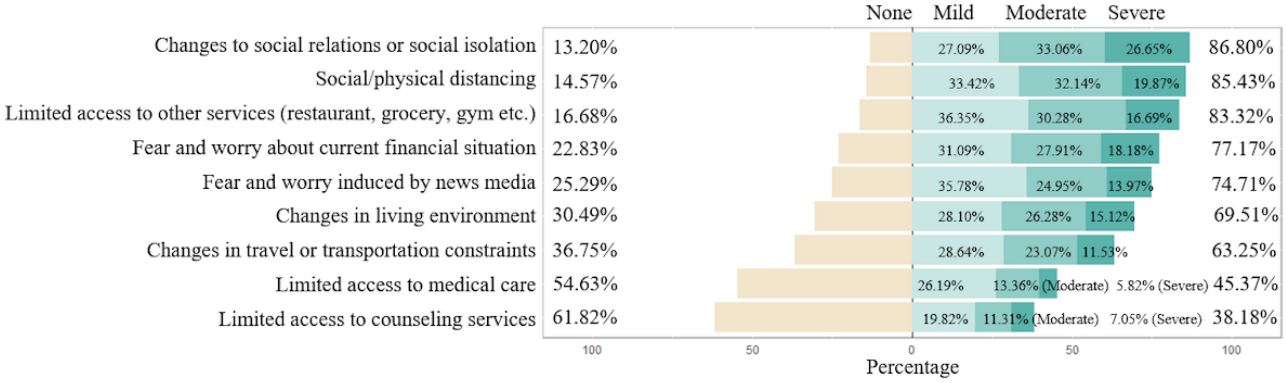


Some participants (n=180) added health-related concerns in the free responses. The biggest concern was physical illness (52/180, 29.9%), including having physical illness (n=31) that worsened during the pandemic (n=6). Some were at higher risk of getting infected with COVID-19 due to asthma (n=9), autoimmune disease (n=4), being immunocompromised (n=3) or other disease (n=3). Four participants reported that they have been infected with COVID-19. Fitness was another concern (49/180, 28.2%), including decreased exercise (n=45), weight gain (n=12), and muscle/back pain due to sedentary lifestyle (n=5). Some participants reported having been diagnosed with mental illness (n=31), which worsened during the pandemic (n=11). Barriers to health care were mentioned by 29 participants, including barriers or reluctance to visit doctors for nonpandemic issues (n=24), barriers to get medication (n=3), and barriers to get tested for COVID-19 (n=2).

Lifestyle-Related Concerns

More than half of the participants reported experiencing the top seven lifestyle concerns shown in Figure 5. It was not surprising that "changes to social relations or social isolation" (n=1775, 86.80%) and "social/physical distancing" (n=1741, 85.43%) were the top two lifestyle concerns.

Figure 5. Lifestyle-related concerns.



Some participants (n=174) mentioned additional concerns related to lifestyle changes in the free responses. The major concern was relationship (71/174, 40.8%), including social activities affected by isolation (n=45) and relationship issues with family members (n=18) and roommates (n=8). Some participants had difficulty working (n=17), due to unsatisfying work environments (n=9) such as internet issues (n=2), or had difficulty following a schedule (n=8). Other concerns included uncertainty of the future (n=12), worry because others do not perform social distancing (n=11), and worry of bringing the virus to their family as essential workers (n=2).

Coping Mechanisms and Barriers

Coping Mechanisms

Nearly half (n=882, 43.25%) of the participants indicated via multiple-choice responses that they were able to cope adequately with the stress related to the current situation, while 323 (15.84%) said they were not able to cope. The rest were unsure (n=834, 40.90%). When asked what coping mechanisms were used to mitigate stress/anxiety, more than half (n=1362, 67.06%) of the participants chose "support from community, family and friends," followed by "technologies (websites, mobile apps, sensors that help monitor health data)" (n=659, 32.45%). Few participants reported using university health services such as counseling service (n=210, 10.34%) or health services outside the university (n=89, 4.38%). Some of the participants (n=387, 19.05%) reported using no coping mechanism.

Some respondents (n=386) indicated that they used other coping mechanisms to reduce stress during the COVID-19 and provided elaboration in the free-response field. Many of these respondents (152/386, 39.0%) mentioned engaging in health lifestyle activities such as exercise (n=130), diet maintenance (n=11),

and self-care activities (n=11). A similar number of respondents (143/386, 37.1%) engaged in relaxing activities including meditation (n=48), reading (n=21), playing with pets (n=13), listening to music (n=12), breathing exercises (n=4), sleeping (n=3), shooting for sport (n=2), gardening (n=1), and other hobbies in general (n=29) alongside general relaxing activities (n=10). Beyond relaxing activities, some respondents engaged in creative activities (31/386, 8.0%), which included creating pieces of art (n=10), writing (n=18), and playing musical instruments (n=3). Additionally, respondents mentioned engaging in spiritual and religious activities (n=69) such as reading sacred texts and praying. A small percentage of the respondents also engaged in negative coping methods (41/386, 10.6%), including distracting themselves (n=20), excessive intake of alcohol (n=5), isolation (n=2), auto-manipulation (n=2), and crying (n=1). Lastly, some respondents maintained focus on their work and professional activities (16/386, 4.2%) by concentrating on school (n=4), continuing work (n=3), or managing their time for productive use.

Smartphone Apps for Coping With COVID-19

A small number of participants (n=290, 14.28%) indicated that they have been using mobile apps for managing added stress related to the pandemic, 278 of whom provided specific names of the apps or features. Most of these respondents (201/278, 72.3%) used an app focused on mindfulness. Most of these mindfulness apps focused on meditation (144), such as Headspace; relaxation apps (n=42) and apps centered on focused breathing (n=15) were also used. Social media (46/278, 16.5%) was used by some of the respondents with most using the typical social media sites (eg, Twitter, Facebook, Instagram, and TikTok) and apps (n=32) for entertainment purposes while others used YouTube (n=14). Some respondents used lifestyle apps (37/278, 13.3%), including exercise apps (n=23), time management (n=9), sleep tracking (n=5), and food tracking (n=2). A small number of respondents (6/278, 2.2%) played video game apps to cope with stress during COVID-19.

Sources of Information for COVID-19

The main sources of information related to the pandemic were university emails (n=1257, 61.89%), newspapers, paper or online periodicals (n=1221, 60.12%), and posts on social media (n=1026, 50.52%). Some participants also received information from medical websites (n=591, 29.10%) or their medical providers (n=315, 15.51%).

A small number of respondents (n=235) elaborated on the other information sources in open-ended responses. About a quarter (60/235, 25.5%) of these respondents received their information from more traditional media including television (n=34), online articles (n=19), and the radio (n=7). A similar number of participants (55/235, 23.4%) received their information from family and friends. Other respondents listened to figures of authority (50/235, 21.3%) for their information regarding COVID-19, split between government officials and organizations, including presidential briefings and local leaders (n=32) on one hand, and prominent scientists and researchers (n=28) on the other. Finally, some respondents preferred to rely on themselves to stay informed (21/235, 8.9%) by conducting searches and reading relevant sources (n=4), or by ignoring the news due to disbelief in or indifference toward the current situation (n=17).

Barriers for Mental Health Care

The major barriers to mental health care were “financial concerns (fees and insurance)” (n=1434, 70.61%), “social stigma” (n=1194, 58.79%), “lack of information about resources available” (n=1099, 54.11%), and “limited access to the services (eg, could not get scheduled)” (n=978, 48.15%).

A few respondents (n=167) indicated that they experienced other barriers and elaborated through the free-response section. Some of the respondents perceived that they themselves (69/167, 41.3%) can be the biggest barrier to reaching out for help: some doubted the efficacy of care (n=8); some mentioned that they or others may not see a problem even if it exists (n=22); others mentioned not wanting help (n=7). Respondents cited an overall feeling of discomfort (25/167, 15.0%) with the topic of mental health and difficulty bringing up the topic with others. Some of the respondents mentioned poor quality (22/167, 13.2%) in their treatment as a barrier to seeking treatment again.

Discussion

Principal Findings

Among the 2031 participants, 48.14% showed a moderate-to-severe level of depression, 38.48% showed a mild-to-severe level of anxiety, and 18.04% had suicidal thoughts in the 2 weeks preceding the survey. Gender and classification had significant effects on depression and anxiety severity (P<.001). Female respondents reported higher scores, while respondents in a higher classification reported lower scores on PHQ-9 and GAD-7. A majority of participants (71.26%) indicated that their stress/anxiety levels had increased during the pandemic. Less than half (43.25%) indicated that they were able to cope adequately with the stress related to the current situation.

The survey had a healthy representation across genders and classifications of undergraduate and graduate students. A vast majority (80.57%) of respondents had scores on the PHQ-9 that indicated some level of depression (defined as a total PHQ-9 score of ≥5), with about 48% in the moderate-to-severe range. This proportion of respondents showing depression is much larger than those found in recent assessments in China. For instance, in their survey of 509 college students, Liu et al [16] found that about 19% of their respondents showed some level of depression. Our findings also show a higher proportion of respondents with depressive symptoms among students than findings in several recent studies in nonpandemic situations [17,18]. Furthermore, nearly 1 in 5 respondents reported having suicidal thoughts. This finding is in line with the increased suicide rates observed during previous pandemics [19]. In comparison, previous research has reported about 3% to 7% of the college student population had suicidal thoughts outside of a pandemic situation [20]. This is an alarming finding warranting immediate attention. Additionally, a majority of our respondents (71.75%) showed some level of anxiety (defined as a total GAD-7 score of ≥5), with over 38% in the moderate-to-severe range. Again, this is a much higher proportion compared to

similar survey-based assessments by Liu et al [16] and Cao et al [9], who found some level of anxiety in 8.8% (out of 509) and 24.9% (out of 7143) of respondents, respectively. Clearly, there is a pressing need to actively provide support to vulnerable students in managing their mental health.

Not surprisingly, given the above findings, a majority of respondents reported that their stress and anxiety had increased during the pandemic. In general, this is consistent with the heightened levels of psychological distress reported among various populations during the current pandemic and previous epidemics such as severe acute respiratory syndrome (SARS) [6,7,9,21]. This finding could be underpinned by the high levels of academic, health, and lifestyle concerns and changes. A vast majority indicated difficulty concentrating, fear, and worry about academic progress and performance, and adjustment to distance learning as dominant academic concerns. To our knowledge, this is the first study that reports these specific effects of the COVID-19 pandemic as related to academic concerns. Given that several universities, including Texas A&M, are continuing partially with distance learning for the remainder of the year, these concerns need to be probed further in order to be adequately addressed.

Among health-related concerns, a majority of students expressed concerns about their own health or the health of loved ones, echoing recent findings [6]. A large proportion (over 80%) of respondents reported changes in eating and sleeping habits. Again, this is not surprising, but certainly concerning, given previous research which has shown that such changes are correlated with depression among college students [22]. Among lifestyle-related concerns, physical distancing and changes in social relations were widely reported, similar to those found earlier among students as well as the general population [6,23]. Additionally, three quarters of respondents indicated fear and worry induced by news outlets. This type of distress may be exacerbated by the large amount of misinformation, including false and fabricated information, distributed through news and social media platforms [24].

More than half the respondents who described coping mechanisms mentioned support from family and friends as a key factor, similar to previous findings [6,9]. Several respondents also mentioned the use of technology, such as mobile apps and other digital platforms, as a means of positive coping practices, such as meditation, echoing recent findings on the positive effect of a mindfulness app on college students' mental health [25]. This indicates some potential for mobile-based technologies to support mental health. Such platforms may have the added benefit of helping overcome the barrier of social stigma related to seeking help from counseling services. Identifying such positive coping behaviors is important in order to enable those behaviors through symptoms-level support.

Limitations and Future Work

Several limitations may impact the generalizability of the findings reported in this paper. Most importantly, some of our findings may be biased due to self-selection by respondents. The higher percentage of respondents with depression/anxiety may be related to this bias. Particularly, the slightly higher level of depression/anxiety among females may be attributed to the slightly higher percentage of female respondents. Additionally, we did not ask if respondents had any existing mental health issues or were receiving treatment before the pandemic. In fact, 31 respondents mentioned on their own that they had a prior diagnosis of mental illness. Therefore, we are unable to clarify whether our findings have been biased by a population of respondents with pre-existing or heightened levels of distress, who may have been more inclined to participate in the survey than those who were less distressed.

The survey is cross-sectional and lacks comparison to a typical semester unaffected by the pandemic, or a different time point of the year. Huckins et al [26] tracked depression and anxiety severity of 217 undergraduate students using PHQ-4 and GAD-2 during Spring 2020. It was found that depression and anxiety levels spiked when the campus switched to remote learning but decreased in the following 2 weeks. It is valuable to keep monitoring the change to understand the long-term effect of the pandemic.

An interesting finding is the differences in depression/anxiety levels among different classifications of students. Undergraduate students may have been more heavily impacted during the pandemic compared to graduate students, probably from adapting to distance learning. Yet the precise factors need further investigation in a future study.

The proportion of respondents showing depression, anxiety, and/or suicidal thoughts is alarming. Respondents reported academic, health, and lifestyle-related concerns caused by the pandemic. Given the unexpected length and severity of the pandemic, these concerns need to be further understood and addressed. Further study on the most at-risk populations and evidence-based interventions should proceed as soon as possible to prevent a secondary epidemic, embedded within the COVID-19 pandemic, of a serious, nationwide mental affliction and potential physical self-harm among vulnerable college students.

At an institutional level, online remote activities and services can be implemented to provide support to students that help address concerns related to the pandemic. For example, Schlesselman et al [27] provided a list of activities that can potentially support students in fitness, socialization, and academic success (eg, virtual group exercise, virtual movie night, and virtual office hours). However, there is no one-size-fits-all solution. More work is needed to identify appropriate ways to implement such support and assess the long-term effects of such interventions.

Conflicts of Interest

None declared.

References

1. Coronavirus Disease 2019 (COVID-19) in the U.S. CDC. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed 2020-08-13]
2. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
3. Zandifar A, Badrfam R. Iranian mental health during the COVID-19 epidemic. Asian J Psychiatr 2020 Jun 04;51:101990 [FREE Full text] [doi: 10.1016/j.ajp.2020.101990] [Medline: 32163908]
4. Center for Collegiate Mental Health 2019 Annual Report. Global Health Education and Learning Incubator at Harvard University. 2020. URL: https://repository.gheli.harvard.edu/repository/collection/resource-pack-college-students-and-mental-health/resource/12372 [accessed 2020-07-23]
5. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun 15;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
6. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e17-e18. [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
7. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):e [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
8. Zhai Y, Du X. Addressing collegiate mental health amid COVID-19 pandemic. Psychiatry Res 2020 Jun;288:113003 [FREE Full text] [doi: 10.1016/j.psychres.2020.113003] [Medline: 32315885]
9. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
10. Kelley K, Clark B, Brown V, Sitzia J. Good practice in the conduct and reporting of survey research. Int J Qual Health Care 2003 Jun;15(3):261-266. [doi: 10.1093/intqhc/mzg031] [Medline: 12803354]
11. Kroenke K, Spitzer RL. The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
12. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
13. Guest G, MacQueen K, Namey E. Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications, Inc; 2012:978-971.
14. Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study. J Med Internet Res 2020 Sep 03;22(9):e21279 [FREE Full text] [doi: 10.2196/21279] [Medline: 32805704]
15. MAXQDA. Berlin, Germany: VERBI Software; 2020.
16. Liu X, Liu J, Zhong X. Psychological State of College Students During COVID-19 Epidemic (Preprint). SSRN Journal 2020:e. [doi: 10.2139/ssrn.3552814]
17. Zeng Y, Wang G, Xie C, Hu X, Reinhardt JD. Prevalence and correlates of depression, anxiety and symptoms of stress in vocational college nursing students from Sichuan, China: a cross-sectional study. Psychol Health Med 2019 Aug;24(7):798-811. [doi: 10.1080/13548506.2019.1574358] [Medline: 30714819]
18. Fata Nahas ARM, Elkalmi RM, Al-Shami AM, Elsayed TM. Prevalence of Depression Among Health Sciences Students: Findings From a Public University in Malaysia. J Pharm Bioallied Sci 2019;11(2):170-175 [FREE Full text] [doi: 10.4103/jpbs.JPBS_263_18] [Medline: 31148894]
19. Gunnell D, Appleby L, Arensman E, Hawton K, John A, Kapur N, et al. Suicide risk and prevention during the COVID-19 pandemic. The Lancet Psychiatry 2020 Jun;7(6):468-471 [FREE Full text] [doi: 10.1016/S2215-0366(20)30171-1] [Medline: 32330430]
20. Zivin K, Eisenberg D, Gollust SE, Golberstein E. Persistence of mental health problems and needs in a college student population. J Affect Disord 2009 Oct;117(3):180-185. [doi: 10.1016/j.jad.2009.01.001] [Medline: 19178949]
21. Moccia L, Janiri D, Pepe M, Dattoli L, Molinaro M, De Martin V, et al. Affective temperament, attachment style, and the psychological impact of the COVID-19 outbreak: an early report on the Italian general population. Brain Behav Immun 2020 Jul;87:75-79 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.048] [Medline: 32325098]
22. Acharya L, Jin L, Collins W. College life is stressful today - Emerging stressors and depressive symptoms in college students. J Am Coll Health 2018 Oct;66(7):655-664. [doi: 10.1080/07448481.2018.1451869] [Medline: 29565759]
23. Nelson LM, Simard JF, Oluyomi A, Nava V, Rosas LG, Bondy M, et al. US Public Concerns About the COVID-19 Pandemic From Results of a Survey Given via Social Media. JAMA Intern Med 2020 Jul 01;180(7):1020-1022 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1369] [Medline: 32259192]

24. Brennen S, Simon F, Howard P, Nielsen R. Types, sources, and claims of COVID-19 misinformation. Reuters Institute / University of Oxford. 2020 Apr 7. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2020-07-23]
25. Lee RA, Jung ME. Evaluation of an mHealth App (DeStressify) on University Students' Mental Health: Pilot Trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
26. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
27. Schlesselman LS, Cain J, DiVall M. Improving and Restoring the Well-being and Resilience of Pharmacy Students during a Pandemic. Am J Pharm Educ 2020 Jun;84(6):ajpe8144 [FREE Full text] [doi: 10.5688/ajpe8144] [Medline: 32665720]

Abbreviations

ANOVA: analysis of variance
GAD-7: General Anxiety Disorder 7-item
PHQ-9: Patient Health Questionnaire 9-item
SARS: severe acute respiratory syndrome
Tukey's HSD: Tukey's honest significant difference

Edited by G Eysenbach; submitted 23.07.20; peer-reviewed by M Elhadi, H Jiang; comments to author 03.08.20; revised version received 02.09.20; accepted 07.09.20; published 17.09.20.

Please cite as:
Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F
Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study
J Med Internet Res 2020;22(9):e22817
URL: http://www.jmir.org/2020/9/e22817/
doi:10.2196/22817
PMID:32897868

Original Paper

New York Inner City Hospital COVID-19 Experience and Current Data: Retrospective Analysis at the Epicenter of the American Coronavirus Outbreak

Vishnu R Mani[1*], MD; Aleksandr Kalabin[1*], MD; Sebastian C Valdivieso[1], MD; Max Murray-Ramcharan[1], MD; Brian Donaldson[1], MD, FACS

Columbia University College of Physicians and Surgeons at Harlem Hospital, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Vishnu R Mani, MD
Columbia University College of Physicians and Surgeons at Harlem Hospital
506 Lenox Avenue
New York, NY,
United States
Phone: 1 2129391000 ext 1641
Email: vishnu.mani@duke.edu

Abstract

Background: In the midst of the coronavirus disease pandemic, emerging clinical data across the world has equipped frontline health care workers, policy makers, and researchers to better understand and combat the illness.

Objective: The aim of this study is to report the correlation of clinical and laboratory parameters with patients requiring mechanical ventilation and the mortality in patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Methods: We did a review of patients with SARS-CoV-2 confirmed infection admitted and managed by our institution during the last month. Patients were grouped into intubated and nonintubated, and subgrouped to alive and deceased. A comprehensive analysis using the following parameters were performed: age, sex, ethnicity, BMI, comorbidities, inflammatory markers, laboratory values, cardiac and renal function, electrocardiogram (EKG), chest x-ray findings, temperature, treatment groups, and hospital-acquired patients with SARS-CoV-2.

Results: A total of 184 patients were included in our study with ages ranging from 28-97 years (mean 64.72 years) and including 73 females (39.67%) and 111 males (60.33%) with a mean BMI of 29.10. We had 114 African Americans (61.96%), 58 Hispanics (31.52%), 11 Asians (5.98%), and 1 Caucasian (0.54%), with a mean of 1.70 comorbidities. Overall, the mortality rate was 17.39% (n=32), 16.30% (n=30) of our patients required mechanical ventilation, and 11.41% (n=21) had hospital-acquired SARS-CoV-2 infection. Pertinent and statistically significant results were found in the intubated versus nonintubated patients with confirmed SARS-CoV-2 for the following parameters: age (P=.01), BMI (P=.07), African American ethnicity (P<.001), Hispanic ethnicity (P=.02), diabetes mellitus (P=.001), creatinine (P=.29), blood urea nitrogen (BUN; P=.001), procalcitonin (P=.03), C-reactive protein (CRP; P=.007), lactate dehydrogenase (LDH; P=.001), glucose (P=.01), temperature (P=.004), bilateral pulmonary infiltrates in chest x-rays (P<.001), and bilateral patchy opacity (P=.02). The results between the living and deceased subgroups of patients with confirmed SARS-CoV-2 (linking to or against mortality) were BMI (P=.04), length of stay (P<.001), hypertension (P=.02), multiple comorbidity (P=.045), BUN (P=.04), and EKG findings with arrhythmias or blocks (P=.02).

Conclusions: We arrived at the following conclusions based on a comprehensive review of our study group, data collection, and statistical analysis. Parameters that were strongly correlated with the need for mechanical ventilation were younger age group, overweight, Hispanic ethnicity, higher core body temperature, EKG findings with sinus tachycardia, and bilateral diffuse pulmonary infiltrates on the chest x-rays. Those intubated exhibited increased disease severity with significantly elevated levels of serum procalcitonin, CRP, LDH, mean glucose, creatinine, and BUN. Mortality was strongly correlated with BMI, African American ethnicity, hypertension, presence of multiple comorbidities (with a mean of 2.32), worsening renal function with acute kidney injury or acute chronic kidney injury, and EKG findings of arrhythmias and heart blocks.

(J Med Internet Res 2020;22(9):e20548) doi:10.2196/20548

KEYWORDS

SARS-CoV-2; COVID-19; pandemic; New York City; coronavirus outbreak; American minority; outbreak; minority; mortality; patient; characteristic; mechanical ventilation

Introduction

There is an emergent need for more research and data sharing to better delineate and understand the disease process of the coronavirus across the world. The ongoing pandemic has struck monumental changes in every aspect of human life [1]. A large amount of coronavirus cases resulted in a massive outbreak in New York State and New York City, making it the epicenter of the world [2]. Since the outbreak, rapid publications from Europe [3] and China [4] helped the rest of the world understand the disease course and severity at varying degrees. New York is one of the most culturally diverse places in the world with varying demographics across the city. Our institution in Manhattan has been at the forefront in this hotspot combating the coronavirus disease (COVID-19) pandemic at its peak. As there is a severe underrepresentation of American minority and underprivileged communities, we felt the urgent need to research and analyze various parameters associated with the disease among our patients that might help a deeper understanding and lead to further research from our peers across the world in mitigating the COVID-19 pandemic.

Methods

Study Design

Ethical approval from the hospital Institutional Review Board for Health Sciences Research was expeditiously sought and approved. This study entails data collection of all patients that were hospitalized and had a confirmed case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection.

Data Collection

Further review using the electronic medical record was undertaken to obtain the parameters included in this study, namely, diagnosis; age; sex; ethnicity; BMI; in-hospital SARS-CoV-2 conversion status (patients without COVID-19–related symptoms upon admission and subsequently tested positive during hospital stay; hospital acquired); need for mechanical ventilation; duration of mechanical ventilation; fever (core body temperature); length of stay; death; comorbidities: hypertension (HTN), chronic kidney disease (CKD), coronary artery disease (CAD), diabetes mellitus (DM), bronchial asthma (BA), and chronic obstructive pulmonary disease (COPD); inflammatory markers: procalcitonin, C-reactive protein (CRP), lactate dehydrogenase (LDH), and ferritin; and laboratory values: creatinine (Cr), blood urea nitrogen (BUN), D-dimers, prothrombin time (PT), partial thromboplastin time, troponin, blood natriuretic peptide (BNP), glucose, electrocardiogram (EKG), chest x-ray (CXR) findings, and treatment received (azithromycin, hydroxychloroquine).

Inclusion and Exclusion Criteria

Out of 204 total patients, 184 were admitted with high suspicion of coronavirus infection or other diagnoses (who later had a conversion) between March and April 2020 and were included in our study. The included patients were confirmed positive for SARS-CoV-2. Of the 204 patients, 20 were excluded, as they tested negative.

Statistical Analysis

Baseline characteristics of all patients with laboratory confirmed SARS-CoV-2 were included in our study and described in detail. Gross numbers with percentages were calculated and used. We summarized the continuous variables using mean and standard deviation, whereas categorical variables were summarized using counts and percentages. Continuous variables were primarily analyzed using the student *t* test with Welch correction, while categorical variable analysis were accomplished using the Fisher exact test; subgroup analyses were performed using a two-tailed chi-square test. $P<.05$ was considered statistically significant. A Microsoft Excel (Microsoft Corporation) spreadsheet was used to display extrapolated data. Graph Pad Prism 8 (Graph Pad Software) and R software version 4.0.0 (R Project for Statistical Computing) were used for statistical analysis.

Results

Definitions

Patients included in our study for the purposes of understanding the disease severity and statistical analysis were further classified into the following groups: the entire group (termed total), nonintubated total (NI-T), nonintubated living (NI-L), nonintubated deceased (NI-D), intubated total (I-T), intubated living (I-L), and intubated deceased (I-D). We compared parameters across all patients and groups to delineate deeper clinical understanding with associations as detailed further in the results and Multimedia Appendix 1.

Patient Characteristics

The total number of patients included in our study was 184 of the total 204 as detailed in the inclusion and exclusion criteria in the Methods section. All of our patients were SARS-CoV-2 confirmed by polymerase chain reaction of the collected nasopharyngeal sample.

The overall mortality in our study was 17.39% (n=32). In our study, 16.30% (n=30) required mechanical ventilation, and the mortality rate was 43.33% (n=13/30) among those intubated. Of the 184 patients, 21 (11.41%) had contracted SARS-CoV-2 in the hospital who were otherwise admitted for other reasons, mostly surgical. The average age was 64.72 (range 28-97, SD 14.87) years. Over one-third of the patients included were female and less than two-thirds were male. Mean age with SD for all groups is reported in Table 1. We found that younger patients were more likely to be intubated than their older counterparts. The comparison between NI-T and I-T yielded a *P* value of .01, and the comparison between NI-L and NI-D yielded a *P* value of .51, and between I-L and I-D yielded a *P* value of .60.

Table 1. Baseline patient characteristics and descriptive statistics.

Patient characteristics	Total (N=184)	NI-T[a] (n=154)	I-T[b] (n=30)	*P* value	NI-L[c] (n=135)	NI-D[d] (n=19)	*P* value	I-L[e] (n=17)	I-D[f] (n=13)	*P* value
Age (years), mean (SD)	64.72 (14.87)	65.75 (15.25)	59.47 (11.56)	.01	64.58 (15.04)	74.05 (14.52)	.51	58.47 (12.19)	60.77 (11.03)	.59
Sex, n (%)				.04			.45			.10
Females	73 (39.67)	61 (39.61)	12 (40)		55 (40.74)	6 (31.58)		9 (52.94)	3 (23.08)	
Males	111 (60.33)	93 (60.39)	18 (60)		80 (59.26)	13 (68.42)		8 (47.06)	10 (76.92)	
BMI (kg/m^2), mean (SD)	29.10 (7.386)	28.45 (7.017)	31.74 (8.645)	.07	29.00 (7.066)	25.71 (6.116)	.04	33.98 (10.07)	28.57 (4.913)	.07
<24.9, n (%)	50 (29.24)	45 (31.69)	5 (17.24)		35 (28.46)	10 (52.63)		3 (17.65)	2 (16.67)	
25.0-29.9 (overweight), n (%)	55 (32.16)	46 (32.39)	9 (31.03)		41 (33.33)	5 (26.32)		2 (1.76)	7 (58.33)	
30.0-34.9 (class 1 obesity), n (%)	39 (22.80)	31 (21.83)	8 (27.59)		29 (23.58)	2 (10.53)		7 (41.18)	1 (8.33)	
35.0-39.9 (class 2 obesity), n (%)	16 (9.36)	11 (7.75)	5 (17.24)		9 (7.32)	2 (10.53)		3 (17.65)	2 (16.67)	
>40.0 (class 3 obesity), n (%)	11 (6.43)	9 (6.34)	2 (6.90)		9 (7.32)	0 (0)		2 (11.76)	0 (0)	
Ethnicity, n (%)										
African American	114 (61.96)	101 (65.58)	13 (43.33)	<.001	86 (63.70)	15 (78.95)	.19	10 (58.82)	3 (23.07)	.05
Hispanic	58 (31.52)	43 (27.92)	15 (50)	.02	39 (28.89)	4 (21.05)	.48	6 (35.30)	9 (69.23)	.07
Asian	11 (5.98)	9 (5.84)	2 (6.67)	N/A[g]	9 (6.67)	0 (0.00)	N/A	1 (5.88)	1 (7.7)	N/A
Caucasians	1 (0.54)	1 (0.65)	0 (0.00)	N/A	1 (0.74)	0 (0.00)	N/A	0 (0.00)	0 (0.00)	N/A
Length of stay (days), mean (SD)	15.15 (13.45)	14.85 (14.42)	16.67 (6.434)	.28	15.76 (15.12)	8.37 (3.833)	<.001	20.88 (4.833)	11.15 (3.288)	<.001
Temperature (°F), mean (SD)	100.9 (1.680)	100.7 (1.568)	101.8 (1.928)	.004	100.7 (1.551)	100.4 (1.694)	.37	102.1 (1.703)	101.5 (2.225)	.49

[a]NI-T: nonintubated total.

[b]I-T: intubated total.

[c]NI-L: nonintubated living.

[d]NI-D: nonintubated deceased.

[e]I-L: intubated living.

[f]I-D: intubated deceased.

[g]Not applicable.

BMI

Of the 184 patients, the mean for BMI was 29.10 (SD 7.386, range 13.38-61.68) kg/m^2, of which more than one-fourth had a BMI<24.0 kg/m^2, less than one-tenth had a BMI of 25-29.9 kg/m^2, less than one-fourth were in class 1, and class 2 and class 3 each had less than one-tenth. We then analyzed to see if the BMI was associated with either mortality or the increased need for mechanical ventilation. We found that BMI had correlations for both NI-T and I-T, with means of 28.56 and 31.74, respectively (*P*=.07); NI-L and NI-D had means of 29 and 25.71, respectively (*P*=.04), and I-L and I-D had means of 33.98 and 28.57, respectively (*P*=.07).

Ethnicity

Of the 184 patients, less than one-third were Hispanic, more than three-fifths were African Americans, less than one-tenth were Asians, and there was 1 Caucasian. For African Americans, the comparison of NI-T and I-T had a *P* value <.001; I-L and I-D yielded a *P* value of .05. For Hispanics, a comparison of NI-T and I-T yielded a *P* value of .02; I-L and I-D had a *P* value of .07. All subgroup analysis findings are listed in Table 1.

Comorbidities

The average comorbidity was 1.70 for all patients admitted with SARS-CoV-2 infection. The average comorbidity among patients in each group were NI-T: 1.73; I-T: 1.57; NI-L: 1.64; NI-D: 2.32; I-L: 1.41; and I-D: 1.77. There was statistical significance with a *P* value of .045 on comparing NI-L and NI-D, clearly indicating that increased comorbidity among patients not requiring mechanical ventilation positively correlated with mortality. We then calculated percent associations for each group based on the presence of existing comorbidities. Overall, almost two-thirds of the overall group had HTN. There was statistical significance when comparing NI-L and NI-D (*P*=.02), indicative of a positive correlation of HTN with mortality; there was no difference between the I-T and NI-T (*P*=0.91). Overall, less than one-fifth of the patients had CKD; there was no statistical significance when comparing all 3 groups for CKD as depicted in Table 2. Overall, about one-fifth had CAD. There were no statistical differences noted among the groups for CAD; however, a comparison of NI-L and NI-D yielded a *P*=.07. Overall, over two-fifths of our patient population had DM. We had statistical significance between the NI-T and I-T for DM (*P*<.001). Both BA and COPD made up less than one-tenth of our patients. There was no significance among any groups for BA and COPD. Table 2 is representative of the percent associations of various comorbidities with *P* values for the groups compared.

Table 2. Association of comorbidities in patients with severe acute respiratory syndrome coronavirus 2.

Comorbidity	Total (N=184), n (%)	NI-T[a] (n=154), n (%)	I-T[b] (n=30), n (%)	*P* value	NI-L[c] (n=135), n (%)	NI-D[d] (n=19), n (%)	*P* value	I-L[e] (n=17), n (%)	I-D[f] (n=13), n (%)	*P* value
Hypertension	12 (65.76)	101 (65.68)	20 (66.67)	.91	84 (62.22)	17 (89.47)	.02	10 (58.82)	10 (76.92)	.30
Chronic kidney disease	32 (17.39)	28 (18.18)	4 (13.33)	.52	22 (16.30)	6 (31.58)	.11	1 (5.89)	3 (23.08)	.17
Coronary artery disease	37 (20.11)	32 (20.78)	5 (16.67)	.61	25 (18.52)	7 (36.84)	.07	2 (11.76)	3 (23.08)	.41
Diabetes mellitus	80 (43.48)	75 (48.70)	5 (16.67)	.001	66 (48.89)	9 (47.37)	.90	2 (11.76)	3 (23.08)	.41
Bronchial asthma	18 (9.78)	17 (11.04)	1 (3.33)	.19	13 (9.63)	4 (21.05)	.14	0 (0.00)	1 (7.69)	.24
Chronic obstructive pulmonary disease	17 (9.24)	16 (10.39)	1 (3.33)	.22	13 (9.63)	3 (15.79)	.41	0 (0.00)	1 (7.69)	.24

[a]NI-T: nonintubated total.

[b]I-T: intubated total.

[c]NI-L: nonintubated living.

[d]NI-D: nonintubated deceased.

[e]I-L: intubated living.

[f]I-D: intubated deceased.

Inflammatory Markers

We then assessed the distribution of inflammatory markers across various subgroups to gain an understanding of them and to see if this would serve as an indirect surrogate marker of the disease severity. Details of inflammatory markers along with other laboratory values with range, mean, SD, and the standard error of the mean (SEM) for each subgroup are depicted in Table 3 along with the results of the statistical analysis.

Table 3. Inflammatory markers and laboratory values in patients with confirmed severe acute respiratory syndrome coronavirus 2.

Parameters	Total	I-T[a]	NI-T[b]	*P* value	NI-L[c]	NI-D[d]	*P* value	I-L[e]	I-D[f]	*P* value
Creatinine (mg/dL)				.03			.10			.77
Mean (SD)	2.738 (3.177)	4.030 (3.481)	2.486 (3.063)		2.286 (2.879)	3.908 (3.953)		4.194 (3.811)	3.815 (3.134)	
Range	0.40-15.80	0.60-12.30	0.40-15.80		0.40-15.60	0.70-15.80		0.60-12.30	0.70-10.90	
SEM[g]	0.2342	0.6355	0.24688		0.2477	0.9068		0.9244	0.8692	
95% CI	2.276-3.200	2.730-5.330	1.999-2.974		1.796-2.776	2.003-5.814		2.235-6.154	1.922-5.71	
Blood urea nitrogen (mg/dL)				.001			.04			.56
Mean (SD)	45.30 (38.07)	68.47 (40.69)	40.79 (35.97)		37.76 (33.21)	62.26 (47.27)		64.53 (38.14)	73.62 (44.84)	
Range	6-234	10-169	6-234		6-234	19-188		22-134	10-169	
SEM	2.807	7.429	2.899		2.858	10.84		9.251	12.44	
95% CI	39.76-50.84	53.27-83.66	35.06-46.51		32.11-43.42	39.48-85.05		44.92-84.14	46.52-100.7	
Procalcitonin (ng/dL)				.03			.20			.23
Mean (SD)	2.463 (7.091)	7.809 (13.62)	1.435 (4.348)		1.237 (4.209)	3.380 (5.363)		4.581 (7.071)	12.65 (19.33)	
Range	0.03-56.42	0.13-56.42	0.03-43.95		0.03-43.95	0.14-16.24		0.13-24.44	0.15-56.42	
SEM	0.5696	2.724	0.3814		0.3875	1.548		1.826	6.112	
95% CI	1.338-3.588	2.186-13.43	0.681-2.190		0.47-2.005	–0.028 to 6.79		0.665-8.497	–1.18 to 26.48	
C-reactive protein (mg/L)				.01			.24			.76
Mean (SD)	17.99 (14.28)	24.06 (11.60)	15.39 (14.61)		13.47	34.97		24.70	23.16	
Range	0.1-91.80	3.3-44.82	0.1-91.80		0.1-34.46	3.64-91.80		3.3-44.82	3.34-35.57	
SEM	1.596	2.368	1.953		1.385	15.41		3.172	3.726	
95% CI	14.81-21.17	19.16-28.96	11.48-19.30		10.69-16.25	–7.81 to 77.76		17.85-31.56	14.73-31.58	
D-dimer (ng/ml)				.45			.47			.17
Mean (SD)	5200 (11,363)	6636 (10,959)	4512 (11,598)		3733 (10,070)	13463 (23,656)		8565 (13,643)	3420 (976.5)	
Range	177-56,475	694-53,952	177-56,475		177-56,475	839-48,935		694-53,952	1236-4857	
SEM	1321	2237	1640		1885	11,828		3523	325.5	
95% CI	2568-7833	2008-11,263	1215-7808		742.6-6724	–24,180 to 51105		1010-16,120	2670-4171	
Lactate dehydrogenase (U/L)				.001			.57			.34
Mean (SD)	572.8 (233.4)	723 (234.2)	519.6 (210.3)		523.8 (215.5)	480.3 (163.1)		678.3 (227.3)	776.6 (242.9)	
Range	165-1345	387-1345	165-1009		165-1009	279-671		387-1069	567-1345	
SEM	25.47	49.94	26.71		28.80	66.57		65.62	76.83	
95% CI	522.2-623.5	619.1-826.9	466.2-573		466-581.5	309.2-651.5		533.9-822.8	602.8-950.4	
Ferritin (ng/ml)				.95			.24			.52

Parameters	Total	I-T[a]	NI-T[b]	*P* value	NI-L[c]	NI-D[d]	*P* value	I-L[e]	I-D[f]	*P* value
Mean (SD)	1678 (2678)	1702 (1932)	1667 (2952)		1717 (3086)	1148 (489.6)		1897 (2405)	1429 (1021)	
Range	36-17,860	84-9366	36-17,860		36-17,860	739-1901		264-9366	84-2899	
SEM	297.6	394.4	391		427.9	218.9		642.9	322.8	
95% CI	1085-2217	885.4-2518	884.1-2451		858.2-2576	540.2-1756		507.8-3285	698.7-2159	
Prothrombin time (sec)				.49			.65			.57
Mean (SD)	14.38 (3.46)	14.87 (2.889)	14.24 (3.612)		14.06 (2.740)	16.58 (10.09)		15.16 (3.208)	14.30 (2.330)	
Range	10.20-31.17	10.50-20.80	10.20-31.17		10.20-25.50	11.10-31.70		10.50-20.80	11.80-17	
SEM	0.4135	0.7459	0.4870		0.3837	5.046		1.015	1.042	
95% CI	13.55-15.20	13.27-16.47	13.27-15.22		13.29-14.83	0.5163-32.63		12.86-17.46	11.41-17.19	
Partial thromboplastin time (sec)				.16			.67			.84
Mean (SD)	39.29 (21.97)	50.86 (37.09)	35.77 (13.47)		35.94 (13.96)	34.05 (7.175)		52.39 (39.49)	48.10 (36.57)	
Range	19.80-120	22.90-120	19.80-108.7		19.80-108.7	27.60-42.80		22.90-120	23.50-109	
SEM	2.837	9.913	1.986		2.155	3.587		13.16	16.35	
95% CI	33.62-44.97	29.44-72.27	31.77-39.77		31.59-40.29	22.63-45.47		22.03-82.75	2.696-93.50	
Troponin (ng/ml)				.72			.29			.49
Mean (SD)	0.04039 (0.09629)	0.04862 (0.1185)	0.03853 (0.09118)		0.03198 (0.07648)	0.09290 (0.1674)		0.02964 (0.03198)	0.06950 (0.1708)	
Range	0.01-0.56	0.01-0.5550	0.01-0.56		0.01-0.4920	0.01-0.56		0.01-0.0970	0.01-0.5550	
SEM	0.009018	0.02587	0.009455		0.008395	0.05292		0.009643	0.05401	
95% CI	0.023-0.058	–0.005 to 0.103	0.02-0.057		0.016-0.049	–0.027 to 0.213		0.008-0.051	–0.053 to 0.19	
Brain natriuretic peptide (pg/ml)				.12			.56			N/A[h]
Mean (SD)	4061 (12,513)	1260 (1223)	4419 (13,254)		4596 (13,975)	2927 (3931)		1295 (1364)	1085 (0)	
Range	14-70,000	23-3379	14-70,000		14-70,000	164-9815		23-3379	1085	
SEM	1719	499.1	1933		2156	1758		609.8	0	
95% CI	612.1-7510	–23.40 to 2543	527-8310		241.5-8951	–1955 to 7808		–398.5 to 2988	0	
Glucose (mg/ml)				.01			.18			.07
Mean (SD)	231.3 (169.8)	297.3 (145.7)	218.4 (171.5)		218.4 (171.5)	287.2 (208.2)		253.9 (124.2)	354.2 (156.8)	
Range	78-1466	135-720	78-1466		78-1466	119-899		135-576	157-720	
SEM	12.51	26.61	13.82		13.82	47.76		30.12	43.49	

Parameters	Total	I-T[a]	NI-T[b]	*P* value	NI-L[c]	NI-D[d]	*P* value	I-L[e]	I-D[f]	*P* value
95% CI	206.6-256	242.9-351.8	191.1-145.7		191.1-145.7	186.8-387.5		190-317.7	259.4-448.9	

[a]I-T: intubated total.

[b]NI-T: nonintubated total.

[c]NI-L: nonintubated living.

[d]NI-D: nonintubated deceased.

[e]I-L: intubated living.

[f]I-D: intubated deceased.

[g]SEM: standard error of the mean.

[h]Not applicable.

Procalcitonin

There was significant statistical difference between the NI-T and I-T groups (95% CI 0.7077-12.04, *P*=.03); however, we found no difference between NI-D and NI-L (95% CI 1.322-5.607, *P*=.20) or between I-D and I-L (*P*=.23).

C-Reactive Protein

Comparison of NI-T and I-T yielded a *P* value of .006 (95% CI 2.516-14.82), and a comparison of NI-D and NI-L had a *P* value of .24 (95% CI –21.20 to 64.19). Finally, I-D compared to I-L had a *P* value of .76 (95% CI –11.77 to 8.871).

Ferritin

There was no statistical significance in all 3 comparison groups, which were as follows: NI-T and I-T (95% CI 1075-1144, *P*=.95), NI-D and NI-L (95% CI –1538 to 400, *P*=.24), I-D and I-L (95% CI –1975 to 1040, *P*=.52).

Lactate Dehydrogenase

We found that there was statistical significance of LDH between NI-T and I-T (95% CI 88.31-318.6, *P*=.001), while no differences were found between I-D and I-L or the NI-D and NI-L groups (*P*=.57 and *P*=.34, respectively).

Fever

The mean temperature was 100.9° F (SD 1.68, SEM 0.1239) across all patients. Plotted values for each subgroup are shown in Table 1. On comparing the temperature between NI-T and I-T, we found statistical significance (*P*=.004) with means of 100.7° F and 101.8° F, respectively. There was no significant difference linking temperature to mortality among the remainder of the subgroup analysis.

Prothrombin Time

There was no statistical significance across all three groups for PT: NI-T vs I-T (95% CI –1.1997 to 2.457, *P*=.49), NI-D vs NI-L (95% CI –13.49 to 18.52, *P*=.65), and I-D vs I-L (*P*=.57).

Partial Thromboplastin Time

No statistical difference between all 3 comparison groups with the following: *P*=.16, *P*=.67, and *P*=.84 comparing NI-T vs I-T and then between the NI and I subgroups, respectively.

D-Dimers

D-dimers were elevated on most of our patients with SARS-CoV-2 confirmed cases. There were no significant differences among the groups; although, the intubated patients had a mean that showed a clear trend of elevation compared to those nonintubated. NI-T vs I-T had a *P* value of .45, with a mean for NI-T of 4512 and for I-T of 6636. Further analysis between the deceased and living patients of the nonintubated and intubated groups yielded *P*=.47 and *P*=.17, respectively.

Glucose

There was statistical significance between the NI-T and I-T groups, who had mean glucose levels of 218.4 and 297.3, respectively (95% CI 18.46-139.3, *P*=.01). There was no significance between the nonintubated subgroups NI-D and NI-L (*P*=.18); however, it is noteworthy that the mean glucose levels among the deceased were 287.2 compared to 218.4 for the living. We noted the same trends among the intubated group with a mean of 354.2 among the deceased vs 253.9 for I-L. The comparison of I-D vs I-L yielded a *P* value of .07.

Creatinine

There was a clear trend of increasing Cr values, indicating acute kidney injury (AKI) proportional to the disease severity regardless of their pre-existing renal function status. NI-T vs I-T was statistically significant (*P*=.03), with a mean Cr of 2.486 and 4.030, respectively (95% CI 0.1641-2.924). Although there was no statistical significance among the NI-D vs NI-L, it is quite clear that up trending Cr was indicative of worsening renal function among the deceased *P*=.09. The patients intubated overall had worse renal function, but there was no significance between groups (*P*=.77).

Blood Urea Nitrogen

We performed the same grouped analysis for BUN. There was significant differences correlating that of Cr among the NI-T and I-T (95% CI 11.54-43.82, *P*=.001). There was statistical significance among the living and deceased in the nonintubated population (95% CI 1.148-47.85, *P*=.04). No statistical differences were detected between the I-L and I-D group (*P*=.56).

Troponin

Although troponin had been elevated, indicating myocardial stress, in most of our patients that had it drawn, a correction

was done for those with and without pre-existing CAD and potential influence with coexisting CKD. We did note an elevation across all groups. NI-T and I-T had means of 0.03853 and 0.04862, respectively (*P*=.72); NI-D vs NI-L had a *P* value of .28, and I-D vs I-L had a *P* value of .48.

Brain Natriuretic Peptide

There was no significance for all three groups. Of note, we had lower sample size and reduced power in the groups after correction; only a few of our patients had BNP drawn. Among those that had it drawn, most of them had their values elevated. NI-T vs I-T had a *P* value of .12, and NI-D vs NI-L had a *P* value of .56.

Electrocardiogram

Our EKG findings among patients and their distribution across various subgroups are described in Multimedia Appendix 1. Predominant EKG findings were sinus rhythm (n=94/160, 58.75%); sinus tachycardia (n=34, 21.25%); sinus tachycardia with prolonged QT (n=3, 1.875%); prolonged QT (n=6, 3.75%); and arrhythmia, blocks, or other (n=23, 14.38%). Comparison of NI-T and I-T yielded a *P* value of .08 for tachycardia; comparison of NI-L and NI-D had a *P* value of .02 for arrhythmias and heart blocks, indicating that patients with pre-existing arrhythmias or heart blocks had positive associations with mortality.

Chest X-Ray

We classified CXR findings into 7 major classes as reported by the radiologist for all the patients in the study group. We sorted out associations as depicted in Table 4 with percent distributions, group analysis, and *P* values. Of significance, we found that NI-T and I-T comparisons were statistically significant (*P*=.02) for bilateral (B/L) patchy opacities, which was a predominant finding among the nonintubated patients and with B/L diffuse pulmonary infiltrates predominantly in the intubated group, for which the comparison of NIT to I-T was statistically significant (*P*<.001).

Table 4. Chest x-ray associations with the coronavirus disease.

CXR[a] findings	Total (N=184), n (%)	Nonintubated total (n=154), n (%)	Intubated total (n=30), n (%)	*P* value	Nonintubated living (n=135), n (%)	Nonintubated deceased (n=19), n (%)	*P* value	Intubated living (n=17), n (%)	Intubated deceased (n=13), n (%)	*P* value
Patchy bilateral opacity	51 (27.72)	48 (31.17)	3 (10)	.02	42 (31.11)	6 (31.58)	.97	1 (5.88)	2 (15.38)	.40
Patchy unilateral opacity	23 (12.5)	21 (13.64)	2 (6.67)	.30	19 (14.08)	2 (10.53)	.67	1 (5.89)	1 (7.69)	.84
Diffuse bilateral infiltrates	84 (45.65)	60 (38.96)	24 (80)	<.001	51 (37.78)	9 (47.37)	.42	14 (82.35)	10 (76.92)	.71
Diffuse unilateral infiltrate	1 (0.54)	1 (0.65)	0 (0)	.66	1 (0.74)	0 (0)	.71	0 (0)	0 (0)	N/A[b]
Ground glass opacity	2 (1.09)	2 (1.30)	0 (0)	.53	2 (1.48)	0 (0)	.59	0 (0)	0 (0)	N/A
Increased interstitial marking	13 (7.07)	12 (7.79)	1 (3.33)	.38	10 (7.41)	2 (10.53)	.64	1 (5.88)	0 (0)	.37
Normal CXR	10 (5.43)	10 (6.49)	0 (0)	.15	10 (7.41)	0 (0)	.22	0 (0)	0 (0)	N/A

[a]CXR: chest x-ray.

[b]Not applicable.

Treatment

Of the total 184 patients, 153 patients received some form of treatment as categorized in Table 5 of our manuscript. The primary treatment groups were group 1 receiving azithromycin and hydroxychloroquine, group 2 receiving Azithromycin only, group 3 receiving hydroxychloroquine only, and group 4 no treatment. They were not grouped based on any specific parameters, but treatments were given with emerging science and potential alleviation of the disease process. Percent distribution and *P* values are described in Table 5. We found that no combination of treatment made any difference in preventing the need for mechanical ventilation or mortality.

Table 5. Treatment groups for the coronavirus disease.

Treatment	Total (N=184), n (%)	Nonintubated (n=154), n (%)	Nonintubated living (n=135), n (%)	Nonintubated deceased (n=19), n (%)	Intubated (n=30), n (%)	Intubated living (n=17), n (%)	Intubated deceased (n=13), n (%)
Azithromycin and hydroxychloroquine	90 (48.91)	71 (46.10)	63 (46.67)	8 (42.11)	19 (63.33)	11 (64.71)	8 (61.54)
Azithromycin	23 (12.5)	23 (14.94)	20 (14.81)	3 (15.79)	0 (0)	0 (0)	0 (0)
Hydroxychloroquine	40 (21.74)	31 (20.13)	27 (20)	4 (21.05)	9 (30)	5 (29.41)	4 (30.77)
No treatment	31 (16.85)	29 (18.83)	25 (18.52)	4 (21.05)	2 (6.67)	1 (5.88)	1 (7.69)

Discussion

Our study population was primarily grouped into patients that required admission with and without the need for mechanical ventilation, we further classified and subgrouped them to deceased and living, this renders a deeper perspective on the entire spectrum of COVID-19 among the hospitalized. Our patient population is unique in that it gives a fair representation of the American minority. It was interesting to note that in our study younger patients were more likely to need mechanical ventilation than their older counterparts, while the older patients had increased mortality, and this positively correlated with increasing comorbidity. We also found that the comparatively lower BMI group had worse prognosis in both needing mechanical ventilation and mortality contrary to our traditional beliefs [5]. Increased core body temperature corresponded to worsening disease. Our study had striking differences in that more African Americans with increased comorbidities had higher mortality and decreased need for mechanical ventilation compared to Hispanics, who were more likely to be intubated and then succumb to the disease. HTN was independently associated with mortality among the nonintubated. Inflammatory markers were increased across the board for all of our patients with significant differences among the I-T and NI-T groups and trends between the deceased and living; this certainly proves that they act as surrogate markers of disease severity in SARS-CoV-2 infection. Recent studies linked AKI to COVID-19 [6], our study not only confirmed their findings but also showed directly proportional worsening of renal function with increasing disease severity. Mean blood glucose followed the same trajectory, corresponding to disease severity indicative of insulin resistance insurgence [7]. Pre-existing conduction defects, atrial fibrillation, and arrhythmias positively correlated with mortality among the nonintubated. We also noted that patients with B/L CXR findings of patchy opacities were more likely to be hospitalized without the need for mechanical ventilation and recovery versus those with B/L diffuse infiltrates, who were more likely to be intubated. Our study suggests azithromycin and hydroxychloroquine treatment when given individually or as combination therapy did not alter disease progression and mortality.

Although our study clearly illustrates associations of age, BMI, ethnicity, body temperature, multiple comorbidities, HTN, DM, CAD, inflammatory markers, mean blood glucose, AKI, EKG findings, and CXR findings with likelihood of needing intubation or mortality, it is not without limitations. This is an early retrospective and not a randomized controlled study. Sample size, demographics, and ethnic composite in our study needs to be taken into account while interpreting results. We urge the readers, clinicians, and researchers to take our early findings, bearing in mind its limitations while making clinical decisions. Further investigation and research on a large scale based on our early findings is essential.

Acknowledgments

We thank the infectious disease department, Department of Surgery, and Hector Osei, MD.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Electrocardiogram findings.
[DOCX File , 17 KB - jmir_v22i9e20548_app1.docx]

References

1. Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen
2. Correal A, Jacobs A. 'A Tragedy is unfolding': inside New York's virus epicenter. The New York Times. 2020 Apr 09. URL: https://www.nytimes.com/2020/04/09/nyregion/coronavirus-queens-corona-jackson-heights-elmhurst.html)

3. Cevik M, Bamford C, Ho A. COVID-19 pandemic-a focused review for clinicians. Clin Microbiol Infect 2020 Apr 25:30231-30237 [FREE Full text] [doi: 10.1016/j.cmi.2020.04.023] [Medline: 32344166]
4. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
5. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, LICORNthe Lille COVID-19 and Obesity study group. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. Obesity (Silver Spring) 2020 Apr 09:1 [FREE Full text] [doi: 10.1002/oby.22831] [Medline: 32271993]
6. Ronco C, Reis T. Kidney involvement in COVID-19 and rationale for extracorporeal therapies. Nat Rev Nephrol 2020 Jun;16(6):308-310 [FREE Full text] [doi: 10.1038/s41581-020-0284-7] [Medline: 32273593]
7. Drucker DJ. Coronavirus infections and type 2 diabetes-shared pathways with therapeutic implications. Endocr Rev 2020 Jun 01;41(3):1 [FREE Full text] [doi: 10.1210/endrev/bnaa011] [Medline: 32294179]

Abbreviations

AKI: acute kidney injury
BA: bronchial asthma
BNP: blood natriuretic peptide
BUN: blood urea nitrogen
B/L: bilateral
CAD: coronary artery disease
CKD: chronic kidney disease
COPD: chronic obstructive pulmonary disease
COVID-19: coronavirus disease
Cr: creatinine
CRP: C-reactive protein
CXR: chest x-ray
DM: diabetes mellitus
EKG: electrocardiogram
HTN: hypertension
I-D: intubated deceased
I-L: intubated living
I-T: intubated total
LDH: lactate dehydrogenase
NI-D: nonintubated deceased
NI-L: nonintubated living
NI-T: nonintubated total
PT: prothrombin time
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SEM: standard error of the mean

Edited by G Eysenbach; submitted 21.05.20; peer-reviewed by A Mitrokhin, A Rajabalan, C Benjamin, S Raviralla; comments to author 10.06.20; revised version received 14.06.20; accepted 14.06.20; published 18.09.20.

Please cite as:
Mani VR, Kalabin A, Valdivieso SC, Murray-Ramcharan M, Donaldson B
New York Inner City Hospital COVID-19 Experience and Current Data: Retrospective Analysis at the Epicenter of the American Coronavirus Outbreak
J Med Internet Res 2020;22(9):e20548
URL: https://www.jmir.org/2020/9/e20548
doi:10.2196/20548
PMID:32540837

Original Paper

Use of Telemedicine for Chronic Liver Disease at a Single Care Center During the COVID-19 Pandemic: Prospective Observational Study

Maria Guarino[1*], MD, PhD; Valentina Cossiga[1*], MD; Andrea Fiorentino[1], MD; Giuseppina Pontillo[1], MD; Filomena Morisco[1], MD, Prof Dr

Gastroenterology Unit, Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy

[*]these authors contributed equally

Corresponding Author:
Maria Guarino, MD, PhD
Gastroenterology Unit
Department of Clinical Medicine and Surgery
University of Naples Federico II
Via Sergio Pansini
Naples
Italy
Phone: 39 0817464746
Email: maria.guarino86@gmail.com

Abstract

Background: The COVID-19 outbreak has overwhelmed and altered health care systems worldwide, with a substantial impact on patients with chronic diseases. The response strategy has involved implementing measures like social distancing, and care delivery modalities like telemedicine have been promoted to reduce the risk of transmission.

Objective: The aim of this study was to analyze the benefits of using telemedicine services for patients with chronic liver disease (CLD) at a tertiary care center in Italy during the COVID-19–mandated lockdown.

Methods: From March 9 to May 3, 2020, a prospective observational study was conducted in the Liver Unit of the University Hospital of Naples Federico II to evaluate the impact of (1) a fully implemented telemedicine program, partially restructured in response to COVID-19 to include video consultations; (2) extended hours of operation for helpline services; and (3) smart-working from home to facilitate follow-up visits for patients with CLD while adhering to social distancing regulations.

Results: During the lockdown in Italy, almost 400 visits were conducted using telemedicine; only patients requiring urgent care were admitted to a non–COVID-19 ward of our hospital. Telemedicine services were implemented not only for follow-up visits but also to screen patients prior to hospital admission and to provide urgent evaluations during complications. Of the nearly 1700 patients with CLD who attended a follow-up visit at our Liver Unit, none contracted COVID-19, and there was no need to alter treatment schedules.

Conclusions: Telemedicine was a useful tool for following up patients with CLD and for reducing the impact of the COVID-19 pandemic. This system of health care delivery was appreciated by patients since it gave them the opportunity to be in contact with physicians while respecting social distancing rules.

(J Med Internet Res 2020;22(9):e20874) doi:10.2196/20874

KEYWORDS

telemedicine; COVID-19; hepatology; telehealth; liver disease; Italy; hospital; chronic disease; liver

Introduction

COVID-19, caused by the recently identified coronavirus SARS-CoV-2, has been spreading rapidly across the world since December 2019. The COVID-19 pandemic constitutes a great health emergency, with 3,090,445 confirmed cases and 217,769 deaths globally, as of April 30, 2020 [1]. Most patients experienced fever, dry cough, and asthenia, but there is a wide spectrum of minor symptoms, such as diarrhea, anosmia, and ageusia. Patients experiencing a severe course of disease had a higher incidence of pneumonia and acute respiratory distress

syndrome, and required oxygen and mechanical ventilation more often, compared to those with nonsevere disease [2].

Patients with comorbidities, such as hypertension, diabetes, and obesity, as well as older adults, have a higher risk of complications and mortality [3]. There is a lack of data on the course of COVID-19 in subjects with pre-existing liver diseases, but it is reasonable to conclude that this category of patients are also at a higher risk of severe outcomes. For example, patients with autoimmune liver disease or liver cancer undergoing immunosuppressive therapy, or cirrhotic patients with an altered immune response would be more susceptible to infections [4].

To contain the rapid spread of COVID-19, social distancing became a necessary prevention strategy [5]. However, pandemic responses are disrupting routine care for non–COVID-19 patients. It has become fundamental to define the prioritization criteria for outpatient visits, to avoid unnecessary in-hospital visits, and to facilitate the management of patients in the home setting. Telemedicine, a virtual care platform that allows communication between health care professionals and patients, has been proposed as an indispensable tool to reduce COVID-19 spread and to maintain care of patients with chronic disease [6-8]. The World Health Organization [9] described telemedicine as:

> *The delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities.*

Telemedicine is now well accepted because it is patient centered, protects patients and clinicians from viral exposure, and provides health outcomes comparable to traditional methods of health care delivery without compromising the patient-clinician relationship and enhancing patient satisfaction [10].

In particular, subjects with chronic disease, such as chronic liver disease (CLD), require assistance from their physician via scheduled outpatient appointments to solve daily clinical issues (ie, monitoring of patient weight, diuresis, and serum electrolytes during a diuretic treatment for decompensated cirrhosis, or checking heart rate in patients undergoing nonselective β-blockers treatment for the first time for variceal bleeding prevention). For these reasons, a telemedicine service has been in effect since 2015 at the Liver Unit of the University Hospital of Naples Federico II. In 2020, it was already fully implemented with well-trained physicians. The only equipment requirements necessary for this service were smartphones, printers, laptops, and tablet computers with an internet connection. This service allows us to stay in contact with patients daily by email, fax, and phone call. During the last 5 years, we improved the quality of clinical care, achieving patient acceptability and trust. Moreover, the use of text messaging, calling, and remote monitoring decreased the number of unscheduled visits. During the COVID-19 outbreak, we further improved telemedicine services to follow international recommendations [11-14] and postpone nonurgent outpatient visits.

The aim of this study was to prospectively analyze the benefits of using telemedicine for patients with CLD in a tertiary care center during the COVID-19 lockdown in Italy.

Methods

A prospective observational study was conducted to evaluate a telemedicine service set up for patients with CLD from March 9 to May 3, 2020, at the Liver Unit of the University Hospital of Naples Federico II during the lockdown. During this period of social distancing, high-acuity care was prioritized, and elective procedures and routine care were postponed.

In the last 2 months, several hospitals in Italy and their liver units became COVID-19 hotspots and designated wards during the pandemic, and patients with CLD were instructed to remain at home to avoid the risk of infection. In the Campania region, there were fewer cases than in regions of Northern Italy (4331 confirmed cases, as of April 25, 2020) [15]. For this reason, our Liver Unit continued to take care of patients with CLD but changed standard procedures for outpatient care. The waiting room, for example, was readapted to facilitate adequate distancing between patients requiring necessary visits, waiting time were reduced, and companions were not permitted.

The telemedicine service, implemented for patients with CLD in 2015, was organized as follows:

- Helpline: available from Monday to Friday between 2 PM and 5 PM, managed by 5 trained physicians. This service collects all requests for medical consultations, concerns with treatment plans, side effects of drugs, visit schedules, and state of disease. This real-time modality allows specialists, general practitioners, and patients to speak with one another in real time to discuss conditions;
- Secured email and fax services: available 24 hours a day, aimed to send treatment schedules, laboratory tests results, medical imaging such as X-rays, photos, ultrasound recordings, or other static and video medical imaging to remote specialists for analysis and future consultation, which is particularly relevant for patients with CLD requiring close monitoring.

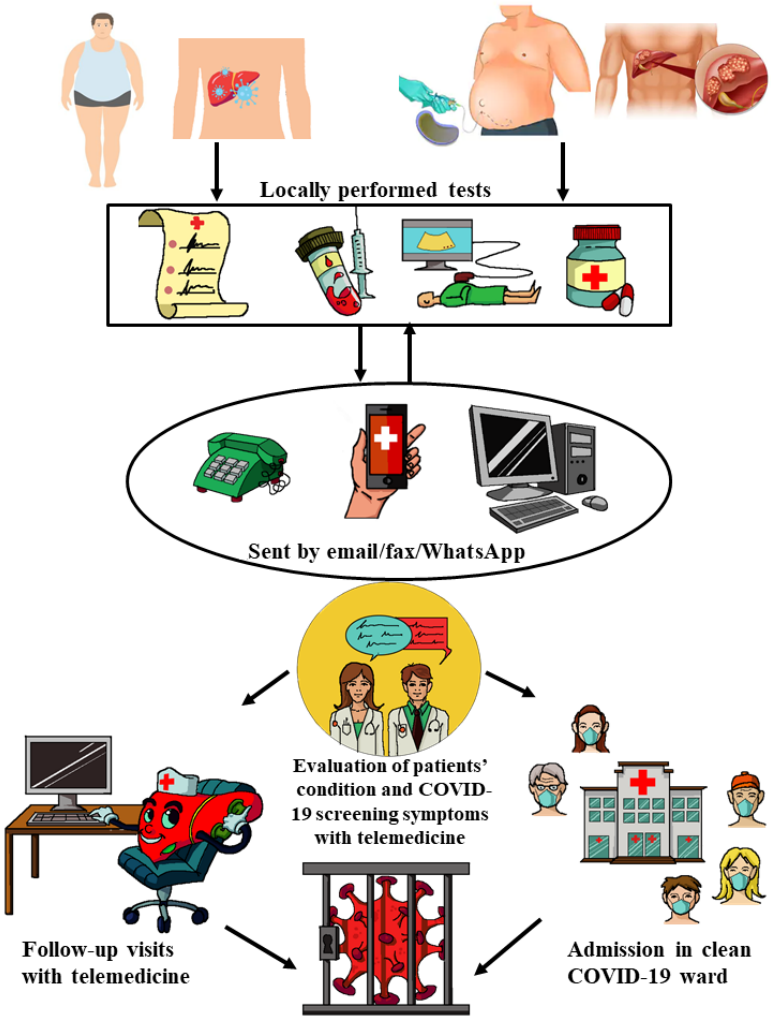
During the lockdown in Italy, these services were further improved to limit face-to-face contact by implementing the following:

- Video consultations (through Skype or WhatsApp). In the beginning, during Italy's state of emergency, Skype and WhatsApp were used for video consultations due to the lack of a secure telemedicine software, which was made available in our Liver Unit starting in May 2020;
- Extended helpline hours, now available from Monday to Friday between 8 AM to 1 PM and 3 PM to 6 PM, as well as during weekends;
- Smart-working from home to increase remote patient monitoring and mobile health care while reducing the risk due to contact with medical staff.

Results

During Italy's lockdown, 480 outpatient visits were scheduled but, in order to avoid hospital admissions, most were conducted via telemedicine (Figure 1). In-hospital visits were limited only to patients with urgent needs or with oncological conditions, since international regulations on social isolation and guidelines from major hepatology associations encouraged conducting appointments via telemedicine [11-14].

Figure 1. Graphical representation of telemedicine activity during the COVID-19 pandemic.



In particular, 80 follow-up visits (in a population of 250 patients) were scheduled for patients with chronic viral hepatitis (hepatitis B virus or hepatitis C virus) undergoing antiviral treatment (nucleos(t)ide analogues or direct-acting antivirals). For these patients, clinical conditions were evaluated by phone and laboratory tests; ultrasound was performed locally and follow-up prescriptions were sent by secure email and/or fax. Similarly, 120 follow-up visits (in a population of 1000 patients) were scheduled for patients who had obtained a sustained virological response after antiviral treatment (interferon or direct-acting antiviral schedules); for these patients, clinical conditions were evaluated by teleconsultations and laboratory tests; ultrasound was performed locally and follow-up prescriptions were sent by email and/or fax.

Subsequently, 40 outpatient visits in a population of 200 patients with alcoholic liver disease, nonalcoholic fatty liver disease, or nonalcoholic steatohepatitis were replaced by teleconsultations and full telemedicine services. In addition, other conditions, such as primary biliary cholangitis, primary sclerosing cholangitis, and autoimmune hepatitis were managed in the same way (approximately 30 outpatient visits in a population of 60 patients). At our center, immunosuppressive treatment for autoimmune hepatitis was not modified in any way.

Additionally, 15 outpatient visits in a population of 40 patients with CLD associated with a rare disease (eg, cystic fibrosis, common variable immunodeficiency disorders, Crigler-Najjar syndrome, hereditary hemochromatosis, Alagille syndrome, congestive hepatopathy, nodular regenerative hyperplasia, etc) were managed with telemedicine services.

Patients with decompensated cirrhosis are at risk of worse outcomes if infected with COVID-19; hence, it is necessary to consider minimizing these patients' exposure to medical staff. During the lockdown period, we managed 50 patients with compensated cirrhosis by teleconsultations, stressing the importance of prophylaxis measures for hepatic encephalopathy and spontaneous bacterial peritonitis to avoid decompensation and reduce hospital admissions. Laboratory tests were performed locally to evaluate electrolytes and renal function in patients

with ascites undergoing diuretic therapy. In patients requiring evacuative paracentesis, this procedure was conducted aseptically in a non–COVID-19 ward of the hospital (15 admissions).

Thirty outpatient visits for patients with a previous diagnosis of hepatocellular carcinoma (HCC), who were "disease-free" at the time, were replaced by telemedicine. Conversely, patients with HCC (a new diagnosis or disease recurrence) were managed as usual in a non–COVID-19 ward of the hospital (30 admissions). In particular, HCC staging and treatment were provided; we were able to offer patients several options according to HCC burden and stage (surgery, ablation techniques, transarterial chemoembolization / transarterial embolization, and systemic treatments). Moreover, for patients with HCC requiring follow-ups to evaluate treatment response, imaging tests were unaffected by the lockdown and were conducted in a timely manner. For HCC screening (performed in ultrasound centers in the Campania Region), ultrasounds was delayed, since according to International Liver Cancer Association guidelines, delays of 1-2 months in HCC screening do not significantly increase risk [13].

Among the approximately 70 liver transplanted patients handled at our center, 50 follow-up visits were scheduled during the lockdown, but the majority was managed via telemedicine. Laboratory tests, including assessment of immunosuppressive drug levels, were performed locally. Only 15 patients had clinical conditions requiring in-hospital visits and close face-to-face monitoring (due to cytomegalovirus infection, immunosuppressive drugs toxicity management, biliary complications, management of comorbidities exacerbation like Crohn disease, and phlebotomy in patients with polyglobulia). Even among transplant patients, immunosuppressive treatment was not modified.

Overall, 200-250 tele- or video consultations (over the course of 9 weeks) were conducted per week and 130-150 phone calls were received, while 100-150 emails and/or faxes were received per week and 150-200 were sent, replacing outpatient visits. Patients expressed satisfaction with this method of management and did not encounter any difficulties with contacting us, since they were strongly aware of the risks associated with leaving home.

In addition to helping patients to manage chronic liver diseases, the primary goals of implementing a telemedicine system were as follows:

1. Screen patients for COVID-19 prior to admission (inpatients; outpatient visits for HCC or decompensated chronic advanced liver disease or liver transplant patients) according to the guidelines [16] implemented to keep our center safe;
2. Provide telemedicine follow-up visits for all patients with CLD and nonurgent conditions for a face-to-face visit;
3. Provide routine care to our patients with CLD, including promptly addressing questions, coordinating complex care, offering caregiver support (eg, in case of hepatic encephalopathy through early recognition and initiation of prophylactic therapies), nutritional advices, and early interventions to ensure drugs compliance (especially for immunosuppression drugs in liver transplant patients), and to prevent decompensation for chronic advanced liver disease while enabling patients to stay at home;
4. Provide urgent evaluations during decompensation events to minimize emergency room visits and admissions.

At the end of this period, none of nearly 1700 patients with CLD who followed up at our Liver Unit contracted COVID-19. This success was possibly due to measures adopted by the Italian government and the medical care guaranteed by our established and further improved telemedicine services.

Discussion

Telemedicine is a rapidly expanding health care delivery modality with increasing utility for health care. During the COVID-19 pandemic, telemedicine became a useful tool for remotely connecting patients with health care providers and facilitating follow-up visits via smartphones or webcam-enabled computers.

In the management of patients with CLD, telemedicine has already been successful in different cases, such as hepatitis C therapy in prison populations or to enhance the efficiency of liver transplant evaluations [17-19]. The current study has shown that a fully implemented telemedicine service, partially restructured for the COVID-19 pandemic, can convert more than 75% of planned outpatients visits to remote appointments.

Moreover, the impact of the COVID-19 pandemic on routine health care should not be underestimated, since it has resulted in the need to completely change usual follow-up procedures, such as frequent patient-physician contact to evaluate disease status, screen for complications, and assess response to therapy. Tapper and Asrani [20] described the ways in which COVID-19 impacted the quality of cirrhosis care and discussed this impact in three waves: (1) a delay in routine care for cirrhotic patients such as deferred screening for HCC and for varices with late diagnosis, or canceled elective therapeutic procedures; (2) a backlog of outpatient visits, resulting in a large workload for physicians and an increase in acute decompensation in patients classified as low risk in the first wave when normal clinical care is resumed; (3) the consequences due to missed diagnosis that will persist for years to come. The authors identified several measures to avoid these complications like the use of telemedicine as a substitute for outpatient visits during social distancing [20].

Our study showed that a sizeable proportion (about 75%) of outpatient visits in a CLD setting can be managed effectively with telemedicine without compromising patients' health or quality of care. Furthermore, with our telemedicine system we did not postpone appointments, preventing any unintended loss to follow-up, since we scheduled a new appointment for each patient according to their specific condition. Above all, we managed to avoid a surge in outpatient activity after the lockdown due to a backlog of postponed care.

No resistance was encountered by patients to using this modality of care delivery since it was both familiar to them and kept both patients and medical staff protected against COVID-19. Indeed, the prevalence of high-speed internet and smartphones makes

it possible to apply this framework easily to set up video teleconsultations from a patient's home. Patients highlighted the faster and increased availability of medical care through telemedicine. Moreover, they were thankful to receive an expert opinion whether it was a reassurance or a clinical evaluation.

After the lockdown, we plan to continue with telemedicine services to keep in contact and manage in a better way patients who cannot come to the hospital for scheduled visits due to work-related obligations or disabling pathologies. Moreover, telemedicine could be used in the future to follow up with specific categories of patients, such as hepatitis C long-term responders, to prevent in-hospital overcrowding.

It should be acknowledged that telemedicine services have some limitations: (1) the absence of physical interactions (eg, eye contact, handshake), which may play a comforting role for patients; (2) interferences, lapses, delays, or interruptions due to service connection problems; and (3) difficulties associated with communicating bad news [21]. In a time of crisis, such as the COVID-19 pandemic, these difficulties can be overcome but must be taken into considerations if the use of telemedicine is to be integrated into routine clinical care. In fact, telemedicine needs to be defined by national regulations and frameworks for public health emergencies. A plan is necessary to increase physicians' expertise and monitor patients remotely, as well as to educate the population on the correct use of this service before it can be globally adopted [8].

The COVID-19 pandemic has dramatically changed several aspects of health care; during the lockdown, it became essential to use telemedicine and virtual software. In Italy, there are no national regulations for telemedicine use; only in some tertiary care centers were there remote services such as secured mail, fax, and dedicated helplines to follow up with patients. During COVID-19 outbreak, each Italian region, including Campania, has attempted to implement remote assistance systems to manage patients with chronic disease.

One limitation of our study is its cross-sectional design. A longitudinal study evaluating the long-term impact of telemedicine on the course of CLD would provide additional data to improve the system and to optimize outcomes for remote patients in order to reduce missed diagnoses, progressive disease outcomes, and loss to follow-up. In this context, additional studies are needed to evaluate the use of telemedicine in the long term to enhance, and not replace, the current standard of care.

In conclusion, the COVID-19 pandemic has deeply strained health care systems around the world. To mitigate the impact of disease, it is fundamental to minimize the risk of patients' and physicians' exposure to the virus; in this scenario, telemedicine played an important role. Although telemedicine will not solve all health-related problems, its well-standardized use in our Liver Unit demonstrated its utility in times of social distancing. Finally, the use of telemedicine systems provided us with the opportunity to conduct scheduled visits remotely for patients with CLD while reducing a postlockdown surge in outpatient visits.

Acknowledgments

We thank Dr Luca Pignata and Dr Mario Capasso for their support with our telemedicine system, and Alfonso De Vivo for his graphical support.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-04-30]
2. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
3. Ruan Q, Yang K, Wang W, Jiang L, Song J. Correction to: Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020 Jun 06;46(6):1294-1297 [FREE Full text] [doi: 10.1007/s00134-020-06028-z] [Medline: 32253449]
4. Zhang C, Shi L, Wang F. Liver injury in COVID-19: management and challenges. The Lancet Gastroenterology & Hepatology 2020 May;5(5):428-430 [FREE Full text] [doi: 10.1016/S2468-1253(20)30057-1] [Medline: 32145190]
5. Sjödin H, Wilder-Smith A, Osman S, Farooq Z, Rocklöv J. Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020. Euro Surveill 2020 Apr;25(13):e [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.13.2000280] [Medline: 32265005]
6. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998 [FREE Full text] [doi: 10.1136/bmj.m998] [Medline: 32165352]
7. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

8. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
9. Telemedicine: Opportunities and developments in member states. World Health Organization. 2010. URL: https://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2020-04-26]
10. Dorsey ER, Topol EJ. Telemedicine 2020 and the next decade. The Lancet 2020 Mar 14;395(10227):859. [doi: 10.1016/S0140-6736(20)30424-4] [Medline: 32171399]
11. Boettler T, Newsome PN, Mondelli MU, Maticic M, Cordero E, Cornberg M, et al. Care of patients with liver disease during the COVID-19 pandemic: EASL-ESCMID position paper. JHEP Rep 2020 Jun;2(3):100113 [FREE Full text] [doi: 10.1016/j.jhepr.2020.100113] [Medline: 32289115]
12. Fix OK, Hameed B, Fontana RJ, Kwok RM, McGuire BM, Mulligan DC, et al. Clinical Best Practice Advice for Hepatology and Liver Transplant Providers During the COVID-19 Pandemic: AASLD Expert Panel Consensus Statement. Hepatology 2020 Jul 16;72(1):287-304 [FREE Full text] [doi: 10.1002/hep.31281] [Medline: 32298473]
13. Meyer T, Chan S, Park J. Management of HCC during COVID-19 pandemic: ILCA guidance. ILCA. URL: https://ilca-online.org/wp-content/uploads/2020/04/webinar-on-management-of-hcc-during-covid-19_-ilca-guidance-pdf.pdf [accessed 2020-04-30]
14. AISF Recommendations. AISF. URL: https://www.webaisf.org/wp-content/uploads/2020/03/Prevenzione-in-epatopatici.pdf [accessed 2020-04-30]
15. The Campania region data. Regione Campania. URL: http://www.regione.campania.it/regione/it/news/primo-piano/covid-19-aggiornamento-4m63?page=2 [accessed 2020-04-30]
16. The Campania region measures to contain COVID-19. Regione Campania. URL: http://regione.campania.it/assets/documents/ord-n-1-24-02-2020.pdf [accessed 2020-04-30]
17. Morey S, Hamoodi A, Jones D, Young T, Thompson C, Dhuny J, et al. Increased diagnosis and treatment of hepatitis C in prison by universal offer of testing and use of telemedicine. J Viral Hepat 2019 Jan;26(1):101-108. [doi: 10.1111/jvh.13017] [Medline: 30315691]
18. Le LB, Rahal HK, Viramontes MR, Meneses KG, Dong TS, Saab S. Patient Satisfaction and Healthcare Utilization Using Telemedicine in Liver Transplant Recipients. Dig Dis Sci 2019 May;64(5):1150-1157. [doi: 10.1007/s10620-018-5397-5] [Medline: 30519848]
19. Serper M, Cubell AW, Deleener ME, Casher TK, Rosenberg DJ, Whitebloom D, et al. Telemedicine in Liver Disease and Beyond: Can the COVID-19 Crisis Lead to Action? Hepatology 2020 Aug 10;72(2):723-728 [FREE Full text] [doi: 10.1002/hep.31276] [Medline: 32275784]
20. Tapper EB, Asrani SK. The COVID-19 pandemic will have a long-lasting impact on the quality of cirrhosis care. J Hepatol 2020 Aug 13;73(2):441-445 [FREE Full text] [doi: 10.1016/j.jhep.2020.04.005] [Medline: 32298769]
21. Wolf I, Waissengrin B, Pelles S. Breaking Bad News via Telemedicine: A New Challenge at Times of an Epidemic. Oncologist 2020 Jun;25(6):e879-e880 [FREE Full text] [doi: 10.1634/theoncologist.2020-0284] [Medline: 32304624]

Abbreviations

CLD: chronic liver disease
HCC: hepatocellular carcinoma

Edited by G Eysenbach; submitted 31.05.20; peer-reviewed by A Berzigotti, R Ohannessian; comments to author 13.07.20; revised version received 01.08.20; accepted 17.08.20; published 21.09.20.

Please cite as:
Guarino M, Cossiga V, Fiorentino A, Pontillo G, Morisco F
Use of Telemedicine for Chronic Liver Disease at a Single Care Center During the COVID-19 Pandemic: Prospective Observational Study
J Med Internet Res 2020;22(9):e20874
URL: http://www.jmir.org/2020/9/e20874/
doi:10.2196/20874
PMID:32896833

Original Paper

Increased Internet Searches for Insomnia as an Indicator of Global Mental Health During the COVID-19 Pandemic: Multinational Longitudinal Study

Yu-Hsuan Lin[1,2,3,4], MD, PhD; Ting-Wei Chiang[1], BSE; Yu-Lun Lin[2], BS

[1]Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan

[2]Institute of Health Behaviors and Community Sciences, College of Public Health, National Taiwan University, Taipei, Taiwan

[3]Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan

[4]Department of Psychiatry, College of Medicine, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Yu-Hsuan Lin, MD, PhD
Institute of Population Health Sciences
National Health Research Institutes
Room A3234, No 35 Keyan Road, Zhunan Town
Miaoli County, 35053
Taiwan
Phone: 1 886 37206166 ext 36383
Email: yuhsuanlin@nhri.edu.tw

Abstract

Background: Real-time global mental health surveillance is urgently needed for tracking the long-term impact of the COVID-19 pandemic.

Objective: This study aimed to use Google Trends data to investigate the impact of the pandemic on global mental health by analyzing three keywords indicative of mental distress: “insomnia,” “depression,” and “suicide.”

Methods: We examined increases in search queries for 19 countries. Significant increases were defined as the actual daily search value (from March 20 to April 19, 2020) being higher than the 95% CIs of the forecast from the 3-month baseline via ARIMA (autoregressive integrated moving average) modeling. We examined the correlation between increases in COVID-19–related deaths and the number of days with significant increases in search volumes for insomnia, depression, and suicide across multiple nations.

Results: The countries with the greatest increases in searches for insomnia were Iran, Spain, the United States, and Italy; these countries exhibited a significant increase in insomnia searches on more than 10 of the 31 days observed. The number of COVID-19–related deaths was positively correlated to the number of days with an increase in searches for insomnia in the 19 countries (ρ=0.64, *P*=.003). By contrast, there was no significant correlation between the number of deaths and increases in searches for depression (ρ=–0.12, *P*=.63) or suicide (ρ=–0.07, *P*=.79).

Conclusions: Our analysis suggests that insomnia could be a part of routine mental health screening during the COVID-19 pandemic.

(J Med Internet Res 2020;22(9):e22181) doi:10.2196/22181

KEYWORDS

internet search; Google Trends; infodemiology; infoveillance; COVID-19; insomnia; mental health

Introduction

The COVID-19 pandemic is the largest global public health challenge of this century. The number of confirmed COVID-19 cases and deaths have been growing exponentially, with over 355,688 deaths and 5,695,115 people infected worldwide, as of May 28, 2020 [1]. In mid-March, countries with the highest prevalence of COVID-19 cases imposed lockdown policies, such as social distancing, bans on nonessential travel, and the temporary closure of almost all businesses, facilities, and places of religious and other gathering, including funerals. There are concerns that the pandemic and the secondary consequences of

the public health response, such as lockdowns and social distancing, may adversely affect mental health. In addition, in the absence of a vaccine or effective treatment, fear of COVID-19 and increased experience of bereavement have left more people vulnerable to mental health problems, such as insomnia, anxiety, depression, posttraumatic stress, and suicide. The World Health Organization (WHO) has recognized the importance of considering population-level psychological well-being and mental health during the COVID-19 pandemic [2].

Surveys on COVID-19–related mental health have been mostly based on self-reported, cross-sectional studies. Most of these surveys investigated a single country; multinational studies have been scarce [3]. Only one study, with a repeated cross-sectional design, has monitored changes in a population's mental health throughout the course of the pandemic [4], but there was no longitudinal follow-up survey. Therefore, current research still does not provide a strong basis for national mental health strategies. However, mental distress can be reflected in Google searches. For example, Google searches for flu symptoms have been found to be real-time indicators of influenza outbreaks [5]. Google Trends has been used previously for population mental health surveillance, and for longitudinal tracking to identify potential risk factors of depression [6] and suicide [7]. In a global crisis like the COVID-19 pandemic, real-time global mental health surveillance is urgently needed for tracking the long-term impact.

This study hypothesized that increases in search terms related to mental health might correspond to the prevalence of COVID-19 in different countries. We aimed to use Google Trends data to investigate the impact of the pandemic on populations' mental health, as indexed by changes in search frequency for three keywords indicative of mental distress: insomnia, depression, and suicide. We delineate a 4-month time course and focus on the period from March 20 to April 19, 2020, to evaluate search behaviors related to insomnia, depression, and suicide in response to a nation's level of COVID-19 transmission.

Methods

Study Design

Using Google Trends data, we obtained search trends including "insomnia," "depression," and "suicide" from December 20, 2019, to April 19, 2020, as a surrogate for the general population's mental health status in 19 countries with differing rates of COVID-19 prevalence. We confirmed the translation of the three terms into the local languages of the 19 countries by using translations from both Chinese and English, with back-translation on Google Translate.

Google Trends does not provide information on the absolute numbers of searches. Instead it provides a relative search value to display search activity for a given term according to a specific period, time, and area. Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. This value is scaled from 0 to 100. A value of 100 is the peak popularity of the term, while a value of 50 means that the term is half as popular in a given time period/area with search volumes for the days given relative to this.

We examined daily searches from March 20 to April 19, using this period since most countries enforced lockdowns in mid-March, to compare observed search volumes with expected search volumes from the 3-month baseline. The 3-month baseline period from December 20, 2019, to March 20, 2020, included two sentinel events: (1) on January 20, Chinese health authorities announced the human-to-human transmission of COVID-19, (ie, the first baseline month prior to any widespread knowledge of the disease worldwide); and (2) on February 20, the number of cases of COVID-19 outside China started to increase rapidly (no countries other than China had accumulated more than 100 cases of COVID-19 prior to February 19). The 3-month baseline time window was selected to inform our prediction since a longer time window could be contaminated by other past relevant events.

Increased Search Volume Estimation

Using search rates from December 20, 2019, to March 19, 2020, for each outcome ("insomnia," "depression," or "suicide"), we forecasted a counterfactual scenario of expected search rates had the COVID-19 rapid outbreak and lockdown policies not occurred in mid-March. The expected relative search volumes were estimated using Hyndman and Khandakar's algorithm for autoregressive integrated moving average (ARIMA) modeling [8], fit to the historical search rates from December 20, 2019, to March 19, 2020, and compared to the observed search rates from March 20 to April 19, 2020. For the residuals from each of the chosen models, we verified that they showed no significant autocorrelation with a Ljung–Box test [9]. Subsequently, we used daily trends from December 20, 2019, to March 19, 2020, to forecast future values with bootstrap CIs, which were computed using R software, version 3.6.3 (R Foundation for Statistical Computing).

Indicators of Increased Searches for Insomnia, Depression, and Suicide

We examined the number of days with increased search volume for insomnia, depression, and suicide. We defined increases in search queries based on both the intensity and duration of increases in population interest, which was applied per our previous study [10]. The intensity of a significant increase was defined as the actual daily Google Trends value during March 20 to April 19 being higher than the expected +1.645 standard error (via ARIMA), that is, the upper limit of the 95% CI in a one-tailed test. We calculated the number of days with significant increases in searches for insomnia higher than the 95% CI expected in each country from March 20 to April 19. We used similar approaches to calculate the indicators for depression and suicide.

Rate of COVID-19 Spread Across Nations

We used the increase in the number of deaths from March 20 to April 19 as an indicator of the speed at which COVID-19 spread within a nation during this time period, since widespread population testing was not yet available in most countries, especially in those with rapid outbreaks of COVID-19. However,

in order to examine the robustness of our findings, we also used three additional indicators: increases in confirmed cases of COVID-19 from March 20 to April 19; the cumulative number of confirmed cases; and the cumulative number of deaths as of March 20.

We used the Spearman rank-order correlation test to examine the correlation between increases in the number of confirmed cases and deaths related to COVID-19 and the number of days with increases in search volume for insomnia, depression, and suicide in the 19 countries from March 20 to April 19. We also examined the temporal correlation between confirmed cases and deaths as of March 20 as well as the number of days with increases in search volumes for insomnia, depression, and suicide from March 20 to April 19.

Results

Table 1 shows the number of days with significant increases in searches for insomnia, depression, and suicide in the local language among the 19 countries. Nations with the highest increases in searches for insomnia were Iran, Spain (Figure 1), the United States, and Italy; these countries exhibited a significant increase in insomnia queries on more than 10 of the 31 days observed (March 20 to April 19). The countries with the greatest increases in searches for depression were Iran (8 days), Australia (6 days), and Hong Kong (6 days). The countries with the greatest increases in searches for suicide were Iran (10 days), Germany (10 days [Figure 2]), and Italy (5 days).

Table 1. Number of days with significant increases in searches for insomnia, depression, and suicide in each country's local language, as well as the cumulative number of confirmed cases as of March 20, 2020; increases in confirmed cases from March 20 to April 19; and the increases in the number of deaths due to COVID-19.

Country	Insomnia		Depression		Suicide		Deaths, n		Confirmed cases, n	
	Translated keyword	Days with significant increases in searches, n	Translated keyword	Days with significant increases in searches, n	Translated keyword	Days with significant increases in searches, n	From March 20 to April 19	As of March 20	From March 20 to April 19	As of March 20
Australia	insomnia	5	depression	6	suicide	0	60	7	5819	791
Brazil	insônia	10	depressão	0	suicídio	0	2451	11	37,861	793
Canada	insomnia	4	depression	4	suicide	1	1552	12	34,676	943
France	insomnie	9	dépression	0	suicide	2	19,244	450	139,196	12,612
Germany	Schlaflosigkeit	4	Depression	0	Selbstmord	10	4519	67	125,336	19,848
Hong Kong	失眠	1	抑鬱	6	自殺	3	0	4	769	256
Iran		17		9		10	3685	1433	62,567	19,644
Italy	insonnia	11	depressione	1	suicidio	5	19,628	4032	131,951	47,021
Japan	眠れない	1	うつ病	0	自殺	0	203	33	9834	963
New Zealand	insomnia	2	depression	1	suicide	1	12	0	1392	39
Russia	бессонница	1	депрессия	0	самоубийство	0	360	1	42,600	253
Singapore	insomnia	3	depression	0	suicide	2	11	0	6203	385
South Korea	불면증	1	우울증	3	자살	2	140	94	2009	8652
Spain	insomnio	16	depresión	4	suicidio	0	19,410	1043	178,264	20,410
Taiwan	失眠	4	憂鬱	2	自殺	4	4	2	285	135
Thailand		6		2		1	46	1	2443	322
Turkey	uykusuzluk	2	depresyon	0	intihar	0	2013	4	85,947	359
United Kingdom	insomnia	4	depression	0	suicide	0	18,298	194	116,084	3983
United States	insomnia	11	depression	4	suicide	1	40,596	349	739,536	19,273

Figure 1. Daily trends for all Google searches for the term "insomnia" alongside expected trends for the days after March 20, 2020, in (A) Spain and (B) Germany.

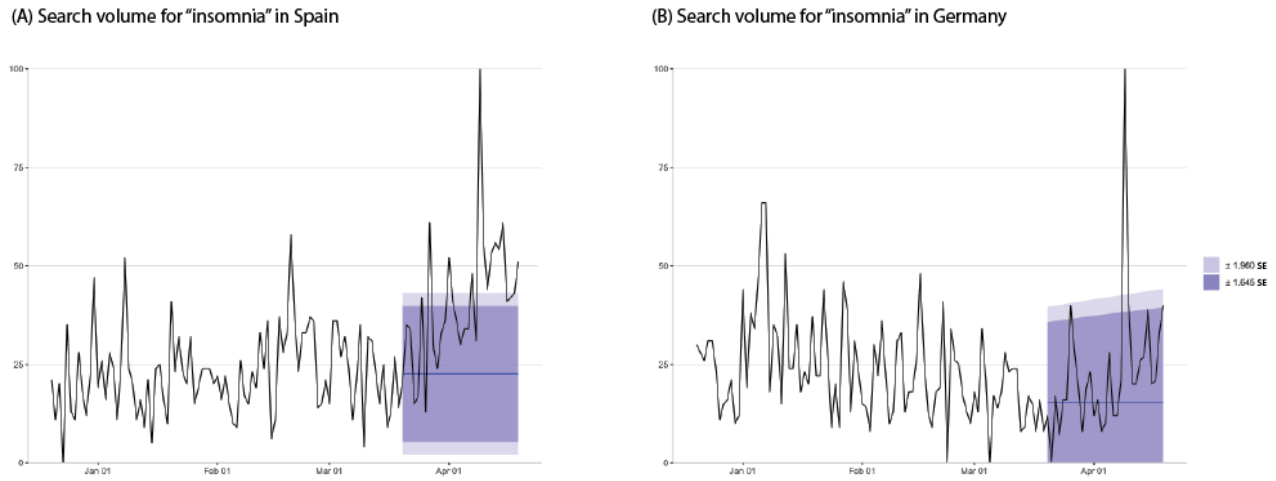


Figure 2. Daily trends for all Google searches with the term "suicide," alongside expected trends for the days after March 20, 2020, in (A) Spain and (B) Germany.

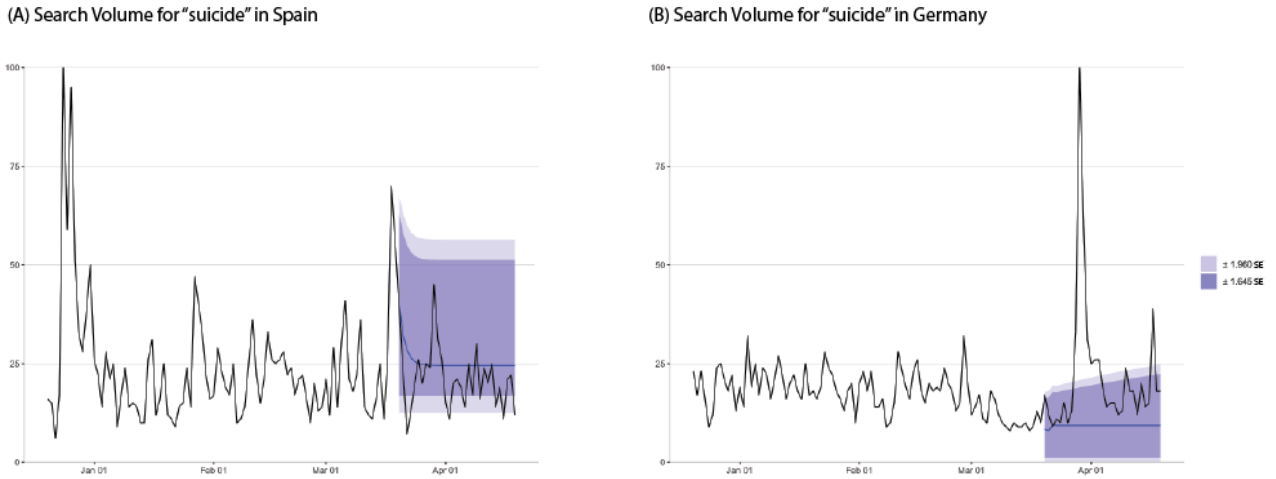
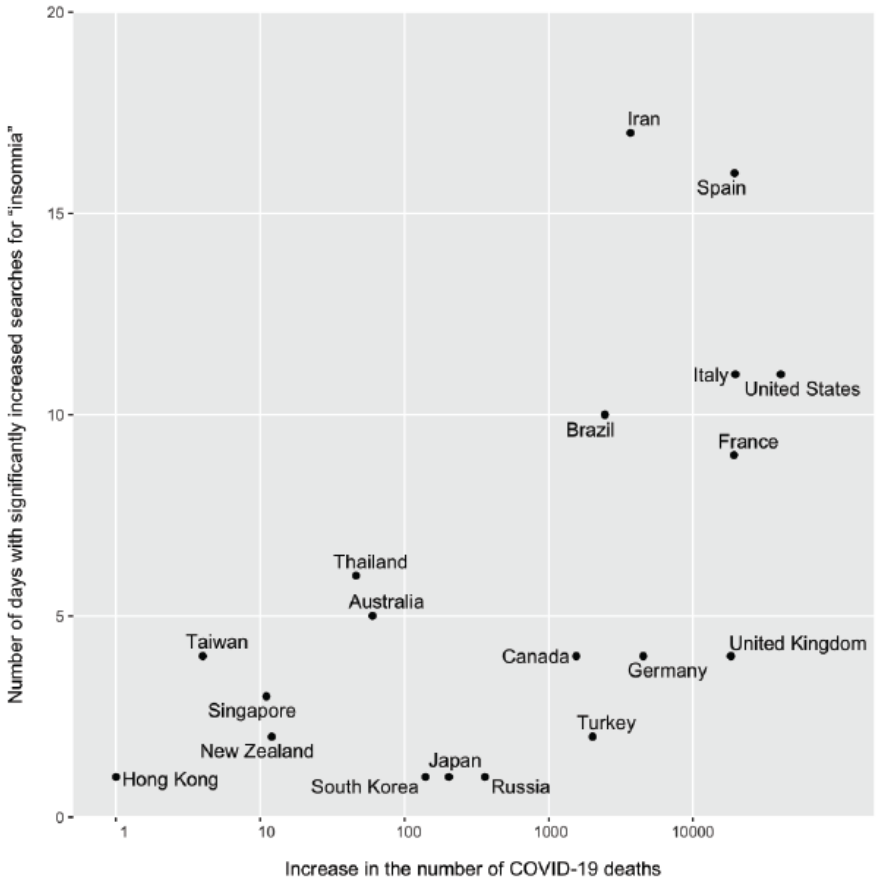


Figure 3 shows that increases in deaths were positively correlated to the number of days with increases in searches for insomnia in the 19 countries from March 20 to April 19 with a Spearman correlation coefficient (ρ) of 0.64 (*P*=.003). Similarly, the number of cumulative death cases as of March 20 was temporally positively correlated to increases in insomnia searches from March 20 to April 19 (ρ=0.60, *P*=.007).

By contrast, there was no significant correlation between the increased searches for depression and increases in deaths (ρ=–0.12, *P*=.63), nor a correlation between the increased searches for suicide and increases in deaths (ρ=–0.07, *P*=.79).

Figure 3. Positive correlation between increases in COVID-19–related deaths and the number of days with an increase in searches for insomnia in 19 countries (Spearman correlation coefficient=0.64, *P*=.003) from March 20 to April 19, 2020.



Discussion

Principal Findings

To the best of our knowledge, this is the first study to use a real-time collection of population-level data to investigate the impacts of the COVID-19 pandemic on mental health across 19 countries. The multinational and longitudinal design demonstrates with a high temporal resolution how the pandemic influences global mental health. Furthermore, this study compared the associations between the impact of the COVID-19 outbreak and its varied impact on mental health indicators, namely insomnia, depression, and suicide. The increase in Google searches for insomnia, rather than for depression and suicide, was significantly correlated to the number of deaths related to COVID-19. These findings indicate the extent of the pandemic's impact on the general population's mental health.

Insomnia is a common, preceding symptom or precipitating factor in new onset mental illness, such as depression, anxiety, and posttraumatic stress disorder [11], whereas suicide is the most severe outcome of psychiatric disorders over a lifetime [12]. Stress‐related sleep problems are common [13], and poor sleep quality has been identified as having mental health consequences as a result of social isolation [14]. In the absence of a vaccine and effective treatment for COVID-19, one of the most vital strategies for slowing the pandemic is social distancing. However, even for households free of the virus, the pandemic is likely to function as a major stressor, especially in terms of economic difficulties. Such effects may be exacerbated by self-isolation policies that can increase social isolation and relationship difficulties. Loneliness and social isolation may worsen the burden of stress and often exacerbate insomnia. Future research should explore search behaviors associated with terms such as fear, anxiety, and stress, which are highly relevant to COVID-19 and insomnia, to strengthen the qualitative correlation between insomnia and the COVID-19 pandemic.

Our results indicate that insomnia is a more sensitive indicator than depression or suicide of populations' mental health during the COVID-19 pandemic. However, unlike the patterns of Google searches for "face mask" and "wash hands," which reached all-time highs during this pandemic [10], searches patterns for insomnia, depression, and suicide usually consisted of multiple fluctuating components (eg, seasonality) and may increase in imperceptible manners. ARIMA models are capable of modeling both seasonal and nonseasonal data and can perform time-series forecasting to quantify whether imperceptible search queries are higher than expected [15,16].

The increased searches for insomnia and suicide demonstrated different patterns in this study (Figures 1 and 2). The surge in suicide-related searches in some countries (Italy, Germany, and Iran) may be attributed to media reports of suicides (ie, an Italian nurse on March 24, a German minister on March 29, and an Iranian student on April 6). These findings are consistent with previous studies that report on the correlation between suicide search trends and actual suicides [17], indicating that media

reports of suicides can lead to spikes in suicides [18]. On the basis of these findings, media professionals should follow the WHO's media guidelines [2] for preventing suicide.

Limitations

There are several methodological limitations that should be noted when interpreting this study's findings. First, we recognize that the Google Trends data do not represent a random sampling of the population and may exclude important vulnerable groups without access to the internet or those who were not actively engaged in searching. Second, we were unable to determine the sociodemographic characteristics of those conducting the searches. COVID-19 disproportionately affects poor and vulnerable populations; additionally, patients with serious mental illness may be among the hardest hit [19]. Our results did not include the entire population, or all internet users, of every country. Third, individual search queries for insomnia, depression, or suicide may not accurately reflect the actual mental health status of internet users. Factors other than the COVID-19 pandemic, including cultural differences that affect the expression and evaluation of symptoms, as well as more complicated bio-psycho-social factors may influence mental health and thus internet search behaviors. However, the collective phenomenon of internet search behavior still can be a meaningful surrogate for mental health across large populations.

Conclusion

In conclusion, our analysis suggests that insomnia could be a part of routine screening for mental health during the COVID-19 pandemic. Monitoring Google Trends has the benefit of allowing for rapid longitudinal tracking at the international level during this unprecedented health crisis.

Acknowledgments

The authors thank Dr Sheng-Hsuan Lin, Ms Hsiao-Han Lin, and Ms Szu-Yu Chen for their excellent technical assistance.

This study was supported by grants from the National Health Research Institutes of Taiwan (09A1-PHPP20-014).

Authors' Contributions

Y-HL designed the study. Y-HL and T-WC conducted the study and analyzed the data. Y-HL and Y-LL drafted the manuscript. All authors contributed to data analysis, and drafting or revising the paper; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. COVID-19 Dashboard. The Center for Systems Science and Engineering JHU. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-05-18]
2. Mental health and psychosocial considerations during the COVID-19 outbreak. World Health Organization. 2020 Mar 18. URL: https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf [accessed 2020-05-18]
3. Chew NW, Lee GK, Tan BY, Jing M, Goh Y, Ngiam NJ, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
4. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
5. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
6. Yang AC, Huang NE, Peng C, Tsai S. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. PLoS One 2010 Oct 28;5(10):e13728 [FREE Full text] [doi: 10.1371/journal.pone.0013728] [Medline: 21060851]
7. Solano P, Ustulin M, Pizzorno E, Vichi M, Pompili M, Serafini G, et al. A Google-based approach for monitoring suicide risk. Psychiatry Res 2016 Dec 30;246:581-586. [doi: 10.1016/j.psychres.2016.10.030] [Medline: 27837725]
8. Hyndman R, Khandakar Y. Automatic Time Series Forecasting: The Forecast Package for R. J Stat Soft 2008;27(3). [doi: 10.18637/jss.v027.i03]
9. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika 1978 Aug 01;65(2):297-303. [doi: 10.1093/biomet/65.2.297]
10. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]

11. Pigeon WR, Bishop TM, Krueger KM. Insomnia as a Precipitating Factor in New Onset Mental Illness: a Systematic Review of Recent Findings. Curr Psychiatry Rep 2017 Aug 14;19(8):44. [doi: 10.1007/s11920-017-0802-x] [Medline: 28616860]
12. Fazel S, Runeson B. Suicide. N Engl J Med 2020 Jan 16;382(3):266-274. [doi: 10.1056/NEJMra1902944] [Medline: 31940700]
13. Altena E, Micoulaud-Franchi J, Geoffroy P, Sanz-Arigita E, Bioulac S, Philip P. The bidirectional relation between emotional reactivity and sleep: From disruption to recovery. Behav Neurosci 2016 Jun;130(3):336-350. [doi: 10.1037/bne0000128] [Medline: 26866361]
14. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct 22;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
15. Ayers JW, Althouse BM, Leas EC, Dredze M, Allem J. Internet Searches for Suicide Following the Release of 13 Reasons Why. JAMA Intern Med 2017 Oct 01;177(10):1527-1529 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3333] [Medline: 28759671]
16. Liu M, Caputi TL, Dredze M, Kesselheim AS, Ayers JW. Internet Searches for Unproven COVID-19 Therapies in the United States. JAMA Intern Med 2020 Aug 01;180(8):1116-1118 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1764] [Medline: 32347895]
17. Yang AC, Tsai S, Huang NE, Peng C. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. J Affect Disord 2011 Jul;132(1-2):179-184. [doi: 10.1016/j.jad.2011.01.019] [Medline: 21371755]
18. Niederkrotenthaler T, Braun M, Pirkis J, Till B, Stack S, Sinyor M, et al. Association between suicide reporting in the media and suicide: systematic review and meta-analysis. BMJ 2020 Mar 18;368:m575. [doi: 10.1136/bmj.m575] [Medline: 32188637]
19. Druss BG. Addressing the COVID-19 Pandemic in Populations With Serious Mental Illness. JAMA Psychiatry 2020 Sep 01;77(9):891-892. [doi: 10.1001/jamapsychiatry.2020.0894] [Medline: 32242888]

Abbreviations

ARIMA: autoregressive integrated moving average
WHO: World Health Organization

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by LA Lee, C Ulmer; comments to author 17.08.20; revised version received 19.08.20; accepted 13.09.20; published 21.09.20.

Please cite as:
Lin YH, Chiang TW, Lin YL
Increased Internet Searches for Insomnia as an Indicator of Global Mental Health During the COVID-19 Pandemic: Multinational Longitudinal Study
J Med Internet Res 2020;22(9):e22181
URL: http://www.jmir.org/2020/9/e22181/
doi:10.2196/22181
PMID:32924951

Original Paper

Dynamic Panel Estimate–Based Health Surveillance of SARS-CoV-2 Infection Rates to Inform Public Health Policy: Model Development and Validation

James Francis Oehmke[1], BSc, MSc, PhD; Theresa B Oehmke[2], BSc, MSc; Lauren Nadya Singh[1], BSc, MPH; Lori Ann Post[1], BA, MA, PhD

[1]Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Civil Engineering, University of California at Berkeley, Berkeley, CA, United States

Corresponding Author:
Lori Ann Post, BA, MA, PhD
Department of Emergency Medicine
Feinberg School of Medicine
Northwestern University
420 E. Superior St
Chicago, IL, 60611
United States
Phone: 1 203 980 7107
Email: lori.post@northwestern.edu

Abstract

Background: SARS-CoV-2, the novel coronavirus that causes COVID-19, is a global pandemic with higher mortality and morbidity than any other virus in the last 100 years. Without public health surveillance, policy makers cannot know where and how the disease is accelerating, decelerating, and shifting. Unfortunately, existing models of COVID-19 contagion rely on parameters such as the basic reproduction number and use static statistical methods that do not capture all the relevant dynamics needed for surveillance. Existing surveillance methods use data that are subject to significant measurement error and other contaminants.

Objective: The aim of this study is to provide a proof of concept of the creation of surveillance metrics that correct for measurement error and data contamination to determine when it is safe to ease pandemic restrictions. We applied state-of-the-art statistical modeling to existing internet data to derive the best available estimates of the state-level dynamics of COVID-19 infection in the United States.

Methods: Dynamic panel data (DPD) models were estimated with the Arellano-Bond estimator using the generalized method of moments. This statistical technique enables control of various deficiencies in a data set. The validity of the model and statistical technique was tested.

Results: A Wald chi-square test of the explanatory power of the statistical approach indicated that it is valid (χ^2_{10}=1489.84, P<.001), and a Sargan chi-square test indicated that the model identification is valid (χ^2_{946}=935.52, P=.59). The 7-day persistence rate for the week of June 27 to July 3 was 0.5188 (P<.001), meaning that every 10,000 new cases in the prior week were associated with 5188 cases 7 days later. For the week of July 4 to 10, the 7-day persistence rate increased by 0.2691 (P=.003), indicating that every 10,000 new cases in the prior week were associated with 7879 new cases 7 days later. Applied to the reported number of cases, these results indicate an increase of almost 100 additional new cases per day per state for the week of July 4-10. This signifies an increase in the reproduction parameter in the contagion models and corroborates the hypothesis that economic reopening without applying best public health practices is associated with a resurgence of the pandemic.

Conclusions: DPD models successfully correct for measurement error and data contamination and are useful to derive surveillance metrics. The opening of America involves two certainties: the country will be COVID-19–free only when there is an effective vaccine, and the “social” end of the pandemic will occur before the “medical” end. Therefore, improved surveillance metrics are needed to inform leaders of how to open sections of the United States more safely. DPD models can inform this reopening in combination with the extraction of COVID-19 data from existing websites.

(J Med Internet Res 2020;22(9):e20924) doi:10.2196/20924

KEYWORDS

COVID-19; models; surveillance; COVID-19 surveillance system; dynamic panel data; infectious disease modeling; reopening America; COVID-19 guidelines; COVID-19 health policy

Introduction

Background

The SARS-CoV-2 pandemic is unprecedented [1,2], with high mortality and morbidity of the virus due to its rapid spread worldwide [3,4]. Without an effective vaccine [5-7], countries are at risk for continued spread [8]. Without good health surveillance, public health leaders are unaware of where and how the disease is spreading. Effective surveillance can inform the safe reopening of economies [9-22] by geographical region [23]. To that end, we submit this proof of concept of the creation of surveillance metrics that correct for measurement error and data contamination. This study applies state-of-the-art statistical modeling to existing data mined from the internet to derive the best available estimates of the state-level dynamics of COVID-19 infection to determine if the sustained decline in SARS-CoV-2 infection that is necessary to reopen is occurring or, conversely, if reopening without applying best public health practices is resulting in a resurgence of SARS-CoV-2.

Public health surveillance is defined as the "ongoing systematic collection, analyses, and interpretation of outcome-specific data for use in the planning, implementation and evaluation of public health practice [18]." Unfortunately, existing surveillance methods suffer from undercounts, bias, and error, and they mostly include more severe cases [24-32]. Research has confirmed that best practices for containment of the COVID-19 pandemic include closing borders between countries [33,34], extreme quarantine measures [35-37], social isolation at home [38], social distancing [39], hand hygiene [40-42], crowd control [43], and wearing a mask in public [44,45]; however, health surveillance must inform where and when to employ these best practices. Due to delays in reporting of new cases, deaths, and testing [46-48], these decisions are made based on partial evidence. Existing models of COVID-19 contagion rely on parameters such as the basic reproduction number (R_0), which are difficult to measure in real time, and they use static statistical methods that do not capture all of the relevant dynamics [49], such as varying specificity and sensitivity of diagnostic testing or asymptomatic individuals who are never tested and are unwittingly carrying SARS-CoV-2 [25,50]. The epidemiological definition of R_0 is the average number of people who contract a disease from a contagious person. It applies specifically to a population of people who were previously free of infection and were not vaccinated [51]. Existing surveillance systems use data that are subject to significant measurement error and other contaminants [52,53]. Moreover, timely information is needed to improve statistical methods that extract information from data sets posted on websites [54-56].

The conventional approach to modeling the spread of diseases such as COVID-19 is to posit an underlying contagion model [57] and then to seek accurate direct measurement of the model parameters, such as reproduction rates or other parameters; these measurements are sometimes inferred through deaths, hospitalizations, and caseloads [58], and they often involve labor-intensive methods that rely on contact tracing to determine the spread of the disease among a sample population [54,59-61]. For viral epidemics with an incubation period of up to 14 days [62], weeks if not months are required to generate accurate parameter estimates, even for simple contagion models. For example, early estimates of COVID-19 were estimated using methods developed by Lipsitch [63] applied to data from contact tracing in Wuhan and Italy; however, the statistical properties were weak [64-70]. For example, Zhao [65] estimated the serial interval distribution and R_0 based on only six pairs of cases [71]. These models also rely on underlying assumptions about immunity, common propensity for infection, and well-mixed populations, among others. Improvements in these models typically focus on relaxing these assumptions, such as disaggregating the population by geography and modeling within-geography and cross-geography personal interactions [3]. Martcheva [76] provides an excellent dynamic analysis of a wide variety of contagion models and their possible dynamics [72-77]. Unfortunately, they provide limited options for the statistical inference of parameter values from actual data [76]. The objective of this study is to derive surveillance metrics using methods that control for data limitations and contamination.

Methods

Model Development

In contrast to previous studies, we used an empirical approach that focuses on statistical modelling of widely available empirical data, such as the number of confirmed cases or the number of tests, which can inform estimates of the current values of critical parameters such as the infection rate or reproduction rate. We explicitly recognized that the data generating process for the reported data contains an underlying contagion component; a politico-economic component, such as availability of accurate test kits; a social component, such as how strongly people adhere to social distancing measures, mask requirements, and shelter-in-place policies; and a sometimes inaccurate data reporting process that may obscure the underlying contagion process. Therefore, we sought to develop a statistical approach that can provide meaningful information despite the complex and sometimes obfuscating data generation process. Our approach is consistent with the principles of evidence-based medicine, including controlling for complex pathways that may include socioeconomic factors such as mediating variables and policy recommendations, and "based on the best available knowledge, derived from diverse sources and methods [5]."

There are two primary advantages to this empirical approach. First, we can apply the empirical model relatively quickly to a

short data set. This advantage stems from the panel nature of the model. We used US states as the cross-sectional variable; therefore, one week of data from 52 states and territories (including Puerto Rico and the District of Columbia) provides a reasonable sample size. In addition to enabling parameter estimation early in a pandemic, using this property, we tested to see if a shift had occurred in the infection or reproduction rates of the contagion process in the past week (ie, whether there is statistical evidence that reopening is associated with an acceleration in the number of cases).

The second advantage of our approach is that it directly measures and informs policy-relevant variables. For example, the White House issued guidance on reopening the US economy that depends on a decrease in the documented number of cases and in the proportion of positive test results over a 14-day period, among other criteria and considerations [23,78-83]. As noted above, the number and proportion of positive test results are the outcomes of a data generating process that includes not only the underlying contagion process but a multitude of mediating factors as well as idiosyncrasies of the data collection and a delayed reporting process. We specifically modeled the number of positive test results in our empirical model, which provides evidence of direct use in policy dialogue.

Herein, we proceed with a brief discussion of the contagion models that informed our selection of an empirical model. We describe the basic dynamic panel data (DPD) approach and its advantages for analyzing the current pandemic. We obtained results that validate the model specification, which is a necessary and important step in the development of a surveillance system [9-11,14,15,18,20]. We then used the validated model to interrogate our research question: is reopening associated with increased infection transmission and a re-emergence of the pandemic? We approached this research question by statistically testing whether R-type contagion parameters and, specifically, the daily and weekly persistence increased during the weeks of June 27-July 3 and July 4-10, 2020.

Representing Contagion as a DPD Model

Transmission models are typically population-based differential equations of the form $dY/dt = f(Y,X)$, where Y is a vector of a population or subpopulation characteristic of interest, such as the number of exposed or infected individuals; X is a vector of mediating factors (often omitted); and f is a transition function. For empirical purposes, we will use difference equations because the data come in discrete time periods, specifically days. For example, the sizes of the susceptible, infected, and recovered populations in the susceptible-infected-recovered (SIR) model in difference equation form are:

$$\blacksquare(S\_it- S\_(it-1) = - (\beta S\_(it-1) I\_(it-1))/N\_(it-1) @I\_it- I\_(it-1) = (\beta S\_(it-1) I\_(it-1))/N\_(it-1) - (\gamma)\_R+\gamma\_D)I\_(it-1)@\blacksquare(R\_it- R\_(it-1)= \gamma\_R I\_(it-1) @D\_it- D\_(it-1)= \gamma\_D I\_(it-1))) \quad (1)$$

where S, I, and R are the sizes of the susceptible, infected, and recovered populations, respectively; D is the number of deaths due to SARS-CoV-2; N is the size of the total population ($S + I + R + D$); and the subscripts denote the time period. The first line represents the change in the susceptible population, which decreases when a susceptible individual becomes infected. This occurs when the susceptible individual interacts with another individual who is infected, in which case the virus is transmitted to the susceptible individual with probability I/N. The number of infected individuals increases by the number of newly infected individuals and decreases by the number of previously infected individuals who either recovered or died. The γ parameters are the probability of recovering or dying. β and the γ are the unknown parameters of the model. Calibration of contagion models requires estimation of the true parameter values.

The availability of state-level data suggests that Equation 1 can be rewritten in panel regression form as

$$\blacksquare(S\_it = \gamma\_Si+S\_(it-1)- (\beta S\_(it-1) I\_(it-1))/N\_(it-1) +\varepsilon\_Sit @I\_it = \gamma\_Ii+((1+ \beta S\_(it-1))/N\_(it-1) - (\gamma)\_R+\gamma\_D))I\_(it-1)+\varepsilon\_Iit@\blacksquare(R\_it = \gamma\_Ri+R)\_(it-1)+\gamma\_R I\_(it-1) +\varepsilon\_Rit @D\_it = \gamma\_Di+D)\_(it-1)+\gamma\_D I\_(it-1) +\varepsilon\_Dit)) \quad (2)$$

The additional index i refers to the state; therefore, I_{it} represents the number of infected people in state i at time t. Consistent with the panel data specifications, we added a state-specific "fixed effect" to each of the equations, γ_i, which represents time-invariant state characteristics such as population rate. The $\varepsilon_{.it}$ represent error terms.

We apply the dynamic panel data approach to the number of positive test results per day as reported on internet sites. To avoid imposing too much specificity, we allowed for some flexibility in the functional form by including the number of tests both linearly and quadratically and as a proportion of the population:

where P_{it} is the number of new positive test results and T_{it} is the number of tests administered in state i on day t; $I_{6.27}$ and $I_{7.04}$ are indicator variables for the time periods from June 27-July 3 and July 4-10, 2020, respectively (latest available data at the time of analysis); and Pop_i is the population of state i (assumed to be constant during the sample). Equation 3 is readily interpretable. The terms containing a β parameter represent the dynamic component of the model. The first term on the right side represents a day-to-day persistence effect (ie, every new case the previous day is a risk factor that contributes β_1 new cases to the current day's caseload). The next two terms allow for shifts in this risk factor (additions or subtractions) for the weeks beginning June 27 and July 4. Analogously, the next three terms represent a 7-day persistence effect and shifts in that effect for the weeks beginning June 27 and July 4. The 7-day persistence effect is the approximate modal time between viral contraction and the appearance of symptoms; therefore, it is related to the reproduction rate (R parameter) in structural contagion models. The final five terms of Equation 3 contain all the contemporaneous effects in the model (the nonhomogeneous component of the difference equation), as in, all the time subscripts occur contemporaneously at time t except for the state fixed effects, which by definition do not change over time. The first of these terms represents state-specific

effects, which are an important control variable in the panel models. The next two terms are linear and quadratic terms of the number of tests administered, while the third term is the number of tests per person. The next three terms represent the effects of the number of tests administered. The fourth term allows for a shift or discontinuity in the level of new infections for the week of July 4-10 because of increasing concern that the pandemic has re-emerged, particularly in the previous 7 days. We would associate a positive shift with an underlying increase in infection rates. The final term is an error term that represents all types of measurement errors.

Data Sources

Case and test data, including the total number of tests administered and the number of positive results, were taken from the COVID Tracking Project [84], which compiles data from multiple sources. Data were accessed from GitHub [85] after 6 PM on July 10, 2020, so that the data would be complete for that day. Population estimates were derived from the 2019 annual state estimates from the US Census Bureau [86].

Estimation

There are three problems with the specification of Equation 3 for estimation purposes. First, the inclusion of lagged dependent variables on the right side means that the errors are autocorrelated and that the usual exogeneity restrictions are violated; therefore, least squares estimates are inappropriate. Second, some variables are omitted, such as all the variables represented in extensions of the SIR model, and other variables that represent socioeconomic factors influencing the contagion, testing, and reporting processes may also have been omitted. Third, our data set has a relatively short time duration, and the asymptotic properties of fixed-effects or random-effects panel data estimators such as statistical efficiency or normality apply as $t \rightarrow \infty$. Use of these estimators with small values of t creates a small-sample problem with unknown or undesirable estimator properties. We applied the Arellano-Bond approach [87,88], which has improved properties for small samples and is appropriate for application to data sets with a small t and large i.

Fortunately, DPD methods can be used to specifically resolve these statistical problems [89-95]. DPD models allow direct estimation of difference equations with panel data, which resolves multiple problems that appear in the COVID-19 data [96]. The technique we used was developed by Arellano and Bond [87], who applied a generalized method of moments (GMM) approach to a dynamic formulation of employment equations, such as the influence of employment levels in a previous period on employment levels in the current period [97-99]. The basic concept translates to the COVID-19 pandemic in the sense that the number of infections in the current period is a function of lagged infection numbers and other variables. In addition, the DPD removes the individual state effects by first differencing the model. Regressions that include a lagged value of the dependent variable violate the exogeneity restrictions for ordinary least squares and panel estimators such as fixed or random effect models because the lagged dependent variable will be correlated with the error term. DPD model estimation is an application of Hansen's GMM approach to difference equations estimated from panel data [97,100-102]. The GMM approach solves the endogeneity problem [103,104]. Rather than minimizing a loss function such as the sum of squared errors or maximizing a distribution-specific likelihood function, the GMM approach focuses on the identification of restrictions, including exogeneity restrictions. In an estimable model, there are more identifying restrictions than parameters, and the GMM selects the parameter values that come closest to satisfying the overidentifying restrictions [105]. In our application, we used 10 explanatory variables as defined in Equation 3 and 940 overidentifying restrictions (ie, the same order of magnitude as the sample size n=1040); therefore, the degrees of freedom were more than sufficient for statistical inference. The GMM procedure requires a set of instrumental variables; in the case of DPDs, the instruments include lags and/or lag differences in the Y variables. These instruments help resolve the endogeneity problem as well as the omitted-variables problem. In addition to addressing the theoretical concerns inherent in the estimation of any difference equation model, the DPD approach addresses multiple statistical issues that are likely to occur in COVID-19 data.

First, the GMM approach is asymptotically efficient; however, it also has good small sample properties, including samples with a large cross-section and a small number of time periods [102]. This is especially important for statistical analysis early in pandemics, when data are not available for a long period of time, as well as for our testing of whether changes in the transmission rate (that may have occurred 1 to 2 weeks ago) have affected the number of positive test results in the past week.

Second, this approach is robust to omitted variables because of its reliance on identifying restrictions and instrumental variables. This is important because we estimate a relatively sparse model that does not include direct controls for mediating factors, data collection issues, or reporting idiosyncrasies.

Third, the approach includes statistical testing of the overidentifying restrictions (ie, whether the empirical model and estimation technique are statistically valid). For this test, we used the Sargan chi-square test.

Fourth, this approach corrects for autocorrelation.

A significant drawback to DPD methods is that they are computationally complex and become very time- and resource-intensive as the number of observations grows.

We used the Arellano-Bond estimation technique developed specifically for DPD applications. We implemented the Arellano-Bond technique using the *xtabond* command in Stata 16.1 (StataCorp LLC).

Model Validation

To validate the significance of the regression, we used a Wald chi-square statistic to test the null hypothesis that the independent variables did not explain the dependent variable (standard goodness-of-fit measures such as R^2 are uninformative in models with a lagged dependent variable). To test the appropriateness of the model, we applied the Sargan chi-square test. This is a test of the null hypothesis that the (over)identifying

restrictions of the model are statistically met; heuristically, this null hypothesis means that the model and estimation procedure are valid. We used $\alpha \leq 5\%$ for tests of statistical significance.

Model Parameters

We report the point estimates and the *P* values for all model parameters in Equation 3 as well as additional statistical test results and *P* values for combinations of parameters when of interest. Of interest are the null hypotheses: $\beta_2 = 0$, $\beta_3 = 0$, $\beta_5 = 0$, and $\beta_6 = 0$. These hypotheses jointly represent the hypothesis that there has been no change in the persistence of the pandemic (ie, the number of new COVID-19 cases over the past two weeks has remained relatively constant). We interpreted rejection of one or more of the hypotheses as evidence that the pandemic is evolving differently, with positive parameter values associated with greater persistence and a re-emergence of the pandemic.

Surveillance Reporting

We translated the estimation results into a surveillance reporting context. The dynamic component (Equation 3) is presented in terms of the persistence rate per 100,000 cases, defined as the number of new COVID-19 cases in every 100,000 cases that remained constant, and this component was applied to the reported infection numbers to determine its effect on the number of cases per state per day. The contemporaneous component was applied to the reported infection numbers to determine its effect on the number of cases per state per day. The two effects were added to obtain a modeled total number of cases per state per day, and this number was multiplied by 52 to obtain a national figure (including the District of Columbia and Puerto Rico but excluding other territories).

Results

Data

The internet data mining effort resulted in a panel (longitudinal data set) with 52 "panels" (50 states, the District of Columbia, and Puerto Rico) using observations from June 13 through July 10, 2020. Before the analysis, outlying and negative values were crosschecked with other reputable COVID-19 data tracking websites, including USA Facts [106] and the Johns Hopkins Coronavirus Resource Center [107]. The data set has $m = 52 \times 28 = 1456$ observations. Because the model requires 8 days of observations to account for various lags and differencing, the model estimation uses $n = 52 \times 20 = 1040$ observations.

Estimation Results

We present the estimation results in Table 1.

Table 1. Arellano-Bond dynamic panel data modeling of the number of daily infections by state from March 20 to July 10, 2020.

Estimation	Coefficient	*P* value
Variables		
Lagged daily positive cases	0.0630	.31
Lagged daily positive shift, June 27-July 03	0.0977	.14
Lagged daily positive shift, July 04-10	–0.1727	.009
Seven-day lagged daily positive cases	0.5188	<.001
Seven-day lagged daily positive shift, June 27-July 03	0.0118	.90
Seven-day lagged daily positive shift, July 04-10	0.2691	.002
Constant	17.7791	.68
Daily tests	0.0520	<.001
Daily tests squared	-1.54×10^{-7}	.002
Daily tests / population	–86,527	<.001
Fitness measurements		
Wald test of regression significance (χ^2_{10})	1489.84	<.001
Sargan test of overidentifying restrictions (χ^2_{946})	935.52	.59
Test of lagged daily positive cases + shift July 04-10 = 0 (χ^2_1)	–9.92	.002

Model Validation

To examine the model fit, we applied a Wald chi-square test of the null hypothesis that there is no explanatory power in the explanatory variables. The model was statistically significant (χ^2_{10}=1489.84, *P*<.001). The Sargan chi-square test failed to reject the null hypothesis of valid overidentifying restrictions (χ^2_{946}=935.52, *P*=.593).

Model Parameter Estimates

The coefficient on the lagged dependent variable of the number of daily cases that tested positive on the previous day was positive and statistically significant (0.0630, *P*<.001). The shift values for this parameter for the weeks beginning June 27 and July 4, 2020, are 0.0977 (*P*=.138) and –0.1727 (*P*=.009), respectively. The effective parameter value for the week of July 4 is 0.0630 – 0.1727 = –0.1097 (*P*=.002).

The coefficient on the 7-day lagged dependent variable, the number of daily cases that tested positive 7 days earlier, was positive and statistically significant (0.5188, $P<.001$). The shift values for this parameter for the weeks beginning June 27 and July 4, 2020, are 0.0118 ($P=.897$) and 0.2691 ($P=.002$), respectively. The effective parameter value for the week of July 4 is 0.5188 + 0.2691 = 0.7879 ($P<.001$).

The coefficient on the linear term in the number of daily tests administered was positive and statistically significant (0.0520, $P<.001$), and the coefficient on the quadratic term was negative and statistically significant (–1.54e-07, $P=.002$). The coefficient on the number of daily tests per person was negative and statistically significant (–86,527, *P<.001*).

Surveillance Results

Table 2 translates the statistical results into a user-friendly, intuitive surveillance reporting template. The first two rows are the reported number of cases and tests, respectively. The third row is the estimated 1-day persistence rate, as in, the number of cases estimated on the current day for every 10,000 cases the previous day. The fourth row is the 7-day persistence rate (ie, the estimated number of cases on the current day for every 10,000 cases 7 days prior). The fifth row is the estimated dynamic component of the model in terms of the number of cases per state per day. This was determined by applying the persistence rates from rows 3 and 4 to the average reported number of cases and adding the effects. The sixth row is the estimated contemporaneous component of the model in terms of number of cases per state per day. The seventh row sums the dynamic and contemporaneous effects to obtain the total estimated effect, as in, the estimated number of new positive test results per state per day. The first column contains the described information as state averages for the period of June 27 to July 03, 2020. The second column contains the information for the United States in aggregate. The third and fourth columns show the same data as the first two columns but for the period of July 4 to 10, 2020.

Table 2. Dynamic panel data estimation results for the United States from June 27 to July 10, 2020.

Variable	June 27-July 3		July 4-10	
	State average	National average	State average	National average
Daily average number of cases for the week	909	47,278	1048	54,491
Daily average number of tests for the week	12,281	638,619	12,630	656,741
Estimated daily persistence rate (per 10,000 cases)	1607	1607	–1106	–1106
Estimated 7-day persistence rate (per 10,000 cases)	5306	5306	7816	7816
Estimated dynamic component (number of cases per day)	499	25,968	595	30,923
Estimated contemporaneous component (number of cases per day)	466	24,254	490	25,464
Total number of estimated cases per day	966	50,221	1084	56,387

Discussion

Principal Findings

Our primary findings are that the 7-day persistence rate is statistically significant and important in magnitude and that the 7-day persistence rate increased by almost 50% from the week of June 27-July 3 to the week of July 4-10 (Table 1). The increase in the 7-day persistence translates into an increase from 5306 new cases per 10,000 cases 7 days prior to 7816 new cases per 10,000 cases (Table 2). On average, this resulted in 95 new cases per state per day. Coupled with a modest increase in the contemporaneous component, the combined result is an estimated increase of 118 new cases per state per day or 6166 new cases nationally per day. The increase in the number of new cases per day is indicative of a shift in the underlying contagion transmission and corroborative of the statement that reopening the US economy has increased the contagion reproduction rate.

The coefficients on the daily lagged dependent variable are small in magnitude and do not indicate strong day-to-day persistence. The negative estimated daily persistence rate for the week of July 4 is indicative of a daily “snaggle-tooth” pattern in the number of daily cases at the state level. This simply indicates that a low number of cases on one day is offset by a high number of cases the next day, probably due to reporting delays and differential testing periods; this pattern appears slightly in the US aggregate data and is strongly evident in the California data. Other states exhibited different snaggle-tooth patterns, including high-incidence states such as Florida, Texas, and Georgia.

The contemporaneous component of the model contributed positively to the number of new daily cases but did not change significantly over the sample period.

Limitations

While DPD is useful in deriving dynamic estimates of the rate of transmission of COVID-19, static numbers using traditional surveillance tools must also be included to obtain a complete understanding of the pandemic.

Conclusions

The DPD model is a statistically validated analysis of reported COVID-19 data and an important addition to the epidemiological toolkit for understanding the progression of the pandemic. It is important to recognize that this is a supplementary tool that does not replace detailed contagion modeling with detailed and specific data for accurate

representation of contagion model parameters. However, there are four salient advantages of the DPD approach. First, this approach enables statistically efficient extraction of information from existing data sets, including statistical validation of results; therefore, it is applicable to the most commonly tracked and reported data in the current pandemic. Second, the tool could be applied relatively quickly after the pandemic started because of its ability to model reported data rather than detailed contract tracing data, which is largely unavailable to date. That is, changes in the evolution of the pandemic can be confirmed much more quickly using panel data than using aggregate data. Third, this approach informs real-time policy decisions, including decisions based on commonly reported data, such as reopening state economies. Fourth, the model results can help inform the parameterization of more traditional contagion models.

This model is consistent in that it shows a higher reproduction rate during the most recent 7 days; this confirms that in general, normal operation should not be resumed in the United States. Rather, empirically validated public health guidelines such as wearing masks, social distancing, social isolation, hand washing, and avoidance of social gatherings should be immediately adopted to reduce the contagion. In fact, White House guidelines recommend 14 sustained days of reduced COVID-19–related deaths, new infection cases, and proportions of positive test results prior to reopening. That threshold has not been met. While these findings reflect the national average, it is possible that some areas within the United States meet the White House guidelines, even though reopening is contraindicated in general.

The opening of America involves two certainties. First, the United States will be COVID-19–free only when there is an effective vaccine. While scientists are working at unprecedented speed worldwide to develop a SARS-CoV-2 vaccine [6,108-113], realistically, it will be necessary to rely on best public health practices to minimize COVID-19 infection and mortality for at least one more year [110,114-116]. Second, the "social" end of the pandemic will occur before the "medical" end [117]; therefore, improved surveillance metrics are needed to inform health policy on opening sections of America more safely.

Acknowledgments

The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the US Agency for International Development.

Conflicts of Interest

None declared.

References

1. Kumari T, Shukla V. COVID-19: Towards Confronting an Unprecedented Pandemic. IJBI 2020;02(01):01-10. [doi: 10.46505/ijbi.2020.2101]
2. Baker S, Bloom N, Davis S, Kost K, Sammon M, Viratyosin T. The unprecedented stock market impact of COVID-19. NBER Working Papers 2020 Apr:1-24. [doi: 10.3386/w26945]
3. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
4. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. COVID-19, SARS and MERS: are they closely related? Clin Microbiol Infect 2020 Jun;26(6):729-734 [FREE Full text] [doi: 10.1016/j.cmi.2020.03.026] [Medline: 32234451]
5. Collins FS, Stoffels P. Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV): An Unprecedented Partnership for Unprecedented Times. JAMA 2020 Jun 23;323(24):2455-2457. [doi: 10.1001/jama.2020.8920] [Medline: 32421150]
6. Corey L, Mascola JR, Fauci AS, Collins FS. A strategic approach to COVID-19 vaccine R&D. Science 2020 May 29;368(6494):948-950. [doi: 10.1126/science.abc5312] [Medline: 32393526]
7. Cohen J. Vaccine designers take first shots at COVID-19. Science 2020 Apr 03;368(6486):14-16. [doi: 10.1126/science.368.6486.14] [Medline: 32241928]
8. Shrivastava S, Shrivastava P. World Health Organization Declares Very High Global Risk of Transmission of Corona Virus Disease 2019. Prim Health Care 2020 May 25;20(2):339 [FREE Full text]
9. Thacker S, Berkelman R. Public health surveillance in the United States. Epidemiol Rev 1988;10(1):164-190. [doi: 10.1093/oxfordjournals.epirev.a036021] [Medline: 3066626]
10. Thacker SB, Berkelman RL, Stroup DF. The Science of Public Health Surveillance. J Public Health Pol 1989;10(2):187. [doi: 10.2307/3342679]
11. Thacker SB, Stroup DF. Public Health Surveillance and Health Services Research. In: Armenian HK, Shapiro S, editors. Epidemiology and Health Services. New York, NY: Oxford University Press; 1998:61-82.
12. Nsubuga P, White ME, Thacker SB, Anderson MA, Blount SB, Broome CV, et al. Public Health Surveillance: A Tool for Targeting and Monitoring Interventions. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB,

et al, editors. Disease Control Priorities in Developing Countries. 2nd edition. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2006.
13. Lee LM, Teutsch SM, Thacker SB, Louis ME. New York, NY: Oxford University Press; 2010.
14. Thacker S, Qualters J, Lee L, Centers for Disease Control and Prevention. Public health surveillance in the United States: evolution and challenges. MMWR Suppl 2012 Jul 27;61(3):3-9. [Medline: 22832990]
15. Thacker S, Stroup D. Future directions for comprehensive public health surveillance and health information systems in the United States. Am J Epidemiol 1994 Sep 01;140(5):383-397. [doi: 10.1093/oxfordjournals.aje.a117261] [Medline: 8067331]
16. Calain P. Exploring the international arena of global public health surveillance. Health Policy Plan 2007 Jan 08;22(1):2-12. [doi: 10.1093/heapol/czl034] [Medline: 17237489]
17. Lee LM, Thacker SB. Public health surveillance and knowing about health in the context of growing sources of health data. Am J Prev Med 2011 Dec;41(6):636-640. [doi: 10.1016/j.amepre.2011.08.015] [Medline: 22099242]
18. Teutsch S, Churchill R, editors. Principles and Practice of Public Health Surveillance, Second Edition. In: Oxford University Press, USA. New York, NY: Oxford University Press; Oct 05, 2000.
19. Romaguera R, German R, Klaucke D, Teutsch S, Churchill R. Evaluating Public Health Surveillance. In: Teutsch S, Churchill R, editors. Principles and Practice of Public Health Surveillance, Second Edition. New York, NY: Oxford University Press; 2000:176-193.
20. Teutsch S. Considerations in Planning a Surveillance System. New York, NY: Oxford University Press; 2010:18-28.
21. Baker S, Bloom N, Davis S, Kost K, Sammon M, Viratyosin T. The Unprecedented Stock Market Reaction to COVID-19. Covid Economics, Vetted and Real-Time Papers 2020 Mar 03;1(3):33-42. [doi: 10.3386/w26945]
22. Yu J, Pang Y. 'Is it wise to open up America again now?' Learning from government responses to COVID-19 and economy reopen points in three Asia-Pacific countries. Br J Health Care Manag 2020 Jul 02;26(7):1-5. [doi: 10.12968/bjhc.2020.0071]
23. Harris J. Reopening Under COVID-19: What to Watch For. NBER Working Papers 2020 May:898-2937. [doi: 10.3386/w27166]
24. Weinberger DM, Chen J, Cohen T, Crawford FW, Mostashari F, Olson D, et al. Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020. JAMA Intern Med 2020 Jul 01:online. [doi: 10.1001/jamainternmed.2020.3391] [Medline: 32609310]
25. Jewell N, Lewnard J, Jewell B. Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic. Ann Intern Med 2020 Aug 04;173(3):226-227. [doi: 10.7326/m20-1565]
26. Fielding-Miller R, Sundaram M, Brouwer K. Social determinants of COVID-19 mortality at the county level. medRxiv 2020 Jul 01:preprint [FREE Full text] [doi: 10.1101/2020.05.03.20089698] [Medline: 32637976]
27. Faust JS, Del Rio C. Assessment of Deaths From COVID-19 and From Seasonal Influenza. JAMA Intern Med 2020 Aug 01;180(8):1045-1046. [doi: 10.1001/jamainternmed.2020.2306] [Medline: 32407441]
28. Wetzler HP, Wetzler EA. COVID-19 Excess Deaths in the United States, New York City, and Michigan During April 2020. medRxiv 2020 May 19:preprint. [doi: 10.1101/2020.04.02.20051532]
29. Hart OE, Halden RU. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. Sci Total Environ 2020 Aug 15;730:138875 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138875] [Medline: 32371231]
30. Quast T, Andel R. Excess mortality and potential undercounting of COVID-19 deaths by demographic group in Ohio. medRxiv 2020 Jul 17:preprint. [doi: 10.1101/2020.06.28.20141655]
31. Ciminelli G, Garcia-Mandicó S. Covid-19 in italy: an analysis of death registry data. VOXEU, Centre for Economic Policy Research. 2020 Apr 22. URL: https://voxeu.org/article/covid-19-italy-analysis-death-registry-data [accessed 2020-09-15]
32. Williamson GD, Weatherby Hudson G. A monitoring system for detecting aberrations in public health surveillance reports. Statist Med 1999 Dec 15;18(23):3283-3298. [doi: 10.1002/(sici)1097-0258(19991215)18:23<3283::aid-sim316>3.0.co;2-z]
33. Yang Y, Peng F, Wang R, Yange M, Guan K, Jiang T, et al. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J Autoimmun 2020 May;109:102434 [FREE Full text] [doi: 10.1016/j.jaut.2020.102434] [Medline: 32143990]
34. Ng Y, Li Z, Chua YX, Chaw WL, Zhao Z, Er B, et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore - January 2-February 29, 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 20;69(11):307-311 [FREE Full text] [doi: 10.15585/mmwr.mm6911e1] [Medline: 32191691]
35. Pastor C. Sentiment Analysis on Synchronous Online Delivery of Instruction due to Extreme Community Quarantine in the Philippines caused by Covid-19 Pandemic. AJMS 2020 Mar 20;3(1):1-6 [FREE Full text]
36. Piguillem F, Shi L. Optimal Covid-19 Quarantine and Testing Policies: CEPR Discussion Paper No. DP14613. SSRN 2020 Apr:1-42.
37. Sjödin H, Wilder-Smith A, Osman S, Farooq Z, Rocklöv J. Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020. Eurosurveillance 2020 Apr 02;25(13):2000280. [doi: 10.2807/1560-7917.es.2020.25.13.2000280]
38. Grant A. The reproduction number R for COVID-19 in England: Why hasn't "lockdown" been more effective? medRxiv 2020:preprint. [doi: 10.1101/2020.07.02.20144840]

39. Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. Lancet Infect Dis 2020 Jun;20(6):631-633. [doi: 10.1016/s1473-3099(20)30190-0]
40. Cavanagh G, Wambier CG. Rational hand hygiene during the coronavirus 2019 (COVID-19) pandemic. J Am Acad Dermatol 2020 Jun;82(6):e211 [FREE Full text] [doi: 10.1016/j.jaad.2020.03.090] [Medline: 32268168]
41. Water, sanitation, hygiene and waste management for COVID-19: technical brief, 03 March 2020. World Health Organization. 2020 Mar 03. URL: https://apps.who.int/iris/handle/10665/331305 [accessed 2020-09-15]
42. Abtahi-Naeini B. Frequent handwashing amidst the COVID-19 outbreak: prevention of hand irritant contact dermatitis and other considerations. Health Sci Rep 2020 Jun 27;3(2):e163 [FREE Full text] [doi: 10.1002/hsr2.163] [Medline: 32346616]
43. Tabari P, Amini M, Moghadami M, Moosavi M. International Public Health Responses to COVID-19 Outbreak: A Rapid Review. Iran J Med Sci 2020 May;45(3):157-169 [FREE Full text] [doi: 10.30476/ijms.2020.85810.1537] [Medline: 32546882]
44. Doung-ngern P, Suphanchaimat R, Panjangampatthana A, Janekrongtham C, Ruampoom D, Daochaeng N, et al. Associations between Wearing Masks, Washing Hands, and Social Distancing Practices, and Risk of COVID-19 Infection in Public: A Cohort-Based Case-Control Study in Thailand. SSRN Journal 2020 Aug 14:preprint. [doi: 10.2139/ssrn.3633162]
45. Infection prevention and control guidance for long-term care facilities in the context of COVID-19: interim guidance, 21 March 2020. World Health Organization. 2020 Mar 21. URL: https://apps.who.int/iris/handle/10665/331508 [accessed 2020-09-15]
46. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
47. Zhang L, Liu Y. Potential interventions for novel coronavirus in China: A systematic review. J Med Virol 2020 May 03;92(5):479-490 [FREE Full text] [doi: 10.1002/jmv.25707] [Medline: 32052466]
48. Stoecklin S, Rolland P, Silue Y. First cases of coronavirus disease 2019 (COVID-19) in Franceurveillance, investigations and control measures, January 2020. Eurosurveillance 2020 Jan;25(6):2000094. [doi: 10.2807/1560-7917.es.2020.25.6.2000094]
49. Tan SXD, Chen L. Real-Time Differential Epidemic Analysis and Prediction for COVID-19 Pandemic. arXiv. 2020 Apr 15. URL: https://arxiv.org/abs/2004.06888 [accessed 2020-09-15]
50. Gandhi M, Yokoe D, Havlir D. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med 2020 May 28;382(22):2158-2160. [doi: 10.1056/nejme2009758]
51. Cespedes MDS, Souza JCRPD. Coronavirus: a clinical update of Covid-19. Rev Assoc Med Bras (1992) 2020 May 15;66(2):116-123. [doi: 10.1590/1806-9282.66.2.116] [Medline: 32428144]
52. Calvo R, Deterding S, Ryan R. Health surveillance during covid-19 pandemic. BMJ 2020 Apr 06;369:m1373. [doi: 10.1136/bmj.m1373] [Medline: 32253180]
53. Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. J Epidemiol Community Health 2020 Jun 24;74(7):612-616. [doi: 10.1136/jech-2018-211654] [Medline: 32332114]
54. Desjardins M, Hohl A, Delmelle E. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Appl Geogr 2020 May;118:102202 [FREE Full text] [doi: 10.1016/j.apgeog.2020.102202] [Medline: 32287518]
55. Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, El Burai Felix S, et al. Coronavirus Disease 2019 Case Surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):759-765 [FREE Full text] [doi: 10.15585/mmwr.mm6924e2] [Medline: 32555134]
56. Barros JM, Duggan J, Rebholz-Schuhmann D. The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]
57. Iacopini I, Petri G, Barrat A, Latora V. Simplicial models of social contagion. Nat Commun 2019 Jun 06;10(1):2485 [FREE Full text] [doi: 10.1038/s41467-019-10431-6] [Medline: 31171784]
58. Stock JH. Data gaps and the policy response to the novel coronavirus. NBER Working Papers 2020 Mar:1-12. [doi: 10.3386/w26902]
59. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19. IEEE Rev Biomed Eng 2020 Apr 16:1-1. [doi: 10.1109/rbme.2020.2987975]
60. Naudé W. Artificial Intelligence against COVID-19: An early review. IZA Institute of Labor Economics. 2020 Apr. URL: http://ftp.iza.org/dp13110.pdf [accessed 2020-09-15]
61. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Liu L, et al. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. arXiv. 2020 Mar 26. URL: https://arxiv.org/abs/2003.11988 [accessed 2020-09-15]
62. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19). US Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html [accessed 2020-07-25]

63. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome. Science 2003 Jun 20;300(5627):1966-1970. [doi: 10.1126/science.1086616] [Medline: 12766207]
64. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 — Studies Needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/nejmp2002125]
65. Zhao S, Gao D, Zhuang Z. Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. medRxiv 2020 Feb 25:preprint. [doi: 10.1101/2020.02.21.20026559]
66. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc Natl Acad Sci USA 2020 May 12;117(19):10484-10491 [FREE Full text] [doi: 10.1073/pnas.2004978117] [Medline: 32327608]
67. You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. Int J Hyg Environ Health 2020 Jul;228:113555 [FREE Full text] [doi: 10.1016/j.ijheh.2020.113555] [Medline: 32460229]
68. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477 [FREE Full text] [doi: 10.3201/eid2607.200282] [Medline: 32255761]
69. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun 22;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
70. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. medRxiv 2020 Feb 18:preprint. [doi: 10.1101/2020.02.16.20023465]
71. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020 Apr;93:284-286 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.060] [Medline: 32145466]
72. Sha A, Samanta S, Martcheva M, Chattopadhyay J. Backward bifurcation, oscillations and chaos in an eco-epidemiological model with fear effect. J Biol Dyn 2019 Dec 03;13(1):301-327. [doi: 10.1080/17513758.2019.1593525] [Medline: 31046638]
73. Chen L. Persistent spatial patterns of interacting contagions. Phys Rev E 2019 Feb 14;99(2):022308. [doi: 10.1103/physreve.99.022308]
74. Kwuimy CAK, Nazari F, Jiao X, Rohani P, Nataraj C. Nonlinear dynamic analysis of an epidemiological model for COVID-19 including public behavior and government action. Nonlinear Dyn 2020 Jul 16:1-15 [FREE Full text] [doi: 10.1007/s11071-020-05815-z] [Medline: 32836814]
75. Li X, Yang J, Martcheva M. Age-Structured Epidemic Models. In: Age Structured Epidemic Modeling. Interdisciplinary Applied Mathematics, vol 52. Cham, Switzerland: Springer; May 29, 2020:23-67.
76. Martcheva M. An Introduction to Mathematical Epidemiology. Cham, Switzerland: Springer; 2015.
77. Martcheva M, Castillo-Chavez C. Diseases with chronic stage in a population with varying size. Mathematical Biosciences 2003 Mar;182(1):1-25. [doi: 10.1016/s0025-5564(02)00184-0]
78. Wiley L. Public health law and science in the community mitigation strategy for Covid-19. J Law Biosci 2020;7(1):lsaa019 [FREE Full text] [doi: 10.1093/jlb/lsaa019] [Medline: 32728464]
79. Lambert C, Stoicu S, Nestsiarovich A, Kumar P, Lauve N, Sloane L, et al. 2020-04-17/18 Daily UNM Global Health COVID-19 Briefing. University of New Mexico Digital Repository. 2020 Apr 24. URL: https://digitalrepository.unm.edu/hsc_covid19_briefings/10/ [accessed 2020-09-15]
80. Alhaery M, Suh E. A COVID-19 Reopening Readiness Index: The Key to Opening up the Economy. SSRN 2020 Jun 07:preprint. [doi: 10.1101/2020.05.22.20110577]
81. Ganz S. Does your state have 14-days of declining COVID-19 cases? Finding straightforward answers to a surprisingly tricky question. American Enterprise Institute. 2020 May 11. URL: https://www.jstor.org/stable/resrep24601?seq=2#metadata_info_tab_contents [accessed 2020-09-15]
82. Baqaee D, Farhi E, Mina M, Stock J. Reopening Scenarios. NBER Working Papers 2020 May:1-42. [doi: 10.3386/w27244]
83. Khan Z, Van BF, Hussain F. A predictive model for Covid-19 spread applied to six US states. arXiv. 2020 Jun 10. URL: https://arxiv.org/abs/2006.05955 [accessed 2020-09-15]
84. The COVID Tracking Project. The Atlantic. 2020. URL: https://covidtracking.com/ [accessed 2020-07-10]
85. COVID 19 Tracking 2020. GitHub. 2020. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2020-09-15]
86. 2019 National and State Population Estimates. United States Census Bureau. 2019 Dec 30. URL: https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html [accessed 2020-08-15]
87. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Rev Econ Stud 1991 Apr;58(2):277. [doi: 10.2307/2297968]
88. Montes-Rojas G, Sosa-Escudero W, Zincenko F. Level-Based Estimation of Dynamic Panel Models. J Econom Method 2019 Aug 01;9(1):1-23. [doi: 10.1515/jem-2018-0015]
89. Fritsch M. On GMM estimation of linear dynamic panel data models. Passauer Diskussionspapiere-Betriebswirtschaftliche Reihe No. B-36-19. 2019 Sep 20. URL: https://www.econstor.eu/bitstream/10419/204581/1/1678187852.pdf [accessed 2020-09-15]

90. Liu L, Moon HR, Schorfheide F. Forecasting With Dynamic Panel Data Models. Econometrica 2020;88(1):171-201. [doi: 10.3982/ecta14952]
91. Kock AB, Tang H. Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. Econom Theory 2018 May 24;35(2):295-359. [doi: 10.1017/s0266466618000087]
92. Bera AK, Doğan O, Taşpınar S, Leiluo Y. Robust LM tests for spatial dynamic panel data models. Regional Science and Urban Economics 2019 May;76:47-66. [doi: 10.1016/j.regsciurbeco.2018.08.001]
93. Judson RA, Owen AL. Estimating dynamic panel data models: a guide for macroeconomists. Econ Lett 1999 Oct;65(1):9-15. [doi: 10.1016/s0165-1765(99)00130-5]
94. Bowsher CG. On testing overidentifying restrictions in dynamic panel data models. Econ Lett 2002 Oct;77(2):211-220. [doi: 10.1016/s0165-1765(02)00130-1]
95. Blundell R, Bond S, Windmeijer F. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator. Adv Econ 2020 Jul 01;15:online. [doi: 10.1016/S0731-9053(00)15003-0]
96. Anser MK, Yousaf Z, Khan MA, Nassani AA, Alotaibi SM, Qazi Abro MM, et al. Does communicable diseases (including COVID-19) may increase global poverty risk? A cloud on the horizon. Environ Res 2020 Aug;187:109668 [FREE Full text] [doi: 10.1016/j.envres.2020.109668] [Medline: 32422482]
97. Ogaki M. Generalized method of moments: Econometric applications. In: Maddala GS, Rao CR, Vinod HD, editors. Handbook of Statistics Volume 11. Amsterdam, Netherlands: Elsevier; Nov 01, 1993:455-488.
98. Roodman D. How to do Xtabond2: An Introduction to Difference and System GMM in Stata. Stata J 2018 Nov 19;9(1):86-136. [doi: 10.1177/1536867x0900900106]
99. Brown BW, Newey WK. Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. J Bus Econ Stat 2002 Oct;20(4):507-517. [doi: 10.1198/073500102288618649]
100. Hansen BE, West KD. Generalized Method of Moments and Macroeconomics. J Bus Econ Stat 2002 Oct;20(4):460-469. [doi: 10.1198/073500102288618603]
101. Stock JH, Wright JH, Yogo M. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. J Bus Econ Stat 2002 Oct;20(4):518-529. [doi: 10.1198/073500102288618658]
102. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 1982 Jul;50(4):1029-1054. [doi: 10.2307/1912775]
103. Windmeijer FAG, Santos Silva JMC. Endogeneity in count data models: an application to demand for health care. J Appl Econ 1997 May;12(3):281-294. [doi: 10.1002/(sici)1099-1255(199705)12:3<281::aid-jae436>3.0.co;2-1]
104. Lewbel A. Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. J Bus Econ Stat 2012 Feb 22;30(1):67-80. [doi: 10.1080/07350015.2012.643126]
105. Newey WK. Generalized method of moments specification testing. J Econom 1985 Sep;29(3):229-256. [doi: 10.1016/0304-4076(85)90154-x]
106. USA Facts. URL: https://usafacts.org/issues/coronavirus/ [accessed 2020-09-16]
107. COVID-19 Dashboard. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-09-16]
108. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. N Engl J Med 2020 May 21;382(21):1969-1973. [doi: 10.1056/nejmp2005630]
109. Shoenfeld Y. Corona (COVID-19) time musings: Our involvement in COVID-19 pathogenesis, diagnosis, treatment and vaccine planning. Autoimmun Rev 2020 Jun;19(6):102538 [FREE Full text] [doi: 10.1016/j.autrev.2020.102538] [Medline: 32268212]
110. Yamey G, Schäferhoff M, Hatchett R, Pate M, Zhao F, McDade KK. Ensuring global access to COVID-19 vaccines. Lancet 2020 May;395(10234):1405-1406. [doi: 10.1016/s0140-6736(20)30763-7]
111. Wu S. Progress and Concept for COVID-19 Vaccine Development. Biotechnol J 2020 Jun;15(6):e2000147 [FREE Full text] [doi: 10.1002/biot.202000147] [Medline: 32304139]
112. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov 2020 May;19(5):305-306. [doi: 10.1038/d41573-020-00073-5] [Medline: 32273591]
113. Cohen J. U.S. 'Warp Speed' vaccine effort comes out of the shadows. Science 2020 May 15;368(6492):692-693. [doi: 10.1126/science.368.6492.692] [Medline: 32409451]
114. Graham BS. Rapid COVID-19 vaccine development. Science 2020 May 29;368(6494):945-946. [doi: 10.1126/science.abb8923] [Medline: 32385100]
115. Schaffer DeRoo S, Pudalov NJ, Fu LY. Planning for a COVID-19 Vaccination Program. JAMA 2020 Jun 23;323(24):2458-2459. [doi: 10.1001/jama.2020.8711] [Medline: 32421155]
116. Smiai H, Mirimi M, Najar M, Boukhatem N. COVID-19 update : what's going on? J Mater Environ Sci 2020 May 13;11(6):877-884.
117. Post LA. When will the Covid-19 pandemic be over? In: Post LA, editor. Buehler Bulletin. Evanston, IL: Buehler Center Media; 2020.

Abbreviations

DPD: dynamic panel data
GMM: generalized method of moments
R_0: basic reproduction number
SIR: susceptible-infected-recovered

Edited by G Eysenbach, G Fagherazzi; submitted 01.06.20; peer-reviewed by J Marquard, M Adenomon, C Moss, G Onel; comments to author 09.07.20; revised version received 19.07.20; accepted 09.09.20; published 22.09.20.

Please cite as:
Oehmke JF, Oehmke TB, Singh LN, Post LA
Dynamic Panel Estimate–Based Health Surveillance of SARS-CoV-2 Infection Rates to Inform Public Health Policy: Model Development and Validation
J Med Internet Res 2020;22(9):e20924
URL: http://www.jmir.org/2020/9/e20924/
doi:10.2196/20924
PMID:32915762

Original Paper

Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19

Shaoxiong Sun[1], PhD; Amos A Folarin[1,2], PhD; Yatharth Ranjan[1], MSc; Zulqarnain Rashid[1], PhD; Pauline Conde[1], BSc; Callum Stewart[1], MSc; Nicholas Cummins[3], PhD; Faith Matcham[4], PhD; Gloria Dalla Costa[5], MD; Sara Simblett[6], PhD; Letizia Leocani[5], MD, PhD; Femke Lamers[7], PhD; Per Soelberg Sørensen[8], MD; Mathias Buron[8], MD; Ana Zabalza[9], MD; Ana Isabel Guerrero Pérez[9], MSc; Brenda WJH Penninx[7], PhD; Sara Siddi[10,11,12], PhD; Josep Maria Haro[10,11,12], MD; Inez Myin-Germeys[13], PhD; Aki Rintala[13], MSc; Til Wykes[6,14], PhD; Vaibhav A Narayan[15], PhD; Giancarlo Comi[16], MD; Matthew Hotopf[4,14], PhD; Richard JB Dobson[1,2], PhD; RADAR-CNS Consortium[17]

[1]The Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]Institute of Health Informatics, University College London, London, United Kingdom

[3]Chair of Embedded Intelligence for Health Care & Wellbeing, University of Augsburg, Augsburg, Germany

[4]The Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[5]Neurorehabilitation Unit and Institute of Experimental Neurology, University Vita Salute San Raffaele, Istituto Di Ricovero e Cura a Carattere Scientifico Ospedale San Raffaele, Milan, Italy

[6]The Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[7]Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam University Medical Centre, Vrije Universiteit and GGZ inGeest, Amsterdam, Netherlands

[8]Danish Multiple Sclerosis Centre, Department of Neurology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

[9]Multiple Sclerosis Centre of Catalonia (Cemcat), Department of Neurology/Neuroimmunology, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[10]Parc Sanitari Sant Joan de Déu, Fundació Sant Joan de Déu, Barcelona, Spain

[11]Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain

[12]Universitat de Barcelona, Barcelona, Spain

[13]Centre for Contextual Psychiatry, Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium

[14]South London and Maudsley National Health Services Foundation Trust, London, United Kingdom

[15]Janssen Research and Development LLC, Titusville, NJ, United States

[16]Institute of Experimental Neurology, Istituto Di Ricovero e Cura a Carattere Scientifico Ospedale San Raffaele, Milan, Italy

[17]The RADAR-CNS Consortium, London, United Kingdom

Corresponding Author:
Shaoxiong Sun, PhD
The Department of Biostatistics and Health Informatics
Institute of Psychiatry, Psychology and Neuroscience
King's College London
Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London
PO Box 80 De Crespigny Park, Denmark Hill
London, SE58AF
United Kingdom
Phone: 44 02078480951
Email: shaoxiong.sun@kcl.ac.uk

Abstract

Background: In the absence of a vaccine or effective treatment for COVID-19, countries have adopted nonpharmaceutical interventions (NPIs) such as social distancing and full lockdown. An objective and quantitative means of passively monitoring the impact and response of these interventions at a local level is needed.

Objective: We aim to explore the utility of the recently developed open-source mobile health platform Remote Assessment of Disease and Relapse (RADAR)–base as a toolbox to rapidly test the effect and response to NPIs intended to limit the spread of COVID-19.

Methods: We analyzed data extracted from smartphone and wearable devices, and managed by the RADAR-base from 1062 participants recruited in Italy, Spain, Denmark, the United Kingdom, and the Netherlands. We derived nine features on a daily basis including time spent at home, maximum distance travelled from home, the maximum number of Bluetooth-enabled nearby devices (as a proxy for physical distancing), step count, average heart rate, sleep duration, bedtime, phone unlock duration, and social app use duration. We performed Kruskal-Wallis tests followed by post hoc Dunn tests to assess differences in these features among baseline, prelockdown, and during lockdown periods. We also studied behavioral differences by age, gender, BMI, and educational background.

Results: We were able to quantify expected changes in time spent at home, distance travelled, and the number of nearby Bluetooth-enabled devices between prelockdown and during lockdown periods (P<.001 for all five countries). We saw reduced sociality as measured through mobility features and increased virtual sociality through phone use. People were more active on their phones (P<.001 for Italy, Spain, and the United Kingdom), spending more time using social media apps (P<.001 for Italy, Spain, the United Kingdom, and the Netherlands), particularly around major news events. Furthermore, participants had a lower heart rate (P<.001 for Italy and Spain; P=.02 for Denmark), went to bed later (P<.001 for Italy, Spain, the United Kingdom, and the Netherlands), and slept more (P<.001 for Italy, Spain, and the United Kingdom). We also found that young people had longer homestay than older people during the lockdown and fewer daily steps. Although there was no significant difference between the high and low BMI groups in time spent at home, the low BMI group walked more.

Conclusions: RADAR-base, a freely deployable data collection platform leveraging data from wearables and mobile technologies, can be used to rapidly quantify and provide a holistic view of behavioral changes in response to public health interventions as a result of infectious outbreaks such as COVID-19. RADAR-base may be a viable approach to implementing an early warning system for passively assessing the local compliance to interventions in epidemics and pandemics, and could help countries ease out of lockdown.

(J Med Internet Res 2020;22(9):e19992) doi:10.2196/19992

KEYWORDS

mobile health; COVID-19; behavioral monitoring; smartphones; wearable devices; mobility; phone use

Introduction

On March 11, 2020, the World Health Organization declared the rapidly spreading SARS-CoV-2 virus outbreak a pandemic. This novel coronavirus is the cause of a contagious acute respiratory disease (COVID-19), which was first reported in Wuhan, Hubei Province, China [1-3]. As of July 1, 2020, it had infected over 10 million people and spread to 213 countries and territories around the world [4]. Although precise statistics on mortality are being determined, COVID-19 can be deadly with an estimated 1% case fatality rate, and this rate increases for older adults and those with underlying health problems [5,6]. The outbreak of COVID-19 has placed an unprecedented burden on health care systems in most-affected countries and has resulted in considerable economic losses and a possible global recession [7,8].

To date, there is no vaccine or highly effective treatment. The widely adopted strategy has been the use of nonpharmaceutical interventions (NPIs) such as social distancing and even full lockdown to control the spread of the virus and ease the pressure on health and care systems [9,10]. NPIs have been implemented in many countries including China, Italy, Spain, the United Kingdom, and the Netherlands. These measures have been shown to considerably reduce the new confirmed cases [9]. Key to the success of NPIs is the timing of these interventions and the response of the population, both of which might differ among countries and could necessitate further interventions in the case of low compliance either nationally or locally. Furthermore, US $11 trillion of fiscal measures have been announced by more than two-thirds of governments worldwide in an attempt to mitigate the fallout from the pandemic and lockdown [11]. Therefore, we urgently require an objective and quantitative way to monitor population behavior to assess the impact and response of such interventions. Additionally, we need to monitor for the potential effects of a rebound in cases in the winter months as social distancing measures are relaxed and to strategize and understand where course corrections are required. Similarly, understanding potential seasonal forcing of COVID-19 will require a good understanding of different NPIs' effects, so they can be factored out.

The increasing availability of wide-bandwidth mobile networks, smartphones, and wearable sensors makes it possible to collect near real-time high-resolution data sets from large numbers of participants and greatly facilitates remote monitoring of behavior [12-14]. By leveraging sensor modalities in smartphones, which includes network and GPS location tracking, and Fitbit devices, which includes step counts and heart rate, it is possible to access mobility and even wellness for the population. To manage the data collected from multiple sensor modalities and mobile devices, platforms such as the open-source Remote Assessment of Disease and Relapse (RADAR)–base [15] mobile health platform have been developed [16]. This platform has been used to enable remote monitoring in a range of use cases including central nervous system diseases (major depressive disorder [MDD], epilepsy, and multiple sclerosis [MS]) as part of the

Innovative Medicines Initiative (IMI2) RADAR–Central Nervous System (CNS) major program [17,18].

In this paper, we explore the utility of the RADAR-base platform as a toolbox to test the effect and response of NPIs aimed at limiting the spread of infectious diseases such as COVID-19 by leveraging participant data already collected from November 2017 onward as part of the ongoing RADAR-CNS studies [16,17,19]. Specifically, we created measures of mobility (as a proxy of physical distancing), phone use (as a proxy of virtual sociality), and physiological measures (heart rate and sleep), and compared these features among the baseline, prelockdown, and during lockdown periods. Furthermore, we also provide a joint analysis of these features to provide a holistic view and interpret these behavioral changes during COVID-19.

Methods

Data Collection

The RADAR-CNS studies were approved by all local ethics committees, and all participants signed informed consent [19]. We included 1062 participants recruited in five European countries: Italy, Spain, Denmark, the United Kingdom, and the Netherlands. Participants in the Netherlands were partially recruited through Hersenonderzoek.nl [20]. The data were collected for the purpose of finding new ways of monitoring MDD (Spain: n=150; the Netherlands: n=103; and the United Kingdom: n=316) and MS (Milan, Italy: n=208; Barcelona, Spain: n=179; and Copenhagen, Denmark: n=106) using wearable devices and smartphone technology to improve patients' quality of life (QoL) and potentially change the treatment of these and other chronic disorders. As we focused on country-level behavioral changes in response to the NPIs, we aggregated data collected in Spain and did not focus on analyzing differences between participants with MDD and MS. Passive participant data, that is data that did not require conscious participant engagement, were collected continuously on a 24/7 basis through a smartphone and a Fitbit device, which included location, Bluetooth, activity, sleep, heart rate, and phone use data. In this study, we used participants' own Android smartphones where available and provided a participant with a Motorola G5, G6, or G7 if participants had an iPhone or did not have a smartphone. For Fitbit devices, Fitbit Charge 2 devices were given to participants, and then Fitbit Charge 3 devices were given to the recently recruited participants when Fitbit Charge 2 devices were no longer available. We asked participants to wear the device on their nondominant hand. Although not used for this study, active data were also collected, which required clinicians or participants to fill out emailed surveys (eg, Inventory of Depressive Symptomatology [Self-Report]), app-delivered questionnaires (eg, Patient Health Questionnaire), or perform short clinical tests (eg, Expanded Disability Status Scale).

The data collection and management were handled by the open-source mHealth platform RADAR-Base [16]. The platform provides high scalability, interoperability, flexibility, and reliability while allowing the freedom for anyone to deploy. Due to the streaming first nature of the platform, it is also easy to aggregate, analyze, and provide insights into the data in real time, hence making the results of this work potentially deployable for localized monitoring and targeted interventions.

Feature Extraction

To study physical-behavioral changes in response to COVID-19 NPIs, we examined participants' mobility by analyzing relative location and Bluetooth data from smartphones and step count data from Fitbit devices. We investigated phone unlock duration and social app use duration to study online social-behavioral changes. Physiological measures such as sleep and heart rate from Fitbit devices were also analyzed to identify possible changes as a result of lockdown. A full list of features is presented in Table 1. These features were extracted for each participant every day. The daily features were calculated using the data from 6 AM on the present day to 6 AM on the next day for all features except total sleep duration and bedtime, where 8 PM was used as the starting time point and 11 AM the finishing. When no data were found in a data modality for a participant on a day due to the participant not wearing the Fitbit device or not using the smartphone, we did not calculate the feature derived from that data modality on that day.

Table 1. A full list of extracted features.

Category and modality	Features	Extraction
Mobility		
Smartphone location	• Maximum travelled distance from home • Homestay	• The maximum distance travelled from home location • The time spent within 200 m radius of home location (determined using DBSCAN[a])
Smartphone Bluetooth	• Maximum number of nearby devices	• The maximum number of Bluetooth-enabled nearby devices
Fitbit step count	• Step count	• Daily total of Fitbit step count
Physiological measures		
Fitbit sleep	• Bedtime • Sleep duration	• The first sleep category at night • Daily total duration of sleep categories (light, deep, and REM[b])
Fitbit heart rate	• Average heart rate	• The daily average heart rate
Phone use		
Smartphone user interaction	• Unlock duration	• The total duration of phone in the unlocked state
Smartphone use event	• Social app use duration	• The total duration spent on social apps (Google Play categories of Social, Communication, and Dating)

[a]DBSCAN: density-based spatial clustering of applications with noise.

[b]REM: rapid eye movement.

The smartphone-derived location data were sampled once every 5 minutes by default, with longer sampling durations dependent on network connectivity. Spurious location coordinates were identified and removed if they differed from preceding and following coordinates by more than five degrees. Home location was determined daily by clustering location data between 8 PM and 4 AM with the mean coordinate of the cluster that the last coordinate belonged to being used. This choice was made because the largest cluster may not be the home location for a single night but the last location before phones shut down had a higher probability to be home location for that night. The clustering was implemented using density-based spatial clustering of applications with noise [21]. A duration gated by two adjacent coordinates was regarded as a valid homestay duration on the condition that both coordinates were no further than 200 meters from the home location. A duration longer than 1 hour was excluded due to the large proportion of missing data when compared to the 5-minute sampling duration. All valid homestay durations between 8 AM and 11 PM were summed to calculate daily homestay. Daily maximum distance from home was also computed based on the coordinates in the same period.

Bluetooth data, including the number of nearby and paired devices, were also collected from smartphones, which were sampled every hour. The daily maximum number of nearby devices was used as a mobility feature. An increased number of nearby devices (typically other phones) detected may indicate other users' presence in the vicinity, which therefore can serve as a proxy of physical distancing.

In addition to mobility features extracted from smartphones, daily step count was taken from the Fitbit device, which was computed as the total steps a participant walked every day. Likewise, daily sleep duration was computed as the summation of three Fitbit-output sleep categories (light, deep, and rapid eye movement) sampled every 30 seconds from 8 PM to 11 AM the next day. Bedtime was defined as the time of the first sleep category reported by Fitbit after 8 PM. Note that the sleep categories referred to the sleep stages provided by the Fitbit application programming interface [22], which are not equivalent to the medical sleep stages. Finally, daily mean heart rate was calculated by averaging the Fitbit-output heart rate readings, sampled every 5 seconds at best. This sampling interval may be longer depending on Fitbit proprietary algorithms for remaining battery level, quality scoring, and network connectivity.

To explore changes in phone use, daily unlock duration was calculated by summing time periods starting with the unlocked state and ending with the standby state. Single intervals longer than 4 hours were excluded, which might result from a missing standby state or unintentionally leaving the phone unlocked. App use was quantified by classifying apps according to categories listed on Google Play [23]. As we were particularly interested in cyber-social interactions, we focused on the daily use time of social apps, including the Google Play categories of Social, Communication, and Dating. Among them are Facebook, Instagram, and WhatsApp.

Data Analysis

We plotted how the features evolved over 1.5 years for each country investigated. The participants' daily median, 25th percentile, and 75th percentile of each feature were calculated and then plotted. A minimum of 20 participants' data points was a prerequisite for calculation for any given day to reduce variance and noise. To facilitate interpretation, we also marked time points of public announcements related to lockdown policies [24].

To examine changes in mobility, physiological measures, and phone use induced by the lockdowns, comparisons among baseline, prelockdown, and during lockdown on the daily median of each feature were carried out using Kruskal-Wallis tests followed by post hoc Dunn tests [25,26]. For the during lockdown phase, we chose the entire period of the national lockdown in each country, which ended when NPIs were eased for the first time. For the prelockdown phase, we chose the period immediately prior to the first restrictive measure with the same length of the entire national lockdown. For the baseline phase, we chose the same period in 2019 as the 2020 national lockdown for countries starting to collect data earlier than 2019, which included Italy, Spain, and the United Kingdom. This was aimed at suppressing seasonal variability. For Denmark and the Netherlands where participant recruitment and data collection started much later, we chose the period that started with the earliest stable date (no considerate missing data or outliers) with the same length of the entire national lockdown. If a significant difference among these three periods was found after Benjamini–Yekutieli correction for the number of features (n=9), post hoc Dunn test was applied with Benjamini–Yekutieli correction for the number of groups (n=3) [27]. Box plots were used to present the results. A $P<.05$ after Benjamini–Yekutieli corrections was deemed statistically significant. It should be noted that we applied corrections resulting from multiple comparisons and multiple features in each country.

We also studied factors that might influence the subpopulation behavioral features during the lockdown period. The investigated factors included age, gender, BMI, and educational background. For age groups, we defined the young group as younger than 45 years and the older adult group as 45 years or older. For BMI groups, the low BMI group was defined as less than 25, and the high BMI group as greater than or equal to 25. For education groups, we defined the degree group as having a bachelor's degree or above and the nondegree group as having lower qualifications. Furthermore, we defined a combined factor group of young men, as this subpopulation was suspected to be less compliant with social distancing measures. Here we focused on features of homestay and daily step count during the entire period of lockdown for each country. We performed Wilcoxon signed rank tests on these two features to examine statistically significant differences. The *P* values were corrected with the number of factors (n=5) and the number of features (n=2) using Benjamini–Yekutieli correction.

Finally, we investigated the effects of different NPIs, in particular immediately after national lockdowns. This was done by comparing the NPIs implemented in the five countries within the first 2 weeks after entering national lockdowns.

Results

Plots showing how the extracted features evolved from February 1, 2019, to July 5, 2020, and box plots of these features are shown in Figures 1-5 and in Figure 6, respectively. Detailed test statistics and *P* values comparing prelockdown and during lockdown measures are presented in Table 2. Figure 7 shows a zoomed in version of Figures 3 and 4.

Figure 1. Behavioral changes for Milan, Italy (208 participants). (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, (i) social app duration. Solid line: median; shade: 25th percentile to 75th percentile. dur: duration; WHO: World Health Organization.

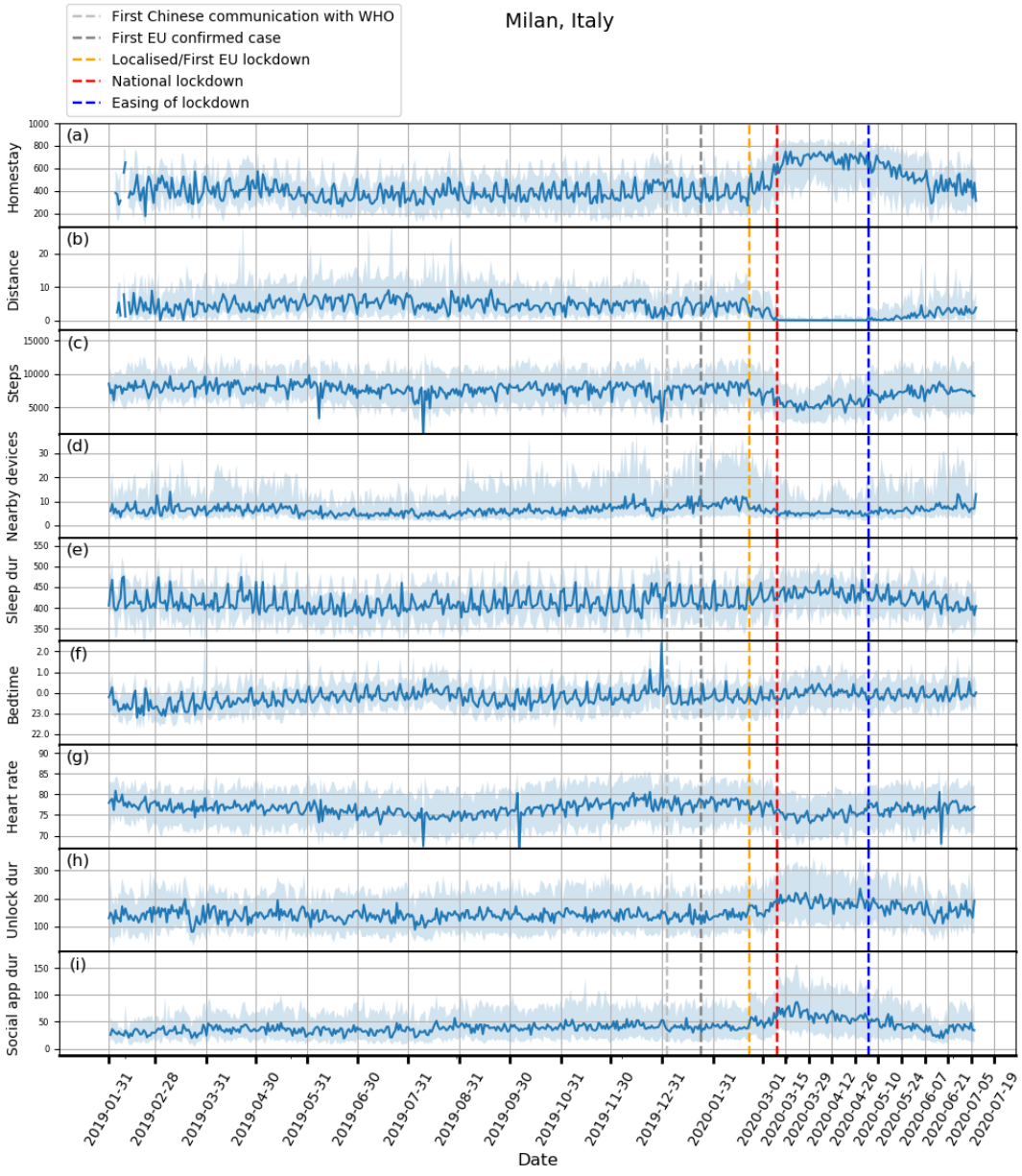
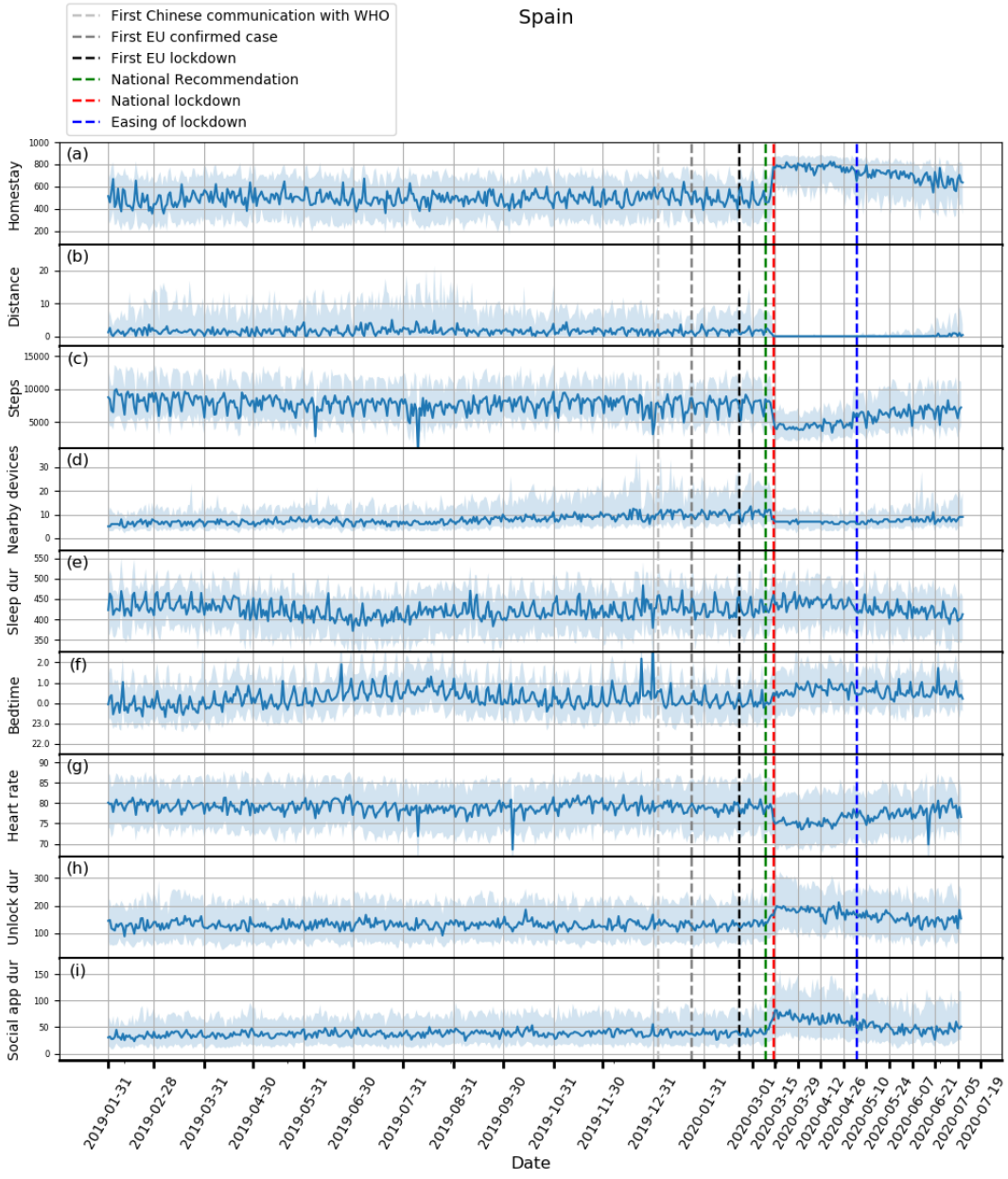


Figure 2. Behavioral changes for Spain (329 participants). (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, (i) social app duration. Solid line: median; shade: 25th percentile to 75th percentile. dur: duration; WHO: World Health Organization.



Through RADAR-base, we quantified changes in mobility, phone use, and physiological measures as a result of NPIs introduced to control COVID-19. As expected, following national lockdowns, participants in all countries stayed at home for longer, travelled shorter distances, walked less, and had fewer Bluetooth-enabled devices in the vicinity.

In contrast to increased physical distancing (reduced sociability) suggested by these mobility features, higher phone use, indicating compensatory sociability, was observed. Italy, Spain, and the United Kingdom saw longer unlock duration, and these 3 countries together with the Netherlands also showed longer social app use duration. Tellingly, both unlock duration and social app use duration saw peaks around the news of national lockdowns in all countries.

Concurrent with the changes in mobility and phone use, changes in physiological measures were observed. Participants in Spain, Italy, and the United Kingdom went to bed later and slept more. Participants in Spain, Italy, and Denmark also had a decrease in heart rate. Although not statistically significant, an increase in sleep duration and bedtime in Denmark and the Netherlands, and a decrease in heart rate in the United Kingdom and the Netherlands can be seen in Figures 3-5.

Figure 3. Behavioral changes for Copenhagen, Denmark (106 participants). (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, (i) social app duration. Solid line: median; shade: 25th percentile to 75th percentile. dur: duration; WHO: World Health Organization.

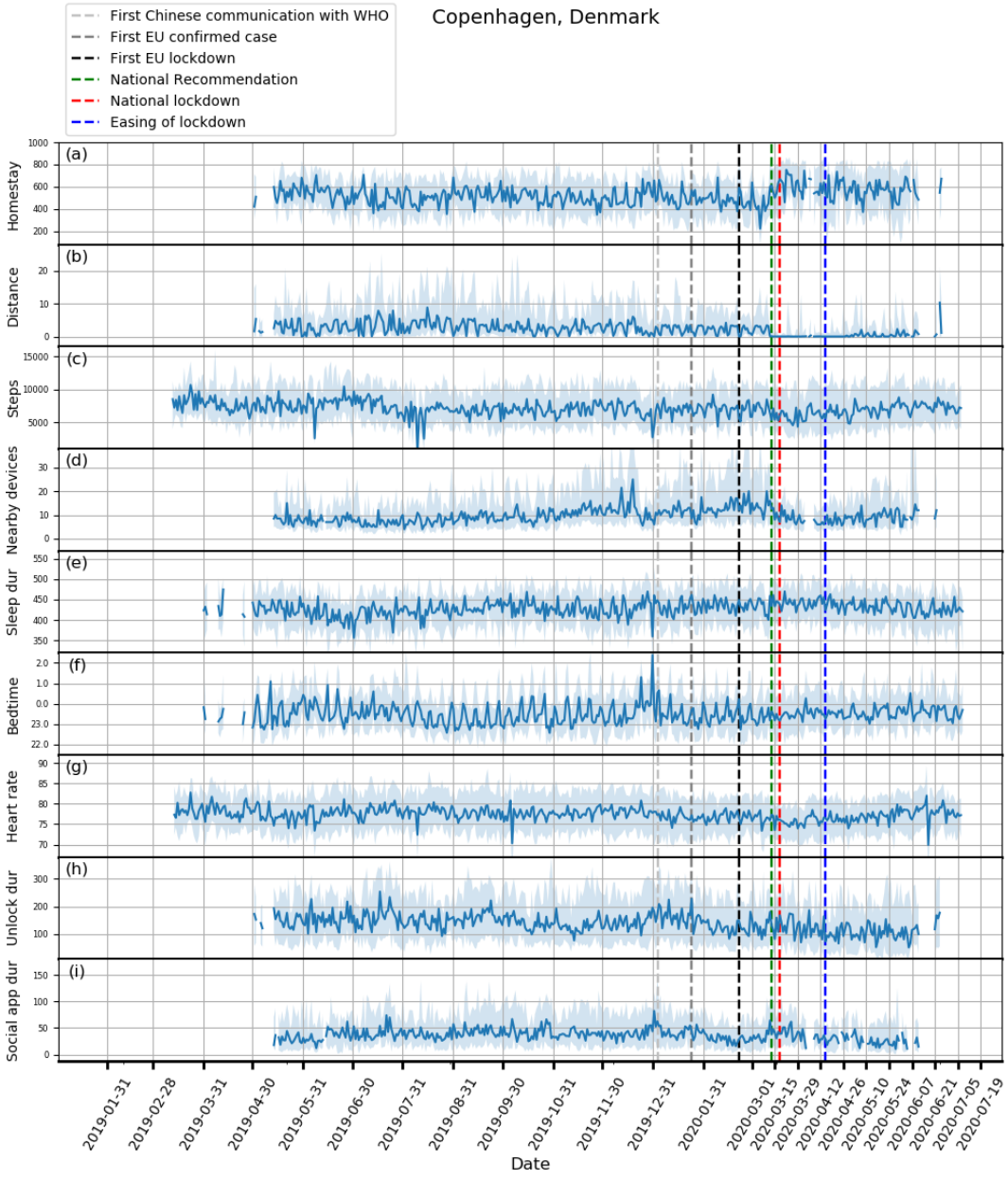


Figure 4. Behavioral changes for the United Kingdom (316 participants). (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, (i) social app duration. Solid line: median; shade: 25th percentile to 75th percentile. dur: duration; WHO: World Health Organization.

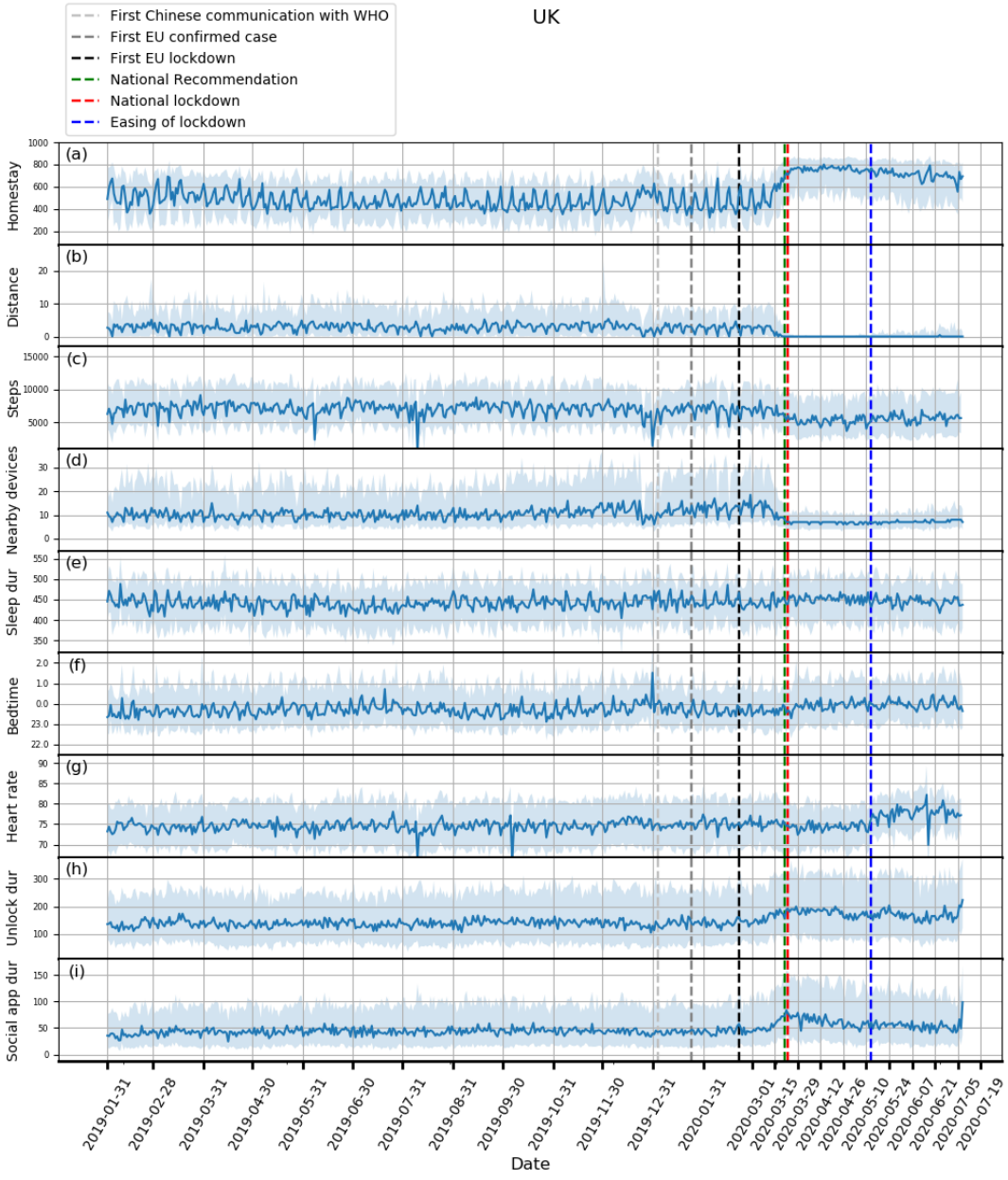


Figure 5. Behavioral changes for the Netherlands (103 participants). (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, (i) social app duration. Solid line: median; shade: 25th percentile to 75th percentile. dur: duration; WHO: World Health Organization.

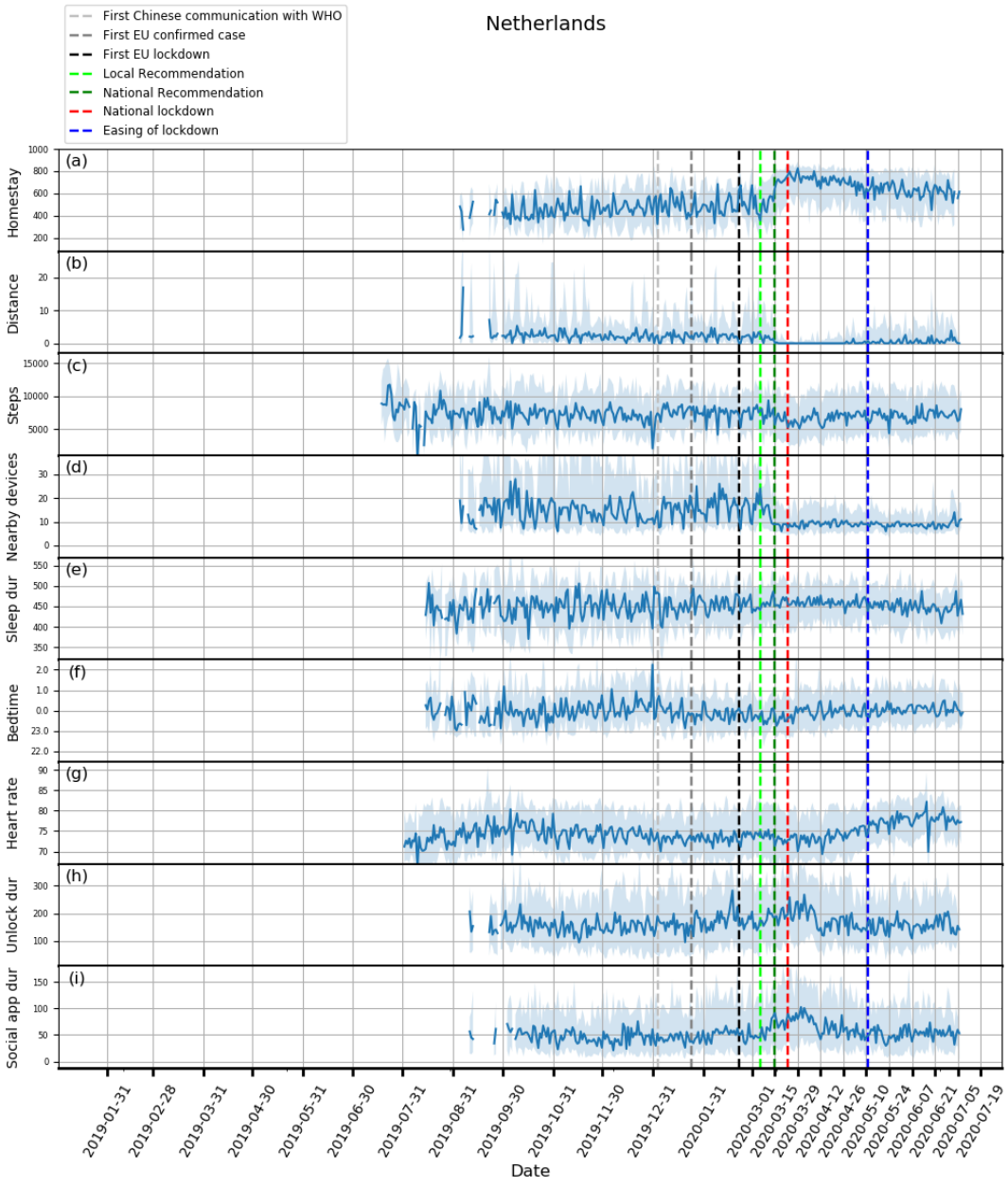


Figure 6. Box plots for comparisons among baseline, prelockdown, and during lockdown phases for different features. (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, and (i) social app duration. *means $P<.05$, **means $P<.01$, ***means $P<.001$.

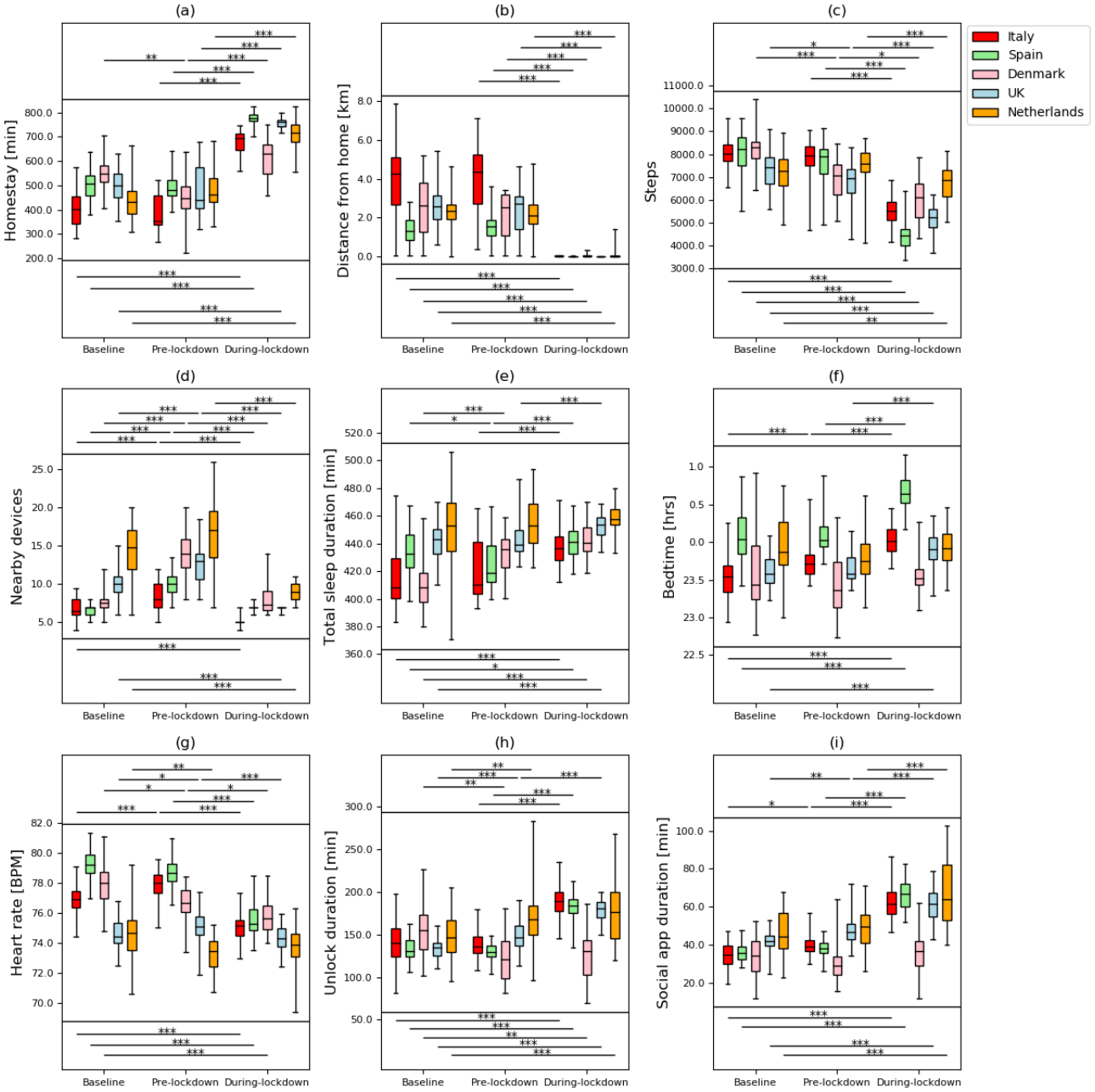
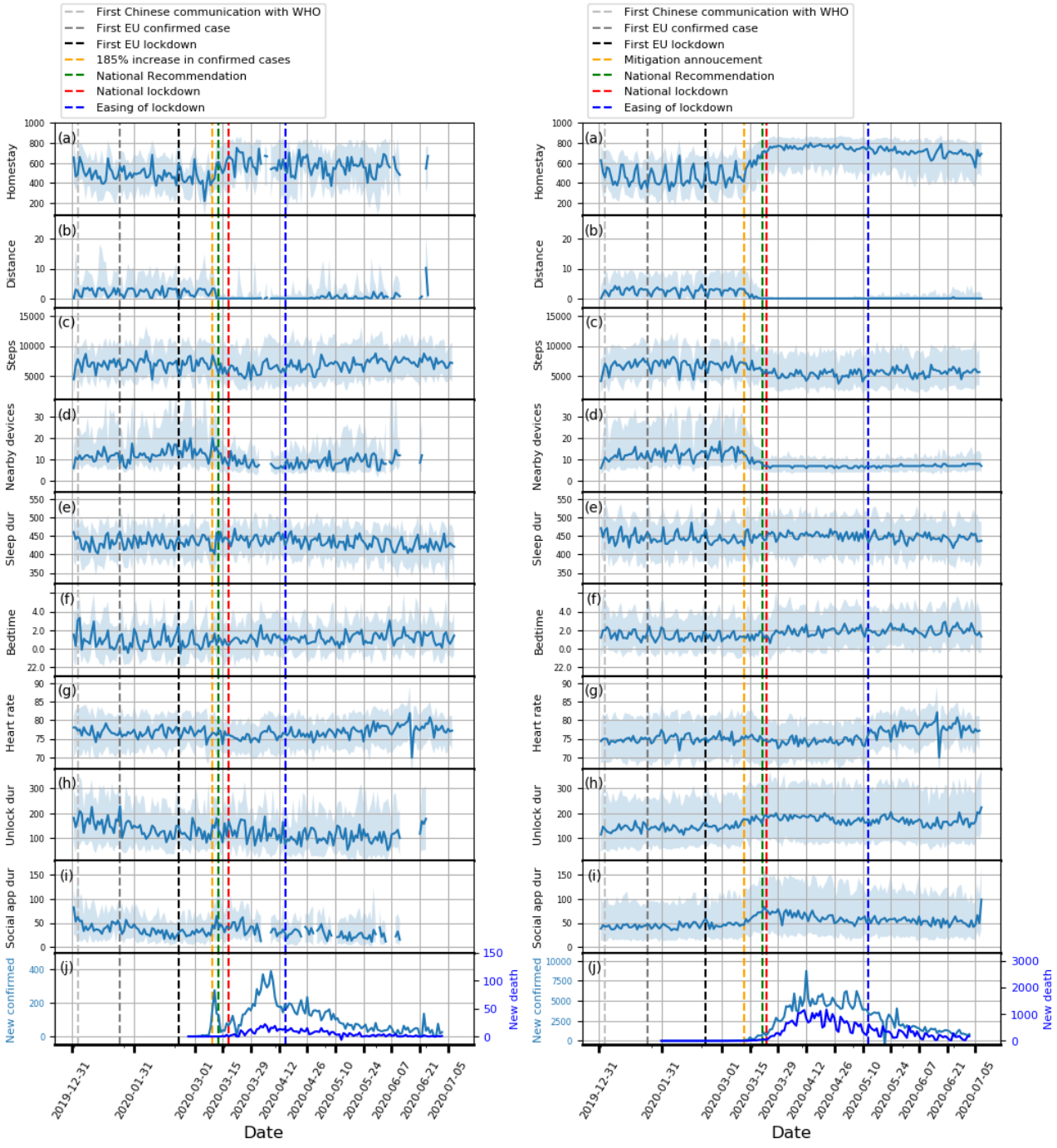


Table 2. Results of post hoc Dunn test (after Kruskal-Wallis tests) on the extracted features between prelockdown and during lockdown periods (only statistically significant differences were reported).

Features	Italy	Spain	Denmark	UK	Netherlands
Homestay					
z test	–9.38	–8.98	–5.44	–9.19	–7.33
P value	<.001	<.001	<.001	<.001	<.001
Maximum distance from home					
z test	9.0	8.91	5.48	8.40	7.58
P value	<.001	<.001	<.001	<.001	<.001
Steps					
z test	8.23	7.72	2.57	6.82	4.78
P value	<.001	<.001	.02	<.001	<.001
Maximum number of nearby devices					
z test	9.68	8.16	5.06	10.2	7.73
P value	<.001	<.001	<.001	<.001	<.001
Total sleep duration					
z test	–4.65	–5.23	—[a]	–4.24	—
P value	<.001	<.001	—	<.001	—
Bedtime					
z test	–4.31	–7.23	—	–5.28	—
P value	<.001	<.001	—	<.001	—
Heart rate					
z test	9.94	7.61	2.68	4.18	—
P value	<.001	<.001	.02	<.001	—
Unlock duration					
z test	–8.8	–8.57	—	–6.0	—
P value	<.001	<.001	—	<.001	—
Social app use duration					
z test	–7.72	–2.36	—	–5.72	–4.98
P value	<.001	<.001	—	<.001	<.001

[a]The difference was not statistically significant.

Figure 7. Zoomed in time series plots for Copenhagen, Denmark (left) and the United Kingdom (right). (a) homestay duration, (b) maximum distance from home, (c) Fitbit step count, (d) maximum number of nearby devices, (e) total sleep duration, (f) bedtime, (g) heart rate, (h) unlock duration, (i) social app duration, (j) COVID-19 confirmed and death cases. Solid line: median; shade: 25th percentile to 75th percentile. dur: duration; WHO: World Health Organization.



The differences across countries existed in the implemented NPIs as well. The requirement of staying at home except for essential trips and the cancellation of public events were implemented in all countries but Denmark where they were only recommended. Working places were required to close for some sectors in Spain, the United Kingdom, and Denmark, and were required to close for all but essential works in the Netherlands and Italy. Public transport was recommended to close in Italy, Spain, and Denmark. Among all countries, Spain had the least strict restrictions on gatherings and school closures (only geographically targeted).

We observed that the young group spent more time at home in Italy, Spain, and the United Kingdom, and degree holders spent more time at home in Italy and Denmark. The young group took fewer daily steps in Italy, the United Kingdom, and the Netherlands; the low BMI group took fewer daily steps in Italy, Spain, Denmark, and the United Kingdom; the young men group took fewer daily steps in Italy, the United Kingdom, and the Netherlands. Participants educated to degree level walked more in the United Kingdom and the Netherlands but less in Italy. The detailed results are presented in Table 3.

Table 3. Wilcoxon signed rank test results on the examined factors during lockdown periods (only statistically significant differences were reported).

Features and factors	Italy	Spain	Denmark	UK	Netherlands
Homestay					
Age					
w test	108	252	—[a]	3	—
P value	<.001	.003	—	<.001	—
Gender					
w test	—	—	—	—	—
P value	—	—	—	—	—
Degree					
w test	270	—	0	—	—
P value	<.001	—	.004	—	—
BMI					
w test	—	—	—	220	—
P value	—	—	—	<.001	—
Young men					
w test	—	—	—	—	—
P value	—	—	—	—	—
Step					
Age					
w test	283.5	—	—	285.5	151
P value	<.001	—	—	.004	<.001
Gender					
w test	67.5	277	—	—	—
P value	<.001	.005	—	—	—
Degree					
w test	76	—	—	6	0
P value	<.001	—	—	<.001	<.001
BMI					
w test	0	70	38	17	—
P value	<.001	<.001	.001	<.001	—
Young men					
w test	149	—	26	—	0
P value	<.001	—	<.001	—	.02

[a]The difference was not statistically significant.

Discussion

Principal Findings

We quantitatively investigated COVID-19 and associated lockdown-related changes in mobility, physiological measures, and phone use features derived from passive data collected through mobile devices (smartphones and wearable Fitbit devices) of participants recruited in five European countries to the RADAR-CNS program. We were able to measure significant changes in behavioral features between baseline, prelockdown, and during lockdown periods. As well as confirming expected changes such as spending more time at home, travelling much less, having far fewer nearby devices, we observed that people were more active on their phones, interacting with others through social apps particularly around major news events such as national lockdown, suggesting physical but maybe less social distancing. Furthermore, participants had lower heart rates, slept more, and went to bed later. In addition, we found that younger people spent more time at home and took fewer daily steps. Participants with lower BMI took more steps while maintaining comparable homestay with the higher BMI group. With 5 billion

global smartphone users and 500 million smartwatch and wearable device users [28,29], we propose that the ability to generate metrics such as these is vital for evaluating NPIs efficacy.

Our mobility analyses are in line with Google mobility reports [30], where substantial reductions in mobility and increase in residential stays during lockdown periods were found in Italy, Spain, and the United Kingdom; Denmark and the Netherlands by comparison showed an increase in mobility trends for parks and a relatively small increase in residential stays. However, in comparison to Google mobility reports, which provide valuable aggregated data for short periods, RADAR-base is an open-source highly configurable platform that supports collection and analysis of participant-level mobile and phone data in near real time with a potential for targeted interventions. Specifically, focused test and tracing may be directed to people perceived to be at high risk based on their behavior. In addition, RADAR-base was also used to collect self-reported questionnaires related to emotional well-being, functional status, and disease symptom severity of its participants [19]. Since April 2020, new questionnaires have been distributed to specifically assess COVID-19 symptoms and diagnosis status of our RADAR-CNS research participants. Our future work will use the entirety of these data to investigate the potential of wearable data, such as digital early warning signs of COVID-19, the impact of COVID-19 on the QoL, and the clinical trajectories of their primary diagnosis (MDD or MS).

The difference in the response across nations may reflect differences in the implementation of NPIs, media communication, and cultural differences. Denmark implemented stricter restrictions on working places and public transports but were less strict on homestay and public events [24]. In contrast, Spain was more flexible on restrictions of group gatherings and school closures. The contrast in the implementation of different NPIs between the two countries showing distinct behaviors during lockdown sheds light on which NPIs might be more productive in promoting physical distancing. This shows the potential utility of RADAR-base for remotely monitoring the effect of different NPIs, and we also saw evidence of this in our data with participants beginning to return to prelockdown routines as NPIs were lifted. Future work will compare these differences within a country and across countries, which may further elaborate on the effect and impact of NPIs on infection rates and potential second waves.

It is interesting to note that the younger group in general stayed at home more and took fewer steps than the older group. Since most countries required staying at home except for essential trips, one reason could be that the older adults, often less experienced in using online shopping, had to go out for groceries. This conjecture requires future work to investigate. Those educated to the degree level stayed at home for significantly longer in Italy and Denmark, possibly reflecting higher employment in sectors better able to work from home. The low BMI group took more steps but retained similar homestay to the high BMI group, which suggested they may have found other means to exercise locally. This information helps us understand the effectiveness of the NPIs at a subpopulation level and may be useful in informed strategies for targeted NPIs.

The ability to simultaneously manage multiple data modalities in RADAR-base facilitates the joint analysis and interpretation of them together with NPIs. The decrease in heart rate may be explained by the concurrent reduction in steps, the increase in homestay, and total sleep duration. The reduction in mobility, coupled with an increase in phone use, could possibly serve as indicators of physical distancing observance and resultant compensated social interaction. The delayed bedtime might be related to children homeschooling as a result of school closure, increased phone use, and a lack of exercise. As such, RADAR-base can also be applied to monitor the population health when jointly interpreting features such as step count, sleep duration, and bedtime, which is vital if the social distancing is implemented for a longer duration.

Finally, it has been shown that an elevated resting heart rate may suggest acute respiratory infections [31]. It may be possible to infer one's infection by continuously monitoring heart rate, especially when the population remains indoor for a vast majority of the time. Such monitoring provides the possibility to generate early warning signals for symptomatic or presymptomatic respiratory infections, thereby aiding timely self-isolation or treatment. The COVID-19–related symptom and diagnosis questionnaires have been added to the study and may provide a means to investigate these relationships further.

Media Effects

In addition to changes in trends over longer periods, we also identified interesting findings in relation to specific events (see Figure 7). A dramatic reduction in total sleep duration was observed in Denmark around March 11, 2020, which may be related to the announcement of the pending lockdown on that day and a 185% increase in the confirmed cases in Denmark on the previous day. Another example can be seen just after the mitigation phase was announced in the United Kingdom on March 12, in which social distancing was not strongly recommended, yet we saw participants isolating themselves voluntarily by staying at home for much longer. These observations highlight the potential role of media and social media in the distribution of information that may precipitate certain behavior. This observation may also explain the significant difference between the baseline and prelockdown phases, and suggests that people may have acted ahead of further government restriction. Furthermore, this is accompanied by a marked loss of week day and weekend periodic structure prelockdown and during the lockdown periods.

Limitations

There are some issues to consider concerning this work. First, the participants included in this study have different medical conditions (MDD or MS), which led to different baseline levels across countries. Nevertheless, as the focus of this study is the changes in the prelockdown and during lockdown phases relative to the baseline, we were still able to identify and compare the changes induced by lockdowns. We also analyzed the data collected in Spain split into MDD and MS separately and found the results only differed marginally. Understanding of any

artifacts or effects introduced into the RADAR-CNS data by the NPIs will be crucial in RADAR-CNS being able to deliver its aim of identifying signals that predict and prevent MDD and MS. Although the medical conditions of the population in this study might not be fully generalizable to a wider population healthy or with other conditions, we were able to demonstrate the utility of RADAR-base in monitoring behavioral changes, which can be readily generalized to other cohorts.

Second, the individual disease status at baseline may be different from that of the during lockdown period in each country, which might complicate the comparisons. To mitigate this, we used the same time period from the previous year to suppress the seasonal variability. We believe this complication on the population level was unlikely to be large, especially compared to the impact of lockdowns.

Third, participants recruited at different times may use different devices for smartphones and Fitbit depending on the availability and enrollment dates, which might make it difficult for interparticipant comparisons. Yet, this work focused on the population-level behavioral changes induced by NPIs where the handset variability was less of a factor.

Fourth, on account of requirements for participants' privacy in the RADAR-CNS studies, location data were purposely obfuscated with a participant-specific random value preventing precise localization of the participants, which limited use of regional geographic factors within a country. It would be interesting to examine how specific regions react to lockdowns when these data are available in future work.

Fifth, limited sample sizes in certain countries and data loss impacted the smoothness of the plots showing how the extracted features evolved over time. The plots for Denmark and the Netherlands showed relatively large variance particularly in the early phase, as these sites have only recently begun recruiting. Several dips and spikes in step counts and heart rate were seen in all countries during July and August. This observation was due to having data loss because of connectivity issues with the Fitbit server during this time.

Last, we only explored a subset of features that can be derived from smartphones and Fitbit wearable devices. Future work will investigate whether other features offer additional information for a more complete description of lifestyle changes.

Conclusions

Using participants' data from smartphones and wearable devices collected and managed by RADAR-base over 1.5 years covering the outbreak and subsequent spread of the COVID-19 pandemic across five European countries, we were able to detect and monitor the physical-behavioral and social-behavioral changes in response to the NPIs. We found that, as well as expected findings (that validated the data collection platform) relating to increased time spent at home, less travel, and fewer nearby Bluetooth-enabled devices, participants were more active on their phone, in particular, interacting with others using social apps, especially around major news events, suggesting increased physical distancing with socialising and interaction moving online. Furthermore, we found that participants had lower heart rates, slept more, and went to bed later. We demonstrated different responses across countries with Denmark showing attenuated responses to NPIs compared to other countries, which may be associated with their different focus of implementation NPIs. We found that younger people stayed at home for longer yet walked less compared to older adults and that the people with lower BMI remained more active during lockdown while having comparable homestay compared to their counterparts with higher BMI. Joint analysis of the extracted features is important for evaluating aspects of NPIs performance during their introduction and any subsequent relaxation of these measures. This work demonstrates the value of RADAR-base for collecting data from wearables and mobile technologies to understand the effect and response of public health interventions implemented in response to infectious outbreaks such as COVID-19. This ability to monitor response to interventions, in near real time, will be particularly important in understanding behavior as social distancing measures are relaxed as part of any COVID-19 exit strategy. Future work will include using participants' responses to COVID-19–related questionnaires together with an expanded feature set to gain more specific understandings into the relationship between mobile devices–derived features and the COVID-19 symptoms.

Acknowledgments

This study was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust, King's College London and EU/European Federation of Pharmaceutical Industries and Associations (EFPIA) IMI Joint Undertaking 2 (RADAR-CNS grant No 115902). This communication reflects the views of the RADAR-CNS consortium, and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. Participant recruitment in Amsterdam, the Netherlands was partially accomplished through Hersenonderzoek.nl, the Dutch online registry that facilitates participant recruitment for neuroscience studies [20]. Hersenonderzoek.nl is funded by ZonMw-Memorabel (project no. 73305095003), a project in the context of the Dutch Deltaplan Dementie, Gieskes-Strijbis Foundation, the Alzheimer's Society in the Netherlands (Alzheimer Nederland), and Brain Foundation Netherlands (Hersenstichting). This study has also received support from Health Data Research UK (funded by the UK Medical Research Council), Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust, and The NIHR University College London Hospitals Biomedical Research Centre.

Authors' Contributions

SS (Shaoxiong Sun), AAF, and RJBD contributed to the study design. SS (Shaoxiong Sun) contributed to the data analysis, figures drawing, and manuscript writing. AAF, NC, TW, VAN, GC, MH, and RJBD contributed to the critical revision of the manuscript. AAF, YR, ZR, PC, CS, and RJBD contributed to the platform design and implementation. AAF, IMG, AR, TW, VAN, GC, MH, and RJBD contributed to the administrative, technical, and clinical support of the study. FM, GDC, SS (Sara Simblett), LL, ALG, AZ, BWJHP, FL, SS (Sara Siddi), and JMH contributed to data collectionn.

Conflicts of Interest

VAN is an employee of Janssen Research & Development LLC and may own equity in the company.

References

1. Bai Y, Yao L, Wei T, Tian F, Jin D, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA 2020 Feb 21:1406-1407 [FREE Full text] [doi: 10.1001/jama.2020.2565] [Medline: 32083643]
2. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med 2020 Apr;8(4):420-422. [doi: 10.1016/s2213-2600(20)30076-x]
3. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020 Mar 13;27(2):taaa021 [FREE Full text] [doi: 10.1093/jtm/taaa021] [Medline: 32052846]
4. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-07-01]
5. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
6. Yu P, Zhu J, Zhang Z, Han Y. A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. J Infect Dis 2020 May 11;221(11):1757-1761 [FREE Full text] [doi: 10.1093/infdis/jiaa077] [Medline: 32067043]
7. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
8. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:A. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
9. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):taaa037 [FREE Full text] [doi: 10.1093/jtm/taaa037] [Medline: 32181488]
10. Lazzerini M, Putoto G. COVID-19 in Italy: momentous decisions and many uncertainties. Lancet Global Health 2020 May;8(5):e641-e642. [doi: 10.1016/s2214-109x(20)30110-8]
11. World economic outlook update. International Monetary Fund. 2020 Jun. URL: https://www.imf.org/~/media/Files/Publications/WEO/2020/Update/June/English/WEOENG202006.ashx?la=en [accessed 2020-07-18]
12. Pandian P, Mohanavelu K, Safeer K, Kotresh T, Shakunthala D, Gopal P, et al. Smart Vest: wearable multi-parameter remote physiological monitoring system. Med Eng Phys 2008 May;30(4):466-477. [doi: 10.1016/j.medengphy.2007.05.014] [Medline: 17869159]
13. Cancela J, Pastorino M, Tzallas A, Tsipouras M, Rigas G, Arredondo M, et al. Wearability assessment of a wearable system for Parkinson's disease remote monitoring based on a body area network of sensors. Sensors (Basel) 2014 Sep 16;14(9):17235-17255 [FREE Full text] [doi: 10.3390/s140917235] [Medline: 25230307]
14. Paradiso R, Alonso A, Cianflone D, Milsis A, Vavouras T, Malliopoulos C. Remote health monitoring with wearable non-invasive mobile system: The Healthwear project. In: IEEE. 2008 Oct 14 Presented at: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 20-25 Aug 2008; Vancouver, BC. [doi: 10.1109/iembs.2008.4649503]
15. RADAR-base. URL: https://radar-base.org/
16. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Hyve, RADAR-CNS Consortium. RADAR-Base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR mHealth uHealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
17. Stewart C, Rashid Z, Ranjan Y, Sun S, Dobson R, Folarin A. RADAR-base: major depressive disorder and epilepsy case studies. In: UbiComp '18: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. 2018 Oct Presented at: UbiComp '18: The 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing; October, 2018; Singapore, Singapore p. 1735-1743. [doi: 10.1145/3267305.3267540]
18. RADAR-CNS. URL: https://radar-cns.org/

19. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
20. hersenonderzoek.nl. URL: https://hersenonderzoek.nl/
21. Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD-96 Proceedings. 1996 Presented at: Second International Conference on Knowledge Discovery and Data Mining; 1996 Aug 2-4; Portland, OR p. 226-231.
22. Sleep logs. Fitbit SDK. URL: https://dev.fitbit.com/build/reference/web-api/sleep/ [accessed 2020-07-18]
23. Communication. Google Play. URL: https://play.google.com/store/apps/category/COMMUNICATION?hl=en_GB [accessed 2020-07-18]
24. Hale T, Petherick A, Phillips T, Webster S, Kira K, Angrist N. Coronavirus government response tracker. Blavatnik School of Government. 2020 Mar. URL: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker [accessed 2020-04-16]
25. Ostertagová E, Ostertag O, Kováč J. Methodology and application of the Kruskal-Wallis test. AMM 2014 Aug;611:115-120. [doi: 10.4028/www.scientific.net/amm.611.115]
26. Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. Stata J 2018 Nov 19;15(1):292-300. [doi: 10.1177/1536867x1501500117]
27. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Statistics 2001 Aug;29(4):1165-1188. [doi: 10.1214/aos/1013699998]
28. Number of mobile phone users worldwide from 2015 to 2020. Statista. URL: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/ [accessed 2020-07-18]
29. Number of connected wearable devices worldwide by region from 2015 to 2022. Statista. URL: https://www.statista.com/statistics/490231/wearable-devices-worldwide-by-region/ [accessed 2020-07-18]
30. COVID-19 community mobile reports. Google. URL: https://www.google.com/covid19/mobility/ [accessed 2020-07-18]
31. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. Lancet Digital Health 2020 Feb;2(2):e85-e93. [doi: 10.1016/s2589-7500(19)30222-5]

Abbreviations

EFPIA: European Federation of Pharmaceutical Industries and Associations
IMI2: Innovative Medicines Initiative
MDD: major depressive disorder
MS: multiple sclerosis
NIHR: National Institute for Health Research
NPI: nonpharmaceutical intervention
QoL: quality of life
RADAR: Remote Assessment of Disease and Relapse
RADAR-CNS: Remote Assessment of Disease and Relapse–Central Nervous System

Edited by G Eysenbach, G Fagherazzi; submitted 08.05.20; peer-reviewed by A Tsanas, A AL-Asadi; comments to author 22.06.20; revised version received 20.07.20; accepted 26.07.20; published 25.09.20.

Please cite as:
Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, Cummins N, Matcham F, Dalla Costa G, Simblett S, Leocani L, Lamers F, Sørensen PS, Buron M, Zabalza A, Guerrero Pérez AI, Penninx BWJH, Siddi S, Haro JM, Myin-Germeys I, Rintala A, Wykes T, Narayan VA, Comi G, Hotopf M, Dobson RJB, RADAR-CNS Consortium
Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19
J Med Internet Res 2020;22(9):e19992
URL: https://www.jmir.org/2020/9/e19992
doi:10.2196/19992
PMID:32877352

Original Paper

Intergroup Contact, COVID-19 News Consumption, and the Moderating Role of Digital Media Trust on Prejudice Toward Asians in the United States: Cross-Sectional Study

Jiun-Yi Tsai[1*], PhD; Joe Phua[2*], PhD; Shuya Pan[3], PhD; Chia-chen Yang[4], PhD

[1]School of Communication, Northern Arizona University, Flagstaff, AZ, United States

[2]Grady College of Journalism and Mass Communication, University of Georgia, Athens, GA, United States

[3]School of Journalism and Communication, Renmin University of China, Beijing, China

[4]School of Educational Foundations, Leadership and Aviation, Oklahoma State University, Stillwater, OK, United States

[*]these authors contributed equally

Corresponding Author:
Shuya Pan, PhD
School of Journalism and Communication
Renmin University of China
No. 59 Zhongguancun Street, Haidian District
Beijing, 100872
China
Phone: 86 10 62511081
Email: shuya@ruc.edu.cn

Abstract

Background: The perceived threat of a contagious virus may lead people to be distrustful of immigrants and out-groups. Since the COVID-19 outbreak, the salient politicized discourses of blaming Chinese people for spreading the virus have fueled over 2000 reports of anti-Asian racial incidents and hate crimes in the United States.

Objective: The study aims to investigate the relationships between news consumption, trust, intergroup contact, and prejudicial attitudes toward Asians and Asian Americans residing in the United States during the COVID-19 pandemic. We compare how traditional news, social media use, and biased news exposure cultivate racial attitudes, and the moderating role of media use and trust on prejudice against Asians is examined.

Methods: A cross-sectional study was completed in May 2020. A total of 430 US adults (mean age 36.75, SD 11.49 years; n=258, 60% male) participated in an online survey through Amazon's Mechanical Turk platform. Respondents answered questions related to traditional news exposure, social media use, perceived trust, and their top three news channels for staying informed about the novel coronavirus. In addition, intergroup contact and racial attitudes toward Asians were assessed. We performed hierarchical regression analyses to test the associations. Moderation effects were estimated using simple slopes testing with a 95% bootstrap confidence interval approach.

Results: Participants who identified as conservatives (β=.08, *P*=.02), had a personal infection history (β=.10, *P*=.004), and interacted with Asian people frequently in their daily lives (β=.46, *P*<.001) reported more negative attitudes toward Asians after controlling for sociodemographic variables. Relying more on traditional news media (β=.08, *P*=.04) and higher levels of trust in social media (β=.13, *P*=.007) were positively associated with prejudice against Asians. In contrast, consuming news from left-leaning outlets (β=–.15, *P*=.001) and neutral outlets (β=–.13, *P*=.003) was linked to less prejudicial attitudes toward Asians. Among those who had high trust in social media, exposure had a negative relationship with prejudice. At high levels of trust in digital websites and apps, frequent use was related to less unfavorable attitudes toward Asians.

Conclusions: Experiencing racial prejudice among the Asian population during a challenging pandemic can cause poor psychological outcomes and exacerbate health disparities. The results suggest that conservative ideology, personal infection history, frequency of intergroup contact, traditional news exposure, and trust in social media emerge as positive predictors of prejudice against Asians and Asian Americans, whereas people who get COVID-19 news from left-leaning and balanced outlets show less prejudice. For those who have more trust in social media and digital news, frequent use of these two sources is associated with lower levels of prejudice. Our findings highlight the need to reshape traditional news discourses and use social media and mobile news apps to develop credible messages for combating racial prejudice against Asians.

(J Med Internet Res 2020;22(9):e22767) doi:10.2196/22767

KEYWORDS

COVID-19; prejudice; news exposure; news trust; infodemic; media bias; racism; social media use; intergroup contact; regression; moderation analysis; cross-sectional survey

Introduction

Background

Since the COVID-19 outbreak, the United States has become an epicenter, surpassing 5 million confirmed cases as of August 2020. The behavioral immune system theory posits that people's tendency to avoid the risk of pathogen contagion is driven by cognitive and affective responses to informational cues [1-3]. Evidence supports that the perceived threat of a contagious virus will provoke strong aversive emotions, leading people to be distrustful of immigrants and out-groups [4,5]. Indeed, reports of anti-Asian racial incidents and hate crimes in the United States reached over 2000 since the term "Chinse virus" was used [6]. A national survey collected in March 2020 found 42% of US residents were likely to engage in discriminatory behaviors toward Asians because of their fear of the virus [7]. Relatedly, the Pew Research Center documented that 40% of Americans believed racial bias against Asians was more common than it was before the outbreak, and 31% of Asian Americans have experienced slurs or racist jokes since the pandemic [8].

Two psychological mechanisms are relevant in explaining the sharp increase in prejudice against out-groups when people cope with high levels of anxiety about infectious diseases. First, the contact hypothesis [9-11] postulates that positive face-to-face intergroup contact will improve out-group attitudes, contributing to prejudice reduction. Conversely, negative intergroup contact could elicit hostility toward out-groups. Because COVID-19 is transmitted through interpersonal interaction, the salience and magnitude of a perceived threat may result in either avoidance of Asians who are historically associated with diseases or hostile attitudes toward Asians [12].

Second, consuming information in news media and social media offers indirect forms of mediated or vicarious contact, thus shaping attitudes toward out-groups when people have minimal direct interaction with minority groups during shelter-in-place orders [13,14]. Early media coverage of the COVID-19 pandemic disproportionately focused on Chinese citizens' consumption of raw bats and other wild animals [15]. Particularly on the internet, conspiracy theories and racist scapegoating of China arose, with widely shared articles, social media posts, and viral videos blaming Chinese people, and Asians in general, for their "dirty" and "unsanitary" eating habits [16]. At the same time, the US government officials' politically consequential messages such as referring to the virus as Chinese virus were prevalent in national press briefings [17]. The constant coverage in mainstream news and sharing on social media likely heightened the salience of infection risk and the impression of Chinese and Asian people as a threat to health in the public's mind. As a majority of Americans have been closely following COVID-19 news from legacy media, local news, and social media [18-21], the prevalent rhetoric blaming China and the negative racial stereotypes can play an influential role in activating prejudice against Asians [22,23].

To date, little research has linked the two aforementioned mechanisms to bias toward stigmatized groups in the context of a global pandemic. Based on a cross-sectional online survey of 430 participants, we jointly examined two psychological mechanisms underlying the activation of prejudice against Asians and Asian Americans in the United States during the COVID-19 pandemic: intergroup contact and indirect contact of media consumption. To reflect the hyper-choice and hyper-partisan news environment, we differentiated three functionally different concepts of media consumption: media use, media trust, and media bias. Next, we compared how traditional news, social media use, and news apps cultivated racial attitudes. Last, this research lent empirical support to the moderating role of media use and trust in prejudice against Asians. Together, the timeliness of our results will yield theoretical and practical insights into developing public health interventions for reducing racial discrimination linked to the COVID-19 pandemic.

Intergroup Contact and Prejudice

Literature on the behavioral immune system has documented that humans' subjective risk perceptions of contagious diseases will trigger a series of cognitive, affective, and behavior mechanisms to avoid pathogen infection [1,2]. The tendency of pathogen avoidance is integral to coping with infectious diseases. The salience of disease threats is often activated by situational cues such as media reports of virus outbreaks and disgusting images, thereby triggering pathogen-avoidant cognitions and behavioral avoidance [3]. Empirical evidence shows that salient disease cues increase the tendency to avoid people who are potentially infectious [24]. Moreover, the activation of the behavior immune activity could motivate aversion and prejudicial attitudes toward out-groups, especially minority members historically associated with dirtiness and diseases [24]. In sum, disease threats can sensitize people to risks of intergroup contact with visibly stigmatized members.

Despite a lower fatality rate than Middle East respiratory syndrome (MERS) and Ebola, the SARS-CoV-2 virus has caused one of the highest death tolls worldwide, and there is still no vaccine available [25]. As the surges of confirmed COVID-19 cases continue to dominate US media attention, the salience and magnitude of the perceived threat is heightened. The sensitivity of intergroup risks activated by constant rhetorical cues of blaming Chinese people can possibly explain an increase in prejudice against Asians. Through three online experiments, Huang et al [4] found that participants felt more negative about immigrants after reading a news report about the 2009 swine flu health risks and shortage of vaccine supply. Kim et al [26] conducted a national sample survey of 1000 American adults during the 2014 Ebola outbreak. Findings

concluded that vulnerable people had more generalized xenophobia. Given that emotional attitudes toward out-groups are more closely related to subsequent discriminatory behaviors, we focus on affective feelings about Asians as a proxy for measuring prejudice [27].

Role of Media Use, Trust, and Source Bias in Out-Group Attitudes

In addition to serving as situational cues to activate aversion of minority groups, news reports offer indirect mediated contact that possibly shapes prejudice toward ethnic minorities [13,14]. Different from prior research, we differentiate three functionally different concepts of news consumption: modalities of media use, media trust, and media bias. We further offer an overview of how each concept relates to negative attitudes toward out-groups.

Decades of cultivation theory and news framing research have proved the direct influence of using traditional media, mostly newspapers and TV newscasts, on cultivating audiences' prejudicial views of out-groups [28]. Persistent stereotypical media coverage of a particular racial group would result in prejudicial beliefs toward individuals of such group. For instance, Dixon [29] found that network TV news exposure resulted in more stereotypical views and negative perceptions of African Americans. Another study [30] examined racialized news framing of President Barack Obama, suggesting that exposure to negative frames about Obama activated underlying prejudicial beliefs and biased evaluations of African Americans in general. Likewise, negative TV portrayals of out-groups such as refugees increased prejudicial attitudes toward refugees based on experimental designs [14]. Therefore, the traditional news coverage of politicized COVID-19 discourses placing blame on China, along with the labeling of the disease as a "Chinese virus" or a "Wuhan virus," might have a negative impact on racial perceptions and stereotypes of Asian Americans and Asians in general, leading to prejudice against them [22,23].

Although the aforementioned studies are insightful, relying on one average scale to measure individuals' exposure to traditional media outlets (eg, print newspaper and TV) limits our understanding of how American audiences receive information from hyper-choice, hyper-partisan, networked news environments [13,14,23,29]. The average index approach assumes that modalities of news media consumption associate out-group attitudes equally. However, it is conceivable that different information channels emphasize distinct coverage of the pandemic and differ in the prevalence of anti-Asian discourses. Moreover, it remains unclear how the use of social media and digital outlets might be related to differential attitudes toward out-groups.

In addition to the well-documented influence of traditional media exposure on prejudice, the role of social media and digital outlets in shaping users' views of ethnic minorities cannot be overlooked. Social media was and is an essential source for coping with the MERS outbreak and the ongoing pandemic [31-33]. At the same time, political leaders' antiminority sentiment tended to circulate widely on social media platforms, contributing to an increase of hate crimes toward minority groups such as Muslims [34]. Although social media provides real-time updates and personally relevant feeds that might reinforce audience's pre-existing biases, users are more likely to be susceptible to misinformation and racially offensive comments [16,20]. Analysis of 2.8 million COVID-19–related English tweets from February and March 2020 revealed that one of the top topics involved racist attacks and rude comments against East Asians [21]. Budhwani and Sun [35] found that tweets containing "Chinese virus" or "China virus" increased ten-fold after political leaders used the insensitive terms on Twitter. However, findings on linking social media use with attitudes toward stereotyped groups remain inconclusive. One study surveying US college students revealed no significant relationship between getting news from social media and prejudice toward undocumented immigrants [36]. In contrast, one recent online survey showed that believing in social media news was positively associated with the perceptions that American identity and the economic situation were threatened by the presence of Chinese people [37]. To address this gap, we will examine how stereotypes propagated on social media and online news may associate with prejudice toward Asians.

Media trust and media use are correlated but conceptually different [38]. A worldwide survey showed that traditional news exposure correlated positively with trust in TV and newspapers, whereas online news use was negatively associated with trust in traditional media [39]. Put differently, when people cope with the evolving situation of COVID-19, not all media types will be trusted equally and have a universal association with activating perceptions of racial stereotypes and prejudicial attitudes. In an increasingly fragmented and polarized media environment, we argue that media use and trust in each information source should be measured separately to generate valid conclusions about their distinct relationships with prejudice. To our knowledge, scant literature has investigated media use and trust simultaneously, nor have they probed the moderation role of viewers' trust in each medium in the exposure-prejudice relationship. Since the outbreak, the use of TV news, online media, and social media sources grew substantially. At the same time, US public trust in news has hit a new low point. Only 29% of American showed trust in news overall in 2020, a significant decrease of 9% compared to 2017, and 14% of them trusted news from social media [40]. The public might trust legacy media more than social media because of journalists' gatekeeping function and reporting accuracy [17,19,20]. Considering the polarization in US newspaper and TV network coverage of COVID-19 [41] and the partisan nature of news trust [42], exploring the moderating role of media trust in determining the direction of the influence of media use on prejudicial attitudes is vital.

Last, relying on conservative or liberal media could also be a significant determinant of negative attitudes toward out-groups. An online survey with 422 Italians demonstrated a positive relationship between exposure to right wing newspapers and TV newscasts, and prejudice against immigrants after controlling for participants' political ideology. Conversely, using left-leaning news outlets showed a negative association with prejudice [43].

Guided by the behavioral immune system theory [1-5], associated outcomes of prejudice toward out-group members

[24-26], and the intergroup contact hypothesis (H) [9-12], we expect that salient perceived risks of contracting COVID-19 when interacting with Asians will positively be related to prejudice.

- H1: The frequency of direct intergroup contact will be positively associated with prejudicial attitudes toward Asians.

As no prior research distinguishes how modalities of news, news trust, and media bias relate to virus-activated prejudice, we propose the following research questions (RQs):

- RQ1: How will media use, media trust, and media bias be associated with prejudicial attitudes toward Asians?
- RQ2: How will media trust moderate the relationship between media use and prejudicial attitudes toward Asians?

Methods

Participants and Procedure

In May 2020, we recruited participants from Amazon's Mechanical Turk (Mturk) program, a dominant internet-based platform for gathering online samples in behavioral science fields [44]. A short description containing the research title, keywords, and expected completion time was generated in a coauthor's Mturk account. Interested participants clicked on an external link that rerouted them to take the Qualtrics survey. Upon completion, the Mturk program generated a random code for participants to receive compensation. The Mturk subject pools demonstrate proven advantages of being more attentive than collegiate populations [44]. Although the representativeness of Mturk samples raise some concerns, Levay et al [45] concluded that Mturk respondents' attitudes toward social issues did not differ fundamentally from the random sample of the American National Election Studies after controlling for key demographics. Similarly, Mturk samples can generate data quality comparable to representative samples [46]. Moreover, Mturk's recruitment speed is excellent for study designs that depend on current social events [47]. For instance, Park et al [48] used Mturk samples to study American adults' information channel preferences during the 2016 Zika virus outbreak. Given that confirmed cases in the United States has surged since March 2020, Mturk samples are suitable for our objectives because of the ability to collect how individuals relied on legacy news and social media to cope with the global pandemic in a naturalistic setting.

To determine the required sample size to achieve desired statistical power, we performed an a priori estimate using G*Power version 3.1 (Heinrich-Heine-Universität Düsseldorf) before data collection [49]. A medium regression effect size (R^2=0.13) with 99% power and 18 predictors indicated a minimum sample of 275 [50]. To ensure the quality of study results, we limited the sample to people residing in the United States. Only participants who passed attention check questions were included in the analysis, yielding a final sample of 430 (mean age 36.75, SD 11.49 years; n=258, 60% male). The majority of participants were White (n=344, 80%), married (n=313, 72.8%), employed full-time (n=318, 74%), and had received Bachelor's degrees (n=267, 62.1%), and 11.4% (n=49) reported having been infected by COVID-19. More than half of the participants (n=224, 52.1%) indicated their family income has been impacted by the outbreak. Table 1 summarizes demographic characteristics of the sample.

Table 1. Demographics of survey participants (N=430).

Variables	Participants
Age (years), mean (SD)	36.75 (11.49)
Male, n (%)	258 (60.0)
Education, n (%)	
High school or less	15 (3.5)
Some college	50 (11.6)
Bachelor's degree	267 (62.1)
Postgraduate	98 (22.8)
Race[a], n (%)	
White	344 (80.0)
Black or African American	52 (12.1)
Hispanic and Latino	34 (7.9)
Other	9 (2.1)
Marital status, n (%)	
Married or domestic partnership	313 (72.8)
Single	100 (23.3)
Other	17 (3.9)
Employment, n (%)	
Employed, ≥40 hours per week	318 (74)
Employed, <40 hours per week	77 (17.9)
Other	35 (8.1)
Family income impacted, n (%)	224 (52.1)
Political ideology, mean (SD)	4.08 (2.13)
Personal infection of COVID-19, n (%)	49 (11.4)

[a]Participants could select one or more self-identified races.

The Institutional Review Board office of Renmin University of China approved the research protocol, and responses were collected via Qualtrics software (Qualtrics International Inc). The first author at a US university in the southwest was provided with an anonymized data set containing no identifiable private information connected to participants. Ethics exemptions were obtained at coauthors' institutions before launching the survey. At the beginning of the survey, the respondents agreed to participate in the research voluntarily for receiving compensation. Next, respondents answered questions related to traditional news exposure, social media use, perceived trust, and their top three news channels for staying informed about the novel coronavirus. In addition, the frequency of intergroup contact and racial attitudes toward Asians were assessed. The survey took approximately 16 minutes to complete (mean 15.95, SD 39.34 minutes).

Measures

Intergroup Contact

We used a common measure [13,37,51] to assess the frequency of direct intergroup contact on a 7-point scale: "How much contact do you have with Asian people (a) at work; (b) as neighbors; (c) as close friends?" (1=none at all, 7=a great deal; Cronbach α=.88, mean 4.11, SD 1.63).

Media Use

Respondents rated how often they get news about COVID-19 from five traditional sources including print newspaper or magazine, radio, local TV, national network television, and cable TV on a 4-point scale with endpoints labeled "never" and "often" (Cronbach α=.72, mean 2.80, SD 0.64). Two single 4-point scales measured social media use (mean 3.12, SD 0.88) and news websites or app use (mean 3.10, SD 0.83).

Media Trust

To gauge different levels of media trust in news sources, we asked participants to indicate how much they trust information about COVID-19 from the aforementioned media outlets (1=not at all, 7=very much) [38]. Trust in traditional media demonstrated good internal consistency (Cronbach α=.86, mean 4.83, SD 1.20). Two single items were measured: social media trust (mean 4.72, SD 1.67) and trust in websites or apps (mean 5.20, SD 1.26).

Media Bias

Exposure to biased news sources was assessed by asking respondents to select the top three news channels they frequently visited for gathering COVID-19 news. We obtained a list of 31 popular news brands across a wide range of platforms from the Pew Research Center's survey of US media polarization and the 2020 election [52]. We then determined the political leanings of each news channel based on ratings provided by AllSides [53]. AllSides employs multiple methods to identify comprehensive media bias ratings for 600 media outlets. Next, we used a 5-point scale to represent each source's ideological placement of left (-2), lean left (-1), center (0), lean right (+1) and right (+2). Exposure to left-leaning sources was calculated by taking the absolute value of the sum of selected channels' bias ratings (mean 1.62, SD 1.05, range 0-5). Exposure to centrist news was calculated by the raw counts that neutral or balanced outlets were selected by each participant (mean 0.73, SD 0.71, range 0-3). Lastly, exposure to right-leaning media was measured by adding up selected sources' bias scores (mean 0.70, SD 0.87, range 0-4)

Prejudicial Attitudes

We modified a measure from [54] to assess prejudicial attitudes toward Asians. The question was rephrased as "Since the COVID-19 outbreak, how much have you felt the following toward Asian people in general?" Respondents indicated, from 1 "not at all" to 7 "very much," seven negative feelings: (1) jealousy, (2) anger, (3) resentment, (4) discomfort, (5) hatred, (6) despise, and (7) shame. These items formed a reliable scale with higher values representing higher levels of prejudice toward Asians (Cronbach α=.97, mean 3.61, SD 1.83). The phrasing of "since the COVID-19 outbreak" was included to set a standard time frame as a basis for participants to respond to news consumption and disease-linked attitudes. Using specific and concrete wording is a recommended practice for survey designs [55]. Specifying a consistent time frame allowed us to pinpoint the relationships between people's exposure and trust in COVID-19 news and negative feelings toward Asians during the pandemic.

Controls

As news consumption is associated with sociodemographic variables, we measured common characteristics for statistical control [13,37,43]. Respondents reported demographic characteristics including age, gender (1=male, 0=female), educational attainment, ethnicity (1=White, 0=others), marital status (1=married or domestic partnership, 0=others), and employment status (1=full-time, 0=others). Personal infection of COVID-19 was measured with one dichotomous item (1=yes, 0=no). Lastly, we measured political ideology on a 7-point scale from 1 "very liberal" to 7 "very conservative" (mean 4.08, SD 2.13).

Statistical Analysis

Table 2 summarizes bivariate correlations of continuous variables. Next, we performed four-step hierarchical linear regression analyses using SPSS version 26 (IBM Corp) to predict prejudicial attitudes toward Asians since the COVID-19 spread. All variance inflation factor scores among predictor variables were less than 3, indicating there was no multicollinearity. The first models included respondents' demographic characteristics and history of personal infection as controls. In the second model, we investigated the main effect of direct intergroup contact to test H1. To answer RQ1, media use, trust in each channel, and exposure to news sources with political leanings were entered in the third model. To explore RQ2, we used mean-centered media use and mean-centered trust scores to compute three separate interaction terms, which were then included in the fourth model. Interaction terms were created by multiplying use of a given medium by trust in the given medium. Subsequently, we conducted significant tests of simple slopes with a 95% bootstrap confidence interval approach in Mplus version 8 (Muthén & Muthén). Moderation effects were tested by creating 5000 bootstrap samples. Such an approach is recommended because it leads to less-inflated type I error rates and provides coverage of confidence intervals [56,57].

Table 2. Bivariate correlations with *P* values.

Variables	1	2	3	4	5	6	7	8	9	10	11
1. Intergroup contact											
r	1	0.44	0.11	0.06	0.42	0.48	0.31	–0.12	0.03	–0.01	0.64
P value	—[a]	<.001	.03	.24	<.001	<.001	<.001	.01	.60	.84	<.001
2. Traditional media use											
r	0.44	1	0.38	0.21	0.66	0.50	0.40	–0.03	0.01	0.04	0.34
P value	<.001	—	<.001	<.001	<.001	<.001	<.001	.61	.86	.46	<.001
3. Social media use											
r	0.11	0.38	1	0.16	0.25	0.45	0.20	0.08	–0.09	–0.06	0.07
P value	.03	<.001	—	.001	<.001	<.001	<.001	.10	.06	.24	.16
4. Websites or apps use											
r	0.06	0.21	0.16	1	0.21	0.07	0.30	0.10	–0.02	0.02	–0.04
P value	.24	<.001	.001	—	<.001	<.001	<.001	.048	.72	.72	.40
5. Traditional media trust											
r	0.42	0.66	0.25	0.21	1	0.51	0.56	0.03	0.05	–0.13	0.31
P value	<.001	<.001	<.001	<.001	—	<.001	<.001	.53	.28	.008	<.001
6. Social media trust											
r	0.48	0.50	0.45	0.07	0.51	1	0.44	–0.06	–0.03	–0.02	0.43
P value	<.001	<.001	<.001	<.001	<.001	—	<.001	.23	.53	.74	<.001
7. Websites or app trust											
r	0.31	0.40	0.20	0.30	0.56	0.44	1	0.05	0.03	–0.09	0.21
P value	<.001	<.001	<.001	<.001	<.001	<.001	—	.29	.54	.06	<.001
8. Left-leaning											
r	–0.12	–0.03	0.08	0.10	0.03	–0.06	0.05	1	–0.45	–0.37	–0.18
P value	.01	.61	.10	.048	.53	.23	.29	—	<.001	<.001	<.001
9. Centrist											
r	0.03	0.01	–0.09	–0.02	0.05	–0.03	0.03	–0.45	1	–0.21	–0.07
P value	.60	.86	.06	.72	.28	.53	.54	<.001	—	<.001	.15
10. Right-leaning											
r	–0.01	0.04	–0.06	0.02	–0.13	–0.02	–0.09	–0.37	–0.21	1	0.05
P value	.84	.46	.24	.72	.008	.74	.06	<.001	<.001	—	.30
11. Prejudice											
r	0.64	0.34	0.07	–0.04	0.31	0.43	0.21	–0.18	–0.07	0.05	1
P value	<.001	<.001	.16	.40	<.001	<.001	<.001	<.001	.15	.30	—

[a]Not applicable.

Results

Table 3 presents the results of the hierarchical regression models predicting prejudicial attitudes toward Asians. Demographic variables and COVID-19 infection history significantly explained 25% of the variance in prejudice ($F_{8,421}$=17.65, *P*<.001). Specifically, participants who were married (β=.13, *P*=.001), held conservative beliefs (β=.08, *P*=.02), and had a personal infection (β=.10, *P*=.004) reported more negative attitudes toward Asians.

Table 3. Hierarchical regression analysis predicting prejudicial attitudes toward Asians since COVID-19 (N=430).

Variables	Model 1[a], β	*P* value	Model 2, β	*P* value	Model 3[b], β	*P* value	Model 4[b], β	*P* value
Age	–.15	.001	–.06	.11	–.05	.18	–.02	.54
Male	.01	.87	–.06	.11	–.06	.10	–.05	.19
Education	.15	.001	.04	.29	.04	.29	.02	.60
White	–.06	.14	–.02	.62	–.03	.46	–.03	.40
Married	.27	<.001	.18	<.001	.14	<.001	.13	.001
Employed full-time	.06	.15	.01	.73	–.01	.83	–.01	.90
Political ideology	.14	.001	.09	.01	.08	.04	.08	.02
Personal infection	.23	<.001	.13	<.001	.12	.001	.10	.004
Intergroup contact	N/A[c]	N/A	.55	<.001	.47	<.001	.46	<.001
Media use								
Traditional media	N/A	N/A	N/A	N/A	.09	.04	.08	.04
Social media	N/A	N/A	N/A	N/A	–.05	.16	–.06	.06
Websites or apps	N/A	N/A	N/A	N/A	–.05	.16	–.06	.13
Media trust								
Traditional media	N/A	N/A	N/A	N/A	.03	.59	.05	.36
Social media	N/A	N/A	N/A	N/A	.14	.005	.13	.007
Websites or apps	N/A	N/A	N/A	N/A	–.02	.60	–.05	.26
Media sources								
Left-leaning	N/A	N/A	N/A	N/A	–.16	.001	–.15	.001
Centrist	N/A	N/A	N/A	N/A	–.14	.003	–.13	.003
Right-leaning	N/A	N/A	N/A	N/A	–.04	.35	–.03	.52
Interaction								
Traditional media use * trust	N/A	N/A	N/A	N/A	N/A	N/A	–.02	.57
Social media use * trust	N/A	N/A	N/A	N/A	N/A	N/A	–.12	.003
Websites/apps use * trust	N/A	N/A	N/A	N/A	N/A	N/A	–.13	<.001
R^2	0.25	<.001	0.49	<.001	0.53	<.001	0.56	<.001
R^2 change	N/A	N/A	0.24	<.001	0.04	<.001	0.03	<.001

[a]Standardized beta coefficients (β) are reported.

[b]In models 3 and 4, scores of media use and trust are mean-centered.

[c]N/A: not applicable.

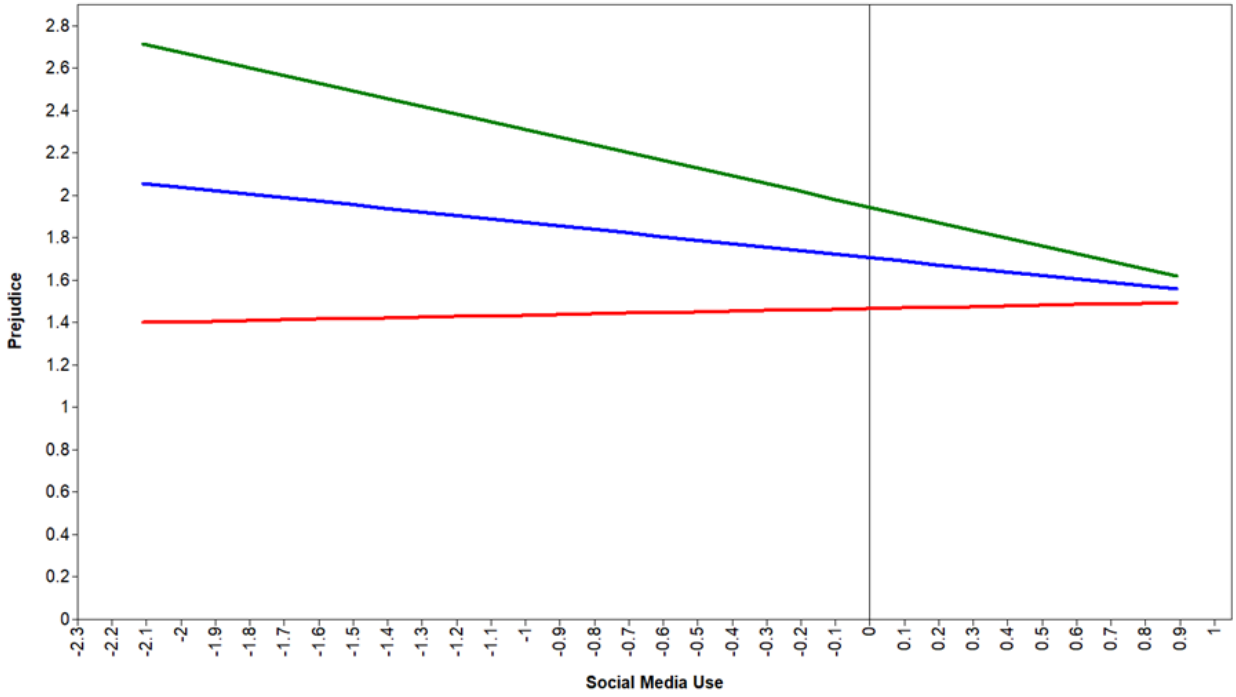
Turning to the first hypothesis, we found a positive association between intergroup contact and prejudice (β=.46, *P*<.001) after controlling for demographic characteristics, supporting H1. Respondents who interacted with Asian people frequently in their daily lives were more inclined to hold unfavorable attitudes toward Asians.

For RQ1, media use, trust, and reliance on biased sources showed significantly divergent relationships with prejudicial attitudes toward Asians (R^2 change 0.04, $F_{18,411}$=25.23, *P*<.001). Reading print newspapers and watching TV newscasts about COVID-19 was positively associated with higher levels of prejudice (β=.08, *P*=.04). Regarding the main effects of media trust, the results suggested that the more people trusted news circulated on social media, the higher their prejudicial attitudes were (β=.13, *P*=.007). Trust in traditional media and digital channels had no significant associations with negative feelings about Asians. Notably, relying on liberal media (β=–.15, *P*=.001) and balanced news sources (β=–.13, *P*=.003) for staying informed about the COVID-19 epidemic was significantly associated with lower levels of negative attitudes toward Asians.

The second RQ explored whether news trust moderated the relationship between media use and prejudice against Asians. Interaction terms explained a 3% increase in variance in prejudice (R^2 change 0.03, $F_{21,408}$=24.33, *P*<.001). The interaction term of trust and social media use was statistically significant (β=–.12, *P*=.003). We also found a similar significant

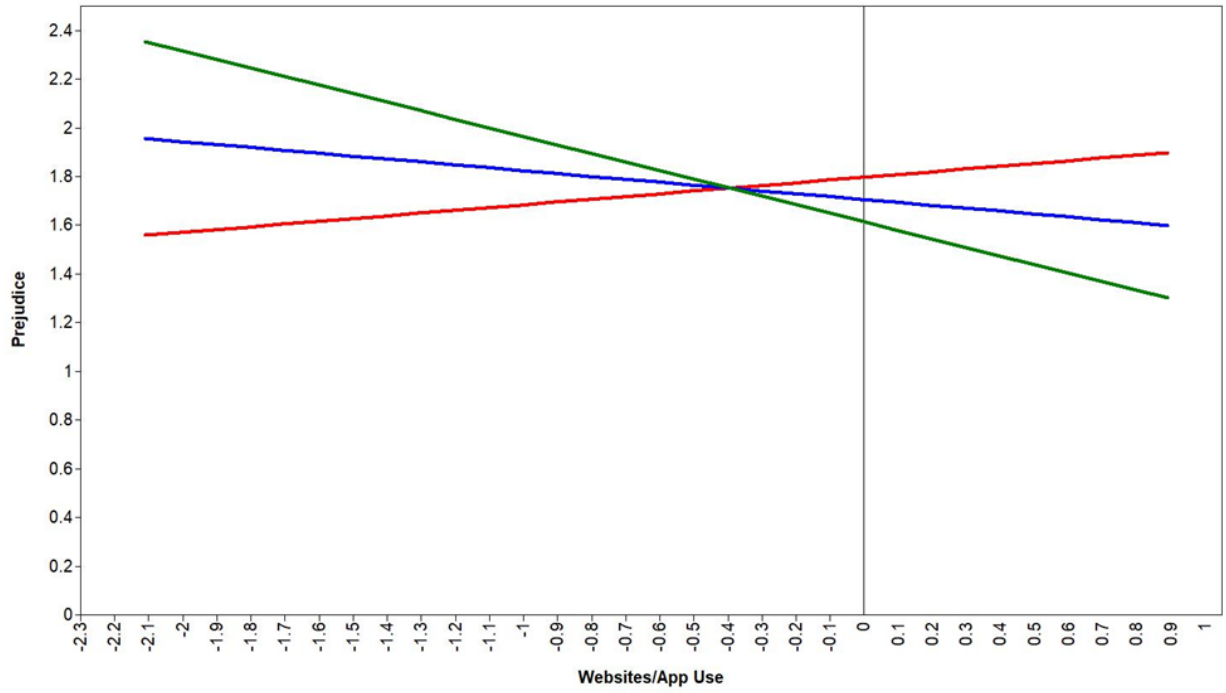
interaction effect of websites and apps trust and use on prejudice (β=–.13, *P*<.001). To explore how variation in media trust changes the relationship between use of social media or digital news and prejudice, we conducted simple slopes analyses. As depicted in Figure 1, the negative relationship between social media use and prejudicial attitudes was significantly greater among participants who had higher levels of trust in social media news, when compared to those with medium and low trust in social media (unstandardized simple slope –0.36, 95% CI –0.61 to –0.12). Additionally, Figure 2 illustrates that, at high levels of trust in online websites and mobile apps, frequent use of digital media was related to less unfavorable attitudes toward Asians (unstandardized simple slope for high trust –0.35, 95% CI –0.56 to –0.14).

Figure 1. Interaction plot showing predicted values of prejudicial attitudes toward Asians as a function of social media use and trust in social media.



Note: Because social media use scores are mean-centered, zero on the x-axis represents mean.

Figure 2. Interaction plot showing predicted values of prejudicial attitudes toward Asians as a function of websites/apps use and trust in websites/apps.



Note: Because websites/apps use scores are mean-centered, zero on the x-axis represents mean.

Discussion

Principal Findings

Drawing from the behavioral immune system theory, intergroup relations, and media psychology literature, this study is a novel attempt to investigate the joint mechanisms of direct intergroup contact and mediated contact offered by diverse news sources on associating with prejudicial attitudes toward Asians during the COVID-19 outbreak. In accordance with previous studies [1-5,7,9-13,24,26], participants who subscribed to conservative beliefs, had personal COVID-19 infection history, and interacted with Asian people frequently in their daily lives reported more negative attitudes toward Asians after controlling for sociodemographic variables. It might be plausible that people with conservative beliefs are more wary of China's influence, making them more likely to blame visible targets with Asian-looking features during the pandemic. Importantly, the medium effect size between intergroup contact and prejudice underscores the role of direct social interaction in relating to people's out-group attitudes triggered by pandemic threats. As expected, human tendency to avoid contagious diseases will stimulate aversive emotions when interacting with minorities explicitly associated with spreading COVID-19. Frequent contact with stigmatized Asian members during the pandemic might increase perceived risks of contracting the virus and, thus, link to negative attitudes.

We empirically examined the associations between three forms of indirect mediated contact and prejudice. The findings reveal the divergent role of news exposure, trust, and political-leaning media in racial attitudes when political orientation and intergroup contact are controlled for. Relying more on traditional news media was positively related to prejudice toward Asians, supporting the TV news' capability to activate or reinforce prevalent racial stereotypes [28-30]. In line with a current report [40], the survey participants watched more national, local, and cable television news rather than receiving information from print media. The positive association can be mostly driven by the fact that TV newscasts live streaming the President's daily press briefings increased viewers' firsthand exposure to the China-blaming discourses. Participants tend to hold unfavorable attitudes toward negatively portrayed groups on TV because repeated exposure to such politicized messages makes accessible the audiences' cognitive schemas of associating Asians or Chinese immigrants with disease threats during an unfamiliar pandemic [41,58].

Regression results suggest habitual selections of news outlets with ideological leanings are a more driven factor than the modalities of media in associating with disease-activated racial attitudes. Consuming news from left-leaning and neutral outlets was associated with less prejudicial attitudes toward Asians holding other factors constant. Journalistic powerhouses with liberal and centrist perspectives are more critical of the current administration's effort in curbing the pandemic and attribute the blame to uncoordinated policy responses at all levels of governments. Although conservative news such as Fox and liberal outlets such as CNN shared similar keyword use of mentioning China and the novel virus, a recent study found CNN news stories featured more scientific statements from Dr Anthony Fauci and Dr Deborah Birx, and emphasized more protective responses such as social distancing and lockdowns than their conservative counterparts did [59]. Liberal and centrist media's emphasis on scientific evidence and prevention strategies possibly explains why those relying on these sources for COVID-19 news hold less prejudice and perceived threats by the presence of Asians.

It is worth highlighting the positive link between social media trust and prejudice. Trust in social media was positively associated with prejudicial attitudes toward Asians. The nature of filter algorithms and selective exposure patterns might afford social media users to be exposed to or spread unfiltered discriminating language about Asians. Using social media as news sources increases users' probability to encounter trending anti-Asian sentiment, even when users do not actively seek such information. People tend to trust information created and shared by friends and family members within homogenous networks [60]. Recent studies [21,35,61] found that hate speech and racist attacks against Asians sharply increased on Twitter after the World Health Organization recognized COVID-19 as a global pandemic in mid-March. Common tweet keywords associated with China were eating habits, animals, virus, blame, and cause, echoing the China-blaming discourses [35,61]. Therefore, the more people trust social media (and thus the stigmatizing comments in this channel), the more likely they would perceive Asian people as the major responsible party for spreading the virus, thereby triggering negative attitudes.

Notably, this study contributes to existing literature by identifying trust in social media and digital websites and apps as a boundary condition of the news exposure-prejudice relationship. Among those who used social media infrequently, those who trusted social media more reported higher prejudice against Asians than those who trusted social media less. However, this difference became less pronounced among heavy users. Likewise, for those who had high levels of trust in digital websites and apps, frequent use was related to less unfavorable attitudes toward Asians. It could be possible that heavy users of social media and digital news have advanced media literacy of filter algorithms, equipping them with higher efficacy to move beyond personalized feeds and search for less politicized information and diverse perspectives. In other words, frequent use of social media and digital sources buffers the relationship between trust in social media and digital apps, and prejudice linked to COVID-19. Together, these results suggest that researchers should consider using nuanced measures of media exposure, media trust, and media bias to generate more accurate conclusions. In a hyper-choice networked news environment, Americans' use of and trust in traditional media, social media, and well-known news brands with distinct political leanings have more complex relationships with their disease-ignited prejudice toward minority groups than previously assumed.

Limitations and Future Directions

We acknowledge limitations of this study. First, the self-selected sample might reflect respondents' heightened concerns about COVID-19–related topics. In addition, the cross-sectional study did not guarantee the causal effects of intergroup contact and news use on influencing racial attitudes. Hence, it is possible that participants who held negative views about the Chinese government or Chinese people selected to consume anti-Asian coverage on traditional outlets and social media because those news reports confirmed their existing viewpoints. Given our correlational findings, longitudinal studies will expand upon the results to establish the causal relationships. Second, the study did not directly measure the prevalence of anti-Asian discourses circulated via all media outlets since the major outbreak in the United States. However, a recent study analyzing Google Trends data between December 2019 to March 2020 showed that the "Chinese virus" rhetoric has led to a sharp increase in search rates for anti-Chinese slurs. Therefore, we can confidently infer that an increase in search rates reflected trending news reports and social media discussion related to anti-Asian sentiment and racist attacks [16].

First, in light of these limitations, future research should content analyze the prevalence of anti-Asian sentiment on different types of media and correlate with public attitudes to replicate these findings. Second, as citizens can access the same news content from multiple modalities, conventional categorizations of media use might not capture a full picture of people's news consumption. A granular examination of website-tracking data to pinpoint what types of news coverage individuals frequently visit for coping with the pandemic will be informative. Last, future inquiries can compare the influence of negative portrayals of Asians circulated through traditional media and social networking sites on out-group attitudes and behavioral avoidance using experimental designs.

Conclusions

Experiencing racial discrimination among the Asian population during a challenging pandemic could cause poor psychological outcomes and exacerbate health disparities [62]. The results suggest that conservative ideology, personal infection history, intergroup contact, traditional news exposure, and trust in social media positively associate with prejudice against Asians and Asian Americans. Conversely, relying on left-leaning and balanced news outlets is related to less prejudice. For those who have more trust in social media and digital news, frequent use of these two sources is associated with lower levels of prejudice. This research highlights the urgent need to reshape part of the traditional news discourses. Emphasizing scientific evidence and political institutions' responsibility for creating effective solutions rather than placing the blame on ethnic minorities will decrease the citizens' perceptions of infection threats, thereby reducing disease-related racial prejudice. It is also critical for

public health organizations to leverage social media and mobile news apps for developing credible messages to combat racial discrimination against Asians linked to the COVID-19 pandemic.

Acknowledgments

This project is supported by the fundamental research funds for the Central Universities and the research funds of Renmin University of China (Project Number: 18XNB006).

Authors' Contributions

JYT and JP contributed to the conceptualization and writing of this paper. JYT conducted the data analysis. SP and CCY contributed to the survey design, manuscript review, and editing.

Conflicts of Interest

None declared.

References

1. Schaller M, Park JH. The behavioral immune system (and why it matters). Curr Dir Psychol Sci 2011 Apr 15;20(2):99-103. [doi: 10.1177/0963721411402596]
2. Schaller M. The behavioural immune system and the psychology of human sociality. Philos Trans R Soc Lond B Biol Sci 2011 Dec 12;366(1583):3418-3426 [FREE Full text] [doi: 10.1098/rstb.2011.0029] [Medline: 22042918]
3. Ackerman J, Hill S, Murray D. The behavioral immune system: current concerns and future directions. Soc Personal Psychol Compass 2018 Jan 26;12(2):e12371. [doi: 10.1111/spc3.12371]
4. Huang JY, Sedlovskaya A, Ackerman JM, Bargh JA. Immunizing against prejudice: effects of disease protection on attitudes toward out-groups. Psychol Sci 2011 Dec;22(12):1550-1556. [doi: 10.1177/0956797611417261] [Medline: 22058107]
5. Prati G, Pietrantoni L. Knowledge, risk perceptions, and xenophobic attitudes: evidence from Italy during the Ebola outbreak. Risk Anal 2016 Oct;36(10):2000-2010. [doi: 10.1111/risa.12537] [Medline: 26916803]
6. Stop AAPI Hate Reporting Center. Asian Pacific Policy and Planning Council. URL: http://www.asianpacificpolicyandplanningcouncil.org/stop-aapi-hate/ [accessed 2020-07-17]
7. Dhanani LY, Franz B. Unexpected public health consequences of the COVID-19 pandemic: a national survey examining anti-Asian attitudes in the USA. Int J Public Health 2020 Jul;65(6):747-754 [FREE Full text] [doi: 10.1007/s00038-020-01440-0] [Medline: 32728852]
8. Ruiz NG, Horowitz JM, Tamir C. Many Black and Asian Americans say they have experienced discrimination amid the COVID-19 outbreak. Pew Research Center. 2020 Jul 01. URL: https://www.pewsocialtrends.org/2020/07/01/many-black-and-asian-americans-say-they-have-experienced-discrimination-amid-the-covid-19-outbreak/ [accessed 2020-08-22]
9. Allport G. Nature of Prejudice. Harlow, England: Longman Higher Education; 1954.
10. Pettigrew TF, Tropp LR. A meta-analytic test of intergroup contact theory. J Pers Soc Psychol 2006 May;90(5):751-783. [doi: 10.1037/0022-3514.90.5.751] [Medline: 16737372]
11. Dovidio J, Love A, Schellhaas F, Hewstone M. Reducing intergroup bias through intergroup contact: twenty years of progress and future directions. Group Processes Intergroup Relat 2017 Jun 25;20(5):606-620. [doi: 10.1177/1368430217712052]
12. Neuberg S, Schaller M. An evolutionary threat-management approach to prejudices. Curr Opin Psychol 2016 Feb;7:1-5. [doi: 10.1016/j.copsyc.2015.06.004]
13. Wojcieszak M, Azrout R. I saw you in the news: mediated and direct intergroup contact improve outgroup attitudes. J Commun 2016 Dec 19;66(6):1032-1060. [doi: 10.1111/jcom.12266]
14. Schemer C, Meltzer CE. The impact of negative parasocial and vicarious contact with refugees in the media on attitudes toward refugees. Mass Commun Soc 2019 Nov 25;23(2):230-248. [doi: 10.1080/15205436.2019.1692037]
15. Yang J. A new virus stirs up ancient hatred. CNN. 2020 Jan 30. URL: https://www.cnn.com/2020/01/30/opinions/wuhan-coronavirus-is-fueling-racism-xenophobia-yang/index.html [accessed 2020-07-20]
16. Vachuska K. Initial effects of the coronavirus pandemic on racial prejudice in the United States: evidence from Google Trends. SocArXiv 2020:A. [doi: 10.31235/osf.io/bgpk3]
17. Gollust SE, Nagler RH, Fowler EF. The emergence of COVID-19 in the U.S.: a public health and political communication crisis. J Health Polit Policy Law 2020 May 28:8641506. [doi: 10.1215/03616878-8641506] [Medline: 32464658]
18. Shearer E. Local news is playing an important role for Americans during COVID-19 outbreak. Pew Research Center. 2020 Jul 02. URL: https://www.pewresearch.org/fact-tank/2020/07/02/local-news-is-playing-an-important-role-for-americans-during-covid-19-outbreak/ [accessed 2020-07-17]

19. De Coninck D, d'Haenens L, Matthijs K. Forgotten key players in public health: news media as agents of information and persuasion during the COVID-19 pandemic. Public Health 2020 Jun;183:65-66 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.011] [Medline: 32438213]
20. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357. [doi: 10.1037/hea0000875] [Medline: 32202824]
21. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
22. Reny T, Barreto M. Xenophobia in the time of pandemic: othering, anti-Asian attitudes, and COVID-19. Polit Groups Identities 2020 May 28:1-24. [doi: 10.1080/21565503.2020.1769693]
23. Sorokowski P, Groyecka A, Kowal M, Sorokowska A, Białek M, Lebuda I, et al. Can information about pandemics increase negative attitudes toward foreign groups? A case of COVID-19 outbreak. Sustainability 2020 Jun 16;12(12):4912. [doi: 10.3390/su12124912]
24. Mortensen CR, Becker DV, Ackerman JM, Neuberg SL, Kenrick DT. Infection breeds reticence: the effects of disease salience on self-perceptions of personality and behavioral avoidance tendencies. Psychol Sci 2010 Mar;21(3):440-447. [doi: 10.1177/0956797610361706] [Medline: 20424082]
25. Lee SWS, Huang JY, Schwarz N. Risk overgeneralization in times of a contagious disease threat. Front Psychol 2020;11:1392. [doi: 10.3389/fpsyg.2020.01392] [Medline: 32612565]
26. Kim HS, Sherman DK, Updegraff JA. Fear of Ebola: the influence of collectivism on xenophobic threat responses. Psychol Sci 2016 Jul;27(7):935-944. [doi: 10.1177/0956797616642596] [Medline: 27207872]
27. Talaska CA, Fiske ST, Chaiken S. Legitimating racial discrimination: emotions, not beliefs, best predict discrimination in a meta-analysis. Soc Justice Res 2008 Sep;21(3):263-396 [FREE Full text] [doi: 10.1007/s11211-008-0071-2] [Medline: 24052687]
28. Dixon TL, Josey CS. Race and news revisited: the content and effects of problematically framing the news. In: The International Encyclopedia of Media Studies. Oxford, England: Blackwell Publishing; 2013:272-291.
29. Dixon TL. Network news and racial beliefs: exploring the connection between national television news exposure and stereotypical perceptions of African Americans. J Commun 2008 Jun;58(2):321-337. [doi: 10.1111/j.1460-2466.2008.00387.x]
30. Ramasubramanian S, Martinez A. News framing of Obama, racialized scrutiny, and symbolic racism. Howard J Commun 2016 Oct 31;28(1):36-54. [doi: 10.1080/10646175.2016.1235519]
31. Liu B, Fraustino J, Jin Y. Social media use during disasters: how information form and source influence intended behavioral responses. Communic Res 2015 Jan 13;43(5):626-646. [doi: 10.1177/0093650214565917]
32. Yoo W, Choi D. Predictors of expressing and receiving information on social networking sites during MERS-CoV outbreak in South Korea. J Risk Res 2019 Feb 11:1-16. [doi: 10.1080/13669877.2019.1569105]
33. Liu Y, Finch BK, Brenneke SG, Thomas K, Le PD. Perceived discrimination and mental distress amid the COVID-19 pandemic: evidence from the understanding America study. Am J Prev Med 2020 Oct;59(4):481-492 [FREE Full text] [doi: 10.1016/j.amepre.2020.06.007] [Medline: 32829968]
34. Müller K, Schwarz C. From hashtag to hate crime: Twitter and anti-minority sentiment. SSRN J 2018 Mar:1. [doi: 10.2139/ssrn.3149103]
35. Budhwani H, Sun R. Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: quantitative analysis of social media data. J Med Internet Res 2020 May 06;22(5):e19301 [FREE Full text] [doi: 10.2196/19301] [Medline: 32343669]
36. Davidson T, Farquhar L. Prejudice and social media: attitudes toward illegal immigrants, refugees, and transgender people. In: Farris DN, Compton DR, Herrera AP, editors. Gender, Sexuality and Race in the Digital Age. New York: Springer; 2020:151-167.
37. Croucher SM, Nguyen T, Rahmani D. Prejudice toward Asian Americans in the COVID-19 pandemic: the effects of social media use in the United States. Front Commun 2020 Jun 12;5:1. [doi: 10.3389/fcomm.2020.00039]
38. Strömbäck J, Tsfati Y, Boomgaarden H, Damstra A, Lindgren E, Vliegenthart R, et al. News media trust and its impact on media use: toward a framework for future research. Ann Int Commun Assoc 2020 Apr 24;44(2):139-156. [doi: 10.1080/23808985.2020.1755338]
39. Tsfati Y, Ariely G. Individual and contextual correlates of trust in media across 44 countries. Communic Res 2013 May 02;41(6):760-782. [doi: 10.1177/0093650213485972]
40. Newman N, Fletcher R, Schulz A, Andi S, Nielsen RK. Reuters Institute Digital News Report. 2020. URL: http://www.digitalnewsreport.org/
41. Hart P, Chinn S, Soroka S. Politicization and polarization in COVID-19 news coverage. Sci Commun 2020 Aug 25:1. [doi: 10.1177/1075547020950735]
42. Motta M, Stecula D, Farhart C. How right-leaning media coverage of COVID-19 facilitated the spread of misinformation in the early stages of the pandemic in the U.S. Can J Pol Sci 2020 May 01;53(2):335-342. [doi: 10.1017/S0008423920000396]
43. Vaes J, Latrofa M, Vieno A, Pastore M. Exposure to politicized media and prejudice against immigrants in Italy: identifying its impact and psychological mediators. Psicologia sociale 2015;10(2):141-160. [doi: 10.1482/80762]

44. Hauser DJ, Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav Res Methods 2016 Mar;48(1):400-407. [doi: 10.3758/s13428-015-0578-z] [Medline: 25761395]
45. Levay K, Freese J, Druckman J. The demographic and political composition of Mechanical Turk samples. SAGE Open 2016 Mar 15;6(1):215824401663643. [doi: 10.1177/2158244016636433]
46. Boas T, Christenson D, Glick D. Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. Polit Sci Res Meth 2018 Aug 08;8(2):232-250. [doi: 10.1017/psrm.2018.28]
47. Christenson DP, Glick DM. Crowdsourcing panel studies and real-time experiments in MTurk. Political Methodologist 2013;20(2):27-32 [FREE Full text]
48. Park S, Boatwright B, Johnson Avery E. Information channel preference in health crisis: exploring the roles of perceived risk, preparedness, knowledge, and intent to follow directives. Public Relations Rev 2019 Dec;45(5):101794. [doi: 10.1016/j.pubrev.2019.05.015]
49. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
50. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
51. Korol L, Fietzer AW, Ponterotto JG. The relationship between multicultural personality, intergroup contact, and positive outgroup attitudes toward Asian Americans. Asian Am J Psychol 2018 Sep;9(3):200-210. [doi: 10.1037/aap0000107]
52. Jurkowitz M, Mitchell A, Shearer E, Walker M. U.S. media polarization and the 2020 election: a nation divided. Pew Research Center. 2020 Jan 24. URL: https://www.journalism.org/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/ [accessed 2020-07-17]
53. Media bias ratings. AllSides. URL: https://www.allsides.com/media-bias/media-bias-ratings [accessed 2020-07-17]
54. Zagefka H, González R, Brown R, Lay S, Manzi J, Didier N. To know you is to love you: effects of intergroup contact and knowledge on intergroup anxiety and prejudice among indigenous Chileans. Int J Psychol 2017 Aug;52(4):308-315. [doi: 10.1002/ijop.12229] [Medline: 26511584]
55. Vannette DL, Krosnick JA, editors. The Palgrave Handbook of Survey Research. Cham, Switzerland: Springer International Publishing; 2019.
56. Liu Y, West SG, Levy R, Aiken LS. Tests of simple slopes in multiple regression models with an interaction: comparison of four approaches. Multivariate Behav Res 2017;52(4):445-464. [doi: 10.1080/00273171.2017.1309261] [Medline: 28463014]
57. Preacher KJ, Rucker DD, Hayes AF. Addressing moderated mediation hypotheses: theory, methods, and prescriptions. Multivariate Behav Res 2007;42(1):185-227. [doi: 10.1080/00273170701341316] [Medline: 26821081]
58. Eichelberger L. SARS and New York's Chinatown: the politics of risk and blame during an epidemic of fear. Soc Sci Med 2007 Sep;65(6):1284-1295 [FREE Full text] [doi: 10.1016/j.socscimed.2007.04.022] [Medline: 17544189]
59. Bermejo F, Reggi L, Tiribelli S, Zuckerman E. Coverage of COVID-19 and political partisanship - comparing across nations. Media Cloud. 2020 Apr 03. URL: https://mediacloud.org/news/2020/4/3/coverage-of-covid-19-and-political-partisanship-comparing-across-nations [accessed 2020-08-22]
60. Gil de Zúñiga H, Weeks B, Ardèvol-Abreu A. Effects of the news-finds-me perception in communication: social media use implications for news seeking and learning about politics. J Comput Mediat Commun 2017 Apr 11;22(3):105-123. [doi: 10.1111/jcc4.12185]
61. Stechemesser A, Wenz L, Levermann A. Corona crisis fuels racially profiled hate in social media networks. EClinicalMedicine 2020 Jun;23:100372 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100372] [Medline: 32368725]
62. Li Y, Galea S. Racism and the COVID-19 epidemic: recommendations for health care workers. Am J Public Health 2020 Jul;110(7):956-957. [doi: 10.2105/ajph.2020.305698]

Abbreviations

H: hypothesis
MERS: Middle East respiratory syndrome
Mturk: Mechanical Turk
RQ: research question

Edited by G Eysenbach; submitted 22.07.20; peer-reviewed by N Martin, O Anikeeva; comments to author 12.08.20; revised version received 26.08.20; accepted 12.09.20; published 25.09.20.

Please cite as:
Tsai JY, Phua J, Pan S, Yang CC
Intergroup Contact, COVID-19 News Consumption, and the Moderating Role of Digital Media Trust on Prejudice Toward Asians in the United States: Cross-Sectional Study
J Med Internet Res 2020;22(9):e22767
URL: http://www.jmir.org/2020/9/e22767/
doi:10.2196/22767
PMID:32924948

Original Paper

Clinical Mortality in a Large COVID-19 Cohort: Observational Study

Mark Jarrett[1*], MD, MBA, MS; Susanne Schultz[2*], RN, MBA; Julie Lyall[2*], RN, MSN; Jason Wang[1*], PhD; Lori Stier[2*], RN, EdD; Marcella De Geronimo[3*], MS; Karen Nelson[2*], RN, MBA

[1]Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hofstra University, Hempstead, NY, United States

[2]Institute for Clinical Excellence and Quality/Safety, Northwell Health, New Hyde Park, NY, United States

[3]Krasnoff Quality Management Institute, Northwell Health, New Hyde Park, NY, United States

[*]all authors contributed equally

Corresponding Author:
Mark Jarrett, MD, MBA, MS
Donald and Barbara Zucker School of Medicine at Hofstra/Northwell
Hofstra University
500 Hofstra University
Hempstead, NY, 11549
United States
Phone: 1 5163216044
Email: MJarrett@northwell.edu

Abstract

Background: Northwell Health, an integrated health system in New York, has treated more than 15,000 inpatients with COVID-19 at the US epicenter of the SARS-CoV-2 pandemic.

Objective: We describe the demographic characteristics of patients who died of COVID-19, observation of frequent rapid response team/cardiac arrest (RRT/CA) calls for non–intensive care unit (ICU) patients, and factors that contributed to RRT/CA calls.

Methods: A team of registered nurses reviewed the medical records of inpatients who tested positive for SARS-CoV-2 via polymerase chain reaction before or on admission and who died between March 13 (first Northwell Health inpatient expiration) and April 30, 2020, at 15 Northwell Health hospitals. The findings for these patients were abstracted into a database and statistically analyzed.

Results: Of 2634 patients who died of COVID-19, 1478 (56.1%) had oxygen saturation levels ≥90% on presentation and required no respiratory support. At least one RRT/CA was called on 1112/2634 patients (42.2%) at a non-ICU level of care. Before the RRT/CA call, the most recent oxygen saturation levels for 852/1112 (76.6%) of these non-ICU patients were at least 90%. At the time the RRT/CA was called, 479/1112 patients (43.1%) had an oxygen saturation of <80%.

Conclusions: This study represents one of the largest reviewed cohorts of mortality that also captures data in nonstructured fields. Approximately 50% of deaths occurred at a non-ICU level of care despite admission to the appropriate care setting with normal staffing. The data imply a sudden, unexpected deterioration in respiratory status requiring RRT/CA in a large number of non-ICU patients. Patients admitted at a non-ICU level of care suffered rapid clinical deterioration, often with a sudden decrease in oxygen saturation. These patients could benefit from additional monitoring (eg, continuous central oxygenation saturation), although this approach warrants further study.

(J Med Internet Res 2020;22(9):e23565) doi:10.2196/23565

KEYWORDS

COVID-19; mortality; respiratory failure; hypoxemia; observational; review; cohort; ICU; intensive care unit

Introduction

Downstate New York was the first epicenter of the SARS-CoV-2 pandemic in the United States [1,2]. Northwell Health, an integrated health system, has treated more than 15,000 inpatients with COVID-19. Comprehensively analyzing the characteristics of patients who die of COVID-19 can help define the clinical nature of COVID-19 infection and potentially suggest new care

protocols. For 7 years, Northwell Health has used a centralized mortality review process with data validated through rigorous internal review and high interrater reliability (92% to 96%). This robust process was applied to a customized database to review all 2634 patients who died of COVID-19 in Northwell Health's adult acute care hospitals between March and April 2020. During this overwhelming surge, documentation was made in various notes as well as in structured fields in the electronic health record (EHR) systems. This study describes the demographic characteristics of patients who died of COVID-19 and the observation of frequent rapid response team/cardiac arrest (RRT/CA) calls for patients not in the intensive care unit (ICU). We also discuss factors that contributed to the RRT/CA calls, which may be a significant element in planning for a resurgence of the pandemic.

Methods

Study Design

Northwell Health is New York State's largest health care provider and private employer. With 23 hospitals (including specialty hospitals) and nearly 800 outpatient practice sites, the organization cares for over 2 million people in greater metropolitan New York. A team of registered nurses in the corporate quality department retrospectively reviewed medical records from 15 acute care hospitals. This team routinely conducts clinical reviews of all adult acute inpatient mortalities (approximately 5000 per year). A physician advisor was available to the team to consult on clinical questions.

Database elements were based on Northwell Health's experience with treating patients with COVID-19, literature review from countries that had early experience in treating patients, and clinical trials being conducted at the Feinstein Institutes for Medical Research. Also, the data were captured in the database established under the direction of critical care intensivists at the epicenter of the pandemic, other subject matter experts, and quality leadership. During data abstraction, modifications and enhancements were made to the database based on trends and emerging information. The demographic data, comorbidities, clinical findings, and management of COVID-19 patients who died were analyzed.

Patient Characteristics

The analyzed cases included inpatients who tested positive for SARS-CoV-2 via polymerase chain reaction before or on admission and who then died between March 13 (first Northwell Health inpatient death) and April 30, 2020. Emergency department (ED) mortalities were excluded. Demographic data and comorbidities were abstracted from the medical records of admitted patients. Initially, data were collected on 10 patient comorbidities that were deemed important and were then narrowed down to 6 comorbidities for inclusion based on our initial analysis. Transfers from one in-system hospital to another were merged and considered as a single visit. Notable patient outcomes that were measured were the level of ICU care (validated and abstracted from the provider order) and a call for RRT/CA. The Institutional Review Board of Northwell Health deemed this study as exempt and waived the requirement for informed consent.

Statistical Analysis

Statistical analyses were performed using chi-square tests for categorical variables and *t* tests for continuous variables. A multivariable logistic regression model was created to determine independent risk factors for the outcome variables. Statistical significance was considered at $P<.05$. All statistical analyses were performed in SAS v9.4 (SAS Institute).

Data Sharing

The data that support the findings of this study are available on request from COVID19@northwell.edu. The data are not publicly available due to restrictions, as this could compromise the privacy of the research participants.

Results

Patient Characteristics

The baseline characteristics of the 2634 patients who died of COVID-19 are described in Tables 1-3. The age range was 21-107 years in the following categories: 21 to 39 years (49/2634, 1.9%), 40 to 59 years (351/2634, 13.3%), 60 to 79 years (1241/2634, 47.1%), and ≥80 years (993/2634, 37.7%). In the patient cohort, 1664/2634 patients (63.2%) were male and 970/2634 (36.8%) were female. Among the 2634 patients, 1256 (47.7%) were White, 463 (17.6%) were Black, 230 (8.7%) were Asian, and 685 (26.0%) were of other/unknown race. The majority of patients (1839/2634, 69.8%) reported Medicare as their insurance. The most common comorbidities among these patients were hypertension (1719/2634, 65.3%), diabetes (1043/2634, 39.6%), and dementia (431/2634, 16.4%). Fewer patients had chronic obstructive pulmonary disease (385/2634, 14.6%), heart failure (291/2634, 11.1%), and end stage renal disease (166/2634, 6.3%). Of these six comorbidities, more than half of the patients (1350/2634, 51.3%) had 2 or more comorbidities, and 445/2634 (16.9%) had 0 comorbidities. The majority of patients with a known BMI, calculated as weight in kilograms divided by height in meters squared, of 25 or more were categorized as follows: 25 to 29.99 (732/2634, 27.8%), 30 to 34.99 (401/2634, 15.2%), 35 to 39.99 (190/2634, 7.2%), and ≥40 (147/2634, 5.6%).

Table 1. Baseline characteristics of patients hospitalized with COVID-19 who died (N=2634), n (%).

Baseline characteristic	Value
Age (years)	
21-39	49 (1.86)
40-59	351 (13.3)
60-79	1241 (47.1)
≥80	993 (37.7)
Sex	
Male	1664 (63.2)
Female	970 (36.8)
Race	
White	1256 (47.7)
Black	463 (17.6)
Asian	230 (8.7)
Other/unknown	685 (26.0)
Payment method	
Commercial insurance	413 (15.7)
Medicaid	341 (13.0)
Medicare	1839 (69.8)
Self-pay	41 (1.6)
Comorbidities	
Hypertension	1719 (65.3)
COPD[a]	385 (14.6)
Diabetes	1043 (39.6)
Heart failure	291 (11.1)
Dementia	431 (16.4)
End stage renal disease	166 (6.3)
Number of comorbidities	
0	445 (16.9)
1	839 (31.9)
2	934 (35.5)
3	343 (13.0)
4	66 (2.5)
5	7 (0.3)
6	0 (0.0)
BMI (kg/m^2)	
Unknown	494 (18.8)
<18.5	82 (3.1)
18.5-24.99	588 (22.3)
25-29.99	732 (27.8)
30-34.99	401 (15.2)
35-39.99	190 (7.2)
≥40	147 (5.6)

[a]COPD: chronic obstructive pulmonary disease.

Table 2. Hospitalization characteristics of patients hospitalized with COVID-19 who died (N=2634), n (%).

Hospitalization characteristic	Value
Admission source	
Home	1895 (72.0)
Rehabilitation	127 (4.8)
Skilled nursing facility	411 (15.6)
Transfer from another acute care hospital	201 (7.6)
Emergency department visit	
Within 48 hours of this admission	51 (1.9)
Within 7 days of this admission	125 (4.8)
Readmission	
Within 24 hours	20 (0.8)
Within 7 days	75 (2.9)
Within 30 days	194 (7.4)
Level of care at time of death	
ICU[a]	1299 (49.3)
Non-ICU	1335 (50.7)
Level of care at time of admission	
ICU	541 (20.5)
Medical/surgical unit	1230 (46.7)
Telemetry/stepdown unit	863 (32.8)
Overall length of stay (days)	
0-7	1420 (53.9)
≥8	1214 (46.1)
ICU length of stay (days)	
0-7	872 (33.1)
≥8	574 (21.8)
Oxygen saturation on presentation (%)	
<80	459 (17.4)
80-89.9	667 (25.3)
≥90	1478 (56.1)
Unable to determine	30 (1.2)
Initial respiratory support on presentation	
None	1397 (53.0)
Nasal cannula	363 (13.8)
Nonrebreather mask	742 (28.2)
Ventilator	24 (0.9)
High-flow nasal cannula	8 (0.3)
Ventimask	11 (0.4)
BiPAP[b]	13 (0.5)
Other	27 (1.0)
Unable to determine	49 (1.9)
RRT/CA[d] while not at ICU level of care	1112 (42.2)

Hospitalization characteristic	Value
Proning	
Yes	756 (28.7)
No	1878 (71.3)
Proning without mechanical ventilation (n=756)	191 (25.3)
Proning prior to mechanical ventilation (n=756)	213 (28.2)
Proning during mechanical ventilation (n=756)	214 (28.3)
Proning prior to and during mechanical ventilation(n=756)	138 (18.3)
DNR[d] complete	1631 (61.9)
Palliative care consult	1014 (38.5)
Clinical trial inclusion	114 (4.3)

[a]ICU: intensive care unit.

[b]BiPAP: bilevel positive airway pressure

[c]RRT/CA: rapid response team/cardiac arrest.

[d]DNR: do not resuscitate.

Table 3. Mechanical ventilation characteristics of patients hospitalized with COVID-19 who died.

Mechanical ventilation characteristic	n	%	
		Total patients (N=2634)	Patients who were ventilated (n=1403)
Traditional ventilator	1259	47.9	89.7
Converted BiPAP[a]	142	0.1	10.1
Anesthesia machine	2	0.08	0.1
Increased oxygen requirement prior to mechanical ventilation	1332	50.6	94.9
Mechanical ventilation length, days			
0-7	851	32.3	60.7
≥8	552	20.9	39.3
Terminal wean	270	10.3	19.2

[a]biPAP: bilevel positive airway pressure.

Patient Outcomes

Most patients were admitted from home (1895/2634, 71.9%). The remaining patients were admitted from a skilled nursing facility (411/2634, 15.6%), an acute care facility (201/2634, 7.6%), or a rehabilitation facility (127/2634, 4.8%). The percentage of patients with a prior ED visit within 7 days of admission was 4.8% (125/2634), and that of patients with a prior ED visit within 48 hours of admission was 1.9% (51/2634). The percentage of patients readmitted within 30 days was 7.4% (194/2634), 2.9% (75/2634) were readmitted within 7 days, and 0.8% (20/2634) were readmitted within 24 hours. On presentation, most patients (1478/2634, 56.1%) had an oxygen saturation level greater than or equal to 90%, and more than half (1397/2634, 53.0%) required no respiratory support. Others required a nasal cannula (363/2634, 13.8%), a nonrebreather mask (742/2634, 28.2%), or mechanical ventilation (24/2634, 0.9%). More than half of the patients who died (1403/2634, 53.2%) required mechanical ventilation during their clinical course. Of those 1403 patients, 1332 (94.9%) had increasing oxygen requirements before intubation, 1259 (89.7%) were on traditional ventilators, 142 (10.1%) were on converted BiPAP machines, and 2 (0.1%) were on anesthesia machines. The length of time on mechanical ventilation was 0 to 7 days for 851/1403 patients (60.7%) and 8 days or more for 552/1403 patients (39.3%).

Prone positioning was documented for 756/2634 patients (28.7%), and 270/2634 patients (10.3%) patients were terminally weaned. Do not resuscitate (DNR) orders were completed for 1631/2634 patients (61.9%). A palliative care consult was provided to 1014/2634 patients (38.5%). At the time of death, the level of care was ICU for 1299/2634 patients (49.3%) and non-ICU for 1335/2634 patients (50.7%).

Patient Outcomes Based on RRT/CA Calls

Of the 2634 patients, 1112 (42.2%) had an RRT/CA call at a non-ICU level of care, while 1522 (57.8%) did not. As shown in Tables 4-6, the RRT/CA group was significantly different from the non-RRT/CA group in terms of age, race, and comorbidities. Among patients between 60 and 79 years of age,

618/1112 (55.6%) were in the RRT/CA group and 623/1522 (40.9%) were in the non-RRT/CA group. In terms of race, there were significantly fewer White patients in the RRT/CA group (404/1112, 36.3%, versus 852/1522, 56.0%; *P*<.001). The RRT/CA cohort had a significantly higher rate of patients with diabetes (491/1112, 44.2%, versus 552/1522, 36.3%; *P*<.001). Patients in the RRT/CA cohort were more likely to be admitted from home (926/1112, 83.3%) than patients in the non-RRT/CA cohort (969/1522, 63.7%). Patients in the RRT/CA cohort were more likely than patients in the non-RRT/CA cohort to be admitted to a medical/surgical unit (576/1112, 51.8%, versus 654/1522, 42.9%) or telemetry/step-down unit (455/1112, 40.9%, versus 408/1522, 26.8%), and to die at an ICU level of care (671/1112, 60.3%, versus 628/1522, 41.3%). An overall length of stay (LOS) of 8 days or more was more common in the RRT/CA cohort (645/1112, 58.0%) than in the non-RRT/CA cohort (569/1522, 37.4%), as was an ICU LOS of 0 to 7 days (472/1112, 42.0%, versus 400/1522, 26.3%) and of 8 days or more (271/1112, 24.4%, versus 303/1522, 19.9%). After adjusting for demographic and clinical characteristics, oxygen saturation levels at presentation were significant for the RRT/CA cohort at oxygen saturation levels of 80% to 89% (odds ratio [OR] 1.988, 95% CI 1.511-2.616) and of ≥90% (OR 2.517, 95% CI 1.962-3.230). For the logistic regression results, see Table 7.

Table 4. Baseline characteristics of patients who died of COVID-19 who experienced an RRT/CA call at a non-ICU level of care (N=2634).

Baseline characteristics	RRT/CA[a] call		
	Yes (n=1112), n (%)	No (n=1522), n (%)	*P* value
Age (years)			<.001
21-39	19 (1.7)	30 (2.0)	
40-59	194 (17.5)	157 (10.3)	
60-79	618 (55.6)	623 (40.9)	
≥80	281 (25.3)	712 (40.8)	
Sex			.35
Male	714 (64.2)	950 (62.4)	
Female	398 (35.8)	572 (37.6)	
Race			<.001
White	404 (36.3)	852 (56.0)	
Black	235 (21.1)	228 (15.0)	
Asian	125 (11.2)	105 (6.9)	
Other/unknown	348 (31.3)	337 (22.1)	
Payment method			<.001
Commercial insurance	226 (20.3)	187 (12.3)	
Medicaid	166 (14.9)	175 (11.5)	
Medicare	702 (63.1)	1137 (74.7)	
Self-pay	18 (1.6)	23 (1.5)	
Comorbidities			
Hypertension			.24
Yes	740 (66.5)	979 (64.3)	
No	372 (33.5)	543 (35.7)	
COPD[b]			.08
Yes	147 (13.2)	238 (15.6)	
No	965 (86.8)	1284 (84.4)	
Diabetes			<.001
Yes	491 (44.2)	552 (36.3)	
No	621 (55.9)	970 (63.7)	
Heart failure			.03
Yes	106 (9.5)	185 (12.2)	
No	1006 (90.5)	1337 (87.8)	
Dementia			<.001
Yes	98 (8.8)	333 (21.9)	
No	1014 (91.2)	1189 (78.1)	
End stage renal disease			.02
Yes	85 (7.6)	81 (5.3)	
No	1027 (92.4)	1441 (94.7)	
Number of comorbidities			.47
0	202 (18.2)	243 (15.9)	
1	355 (31.9)	484 (31.8)	

Baseline characteristics	RRT/CA[a] call		
	Yes (n=1112), n (%)	No (n=1522), n (%)	*P* value
2	388 (34.9)	546 (35.9)	
3	134 (12.1)	209 (13.7)	
4	31 (2.8)	35 (2.3)	
5	2 (0.2)	5 (0.3)	
BMI (kg/m^2)			<.001
Unknown	136 (12.2)	358 (23.5)	
<18.5	22 (1.9)	60 (3.9)	
18.5-24.99	236 (21.2)	352 (23.1)	
25-29.99	352 (31.7)	380 (24.9)	
30-34.99	206 (18.5)	195 (12.8)	
35-39.99	88 (7.9)	102 (6.7)	
≥40	72 (6.5)	75 (4.9)	

[a]RRT/CA: rapid response team/cardiac arrest.

[b]COPD: chronic obstructive pulmonary disease.

Table 5. Hospitalization characteristics of patients who died of COVID-19 who experienced an RRT/CA call at a non-ICU level of care (N=2634).

Baseline characteristics	RRT/CA[a] call		
	Yes (n=1112), n (%)	No (n=1522), n (%)	*P* value
Admission source			<.001
Home	926 (83.3)	969 (63.7)	
Rehabilitation	34 (3.0)	93 (6.1)	
Skilled nursing facility	80 (7.2)	331 (21.7)	
Transfer from another acute care hospital	72 (6.5)	129 (8.5)	
Emergency department visit			
Within 48 hours of this admission			.03
Yes	29 (2.6)	22 (1.5)	
No	1083 (97.4)	1500 (98.6)	
Within 7 days of this admission			.13
Yes	61 (5.5)	64 (4.2)	
No	1051 (94.5)	1458 (95.8)	
Readmission			
Within 24 hours			.51
Yes	7 (0.6)	13 (0.9)	
No	1105 (99.4)	1509 (99.2)	
Within 7 days			.88
Yes	31 (2.8)	44 (2.9)	
No	1081 (97.2)	1478 (97.1)	
Within 30 days			.10
Yes	71 (6.4)	123 (8.1)	
No	1041 (93.6)	1399 (91.9)	
Level of care at time of death			N/A[b]
ICU[c]	671 (60.3)	628 (41.3)	
Non-ICU	441 (39.7)	894 (58.7)	
Level of care at time of admission			<.001
ICU	81 (7.3)	460 (30.2)	
Medical/surgical unit	576 (51.8)	654 (42.9)	
Telemetry/stepdown unit	455 (40.9)	408 (26.8)	
Overall length of stay (days)			<.001
0-7	467 (42.0)	953 (62.6)	
≥8	645 (58.0)	569 (37.4)	
ICU length of stay (days)			<.001
0-7	472 (42.4)	400 (26.3)	
≥8	271 (24.4)	303 (19.9)	
Oxygen saturation on presentation (%)			<.001
<80	152 (13.7)	307 (20.2)	
80-89.9	289 (26.0)	378 (24.8)	
≥90	664 (59.7)	814 (53.5)	
Unable to determine	7 (0.6)	23 (1.5)	

Baseline characteristics	RRT/CA[a] call		
	Yes (n=1112), n (%)	No (n=1522), n (%)	*P* value
Initial respiratory support on presentation			<.001
None	687 (61.8)	710 (46.7)	
Nasal cannula	161 (14.5)	202 (13.3)	
High-flow nasal cannula	0 (0.0)	8 (0.5)	
Ventimask	2 (0.2)	9 (0.6)	
BiPAP[d]	2 (0.2)	11 (0.7)	
Nonrebreather mask	239 (21.5)	503 (33.1)	
Ventilator	1 (0.1)	23 (1.5)	
Other	4 (0.4)	23 (1.5)	
Unable to determine	16 (1.4)	33 (2.2)	
Mechanical ventilation	723 (65.0)	680 (44.7)	<.001
Type of mechanical ventilation			
Traditional ventilator	650 (58.5)	609 (40.0)	
Converted BiPAP	71 (6.4)	71 (4.7)	
Anesthesia machine	2 (0.2)	0 (0.0)	
Increased oxygen requirement before mechanical ventilation	699 (62.9)	633 (41.6)	<.001
Mechanical ventilation length (days)			
0-7	461 (41.5)	390 (25.6)	
≥8	262 (23.6)	290 (19.1)	
Terminal wean			.52
Yes	109 (9.8)	161 (10.6)	
No	1003 (90.2)	1361 (89.4)	
Proning			<.001
Yes	500 (45.0)	256 (16.8)	
No	612 (54.9)	1266 (83.2)	
Proning without mechanical ventilation	116 (10.4)	75 (4.9)	
Proning before mechanical ventilation	171 (15.4)	42 (2.7)	
Proning during mechanical ventilation	99 (8.9)	115 (7.5)	
Proning before and during mechanical ventilation	114 (10.3)	24 (1.6)	
DNR[e] complete			<.001
Yes	558 (50.2)	1073 (70.5)	
No	554 (49.8)	449 (29.5)	
Palliative care consult			<.001
Yes	385 (34.6)	629 (41.3)	
No	727 (65.4)	893 (58.7)	
Clinical trial inclusion			N/A
Yes	91(8.2)	23(1.5)	

Baseline characteristics	RRT/CA[a] call		
	Yes (n=1112), n (%)	No (n=1522), n (%)	*P* value
No	1021(91.8)	1499 (98.5)	

[a]RRT/CA: rapid response team/cardiac arrest.

[b]N/A: not applicable.

[c]ICU: intensive care unit.

[d]BiPAP: bilevel positive airway pressure.

[e]DNR: do not resuscitate.

Table 6. Additional characteristics associated with RRT/CA calls for patients at a non–intensive care unit level of care (n=1112), n (%).

Characteristic	Value
Required escalation in level of care following initial RRT/CA[a] call	716 (64.4)
Oxygen saturation at time RRT/CA call initiated (%)	
<80	479 (43.1)
80-89	407 (36.6)
≥90	128 (11.5)
Unable to determine	98 (8.8)
Oxygen supplement at time RRT/CA call initiated	
Nonrebreather mask with or without nasal cannula	868 (78.1)
Nasal cannula	147 (13.2)
Room air	40 (3.6)
Ventimask	18 (1.6)
Ventilator	11 (1.0)
High-flow nasal cannula	9 (0.8)
BiPAP[b]	5 (0.4)
Unable to determine	14 (1.3)
Most recent oxygen saturation before RRT/CA initiated (%)	
<80	43 (3.9)
80-89	211 (18.9)
90≤	852 (76.6)
Unable to determine	6 (0.5)
Documented timing of most recent oxygen saturation before RRT/CA initiated (hours)	
<1	263 (23.7)
1-2	191 (17.2)
2-3	140 (12.6)
3-4	109 (9.8)
>4	409 (36.8)

[a]RRT/CA: rapid response team/cardiac arrest.

[b]BiPAP: bilevel positive airway pressure.

Table 7. Regression analysis of patients who died of COVID-19 who experienced a rapid response team/cardiac arrest call at a non–intensive care unit level of care (N=2634).

Baseline characteristics	Estimate	*P* value	Odds ratio	95% CI
Age (years)				
50-69	0.2653	.20	1.304	0.872-1.949
70-79	0.1721	.44	1.188	0.766-1.842
≥80	–0.3179	.17	0.728	0.460-1.151
Sex				
Male	–0.2299	.02	0.795	0.658-0.960
Race				
Black	0.6134	<.001	1.847	1.445-2.361
Asian	0.6548	<.001	1.925	1.395-2.655
Other/unknown	0.5333	<.001	1.704	1.362-2.133
Payment method				
Medicaid	–0.0458	.78	0.955	0.691-1.321
Medicare	–0.0107	.94	0.989	0.750-1.305
Self-pay	–0.3020	.40	0.739	0.367-1.488
Comorbidities				
Heart failure	0.1429	.34	1.154	0.860-1.547
End stage renal disease	0.6184	.002	1.856	1.262-2.729
COPD[a]	–0.1216	.35	0.886	0.687-1.141
Hypertension	0.1239	.21	1.132	0.931-1.376
Diabetes mellitus	0.0833	.38	1.087	0.902-1.310
BMI (kg/m^2)				
Unknown	–0.4645	<.001	0.628	0.491-0.804
≥30	–0.0545	.62	0.947	0.765-1.173
Admit source				
Home	0.9060	<.001	2.474	1.850-3.310
Rehabilitation	0.2904	.25	1.337	0.813-2.199
Transfer from acute care hospital	0.0544	.80	1.056	0.691-1.614
Oxygen saturation on presentation (%)				
80-89	0.6871	<.001	1.988	1.511 2.616
≥90	0.9232	<.001	2.517	1.962 3.230
Proning	1.1840	<.001	3.267	2.667 4.003

[a]COPD: chronic obstructive pulmonary disease.

Discussion

Summary of Findings

This study represents a review of one of the largest cohorts of COVID-19 mortality that includes data documented in nonstructured fields within the EHR. An experienced team of registered nurses was able to extract detailed information from the medical record that is typically not included in a structured data set analysis. The demographics of the patients who died are similar to those in other published studies: age predominately over 69, male majority, payor mix (reflecting age and Medicare along with a low number of self-paying patients, namely 41/2634, 1.6%), and multiple comorbidities [3-12].

Circumstances Preceding Patient Deterioration

This study provides a detailed clinical picture of the circumstances that precede the sudden deterioration in non-ICU patients reported by clinicians, which have not been fully examined in the literature. A striking reported feature of COVID-19 is the rapid progression of respiratory failure soon after the onset of dyspnea and hypoxemia [13]. The US National Institutes of Health (NIH) has reported that hypoxemia is

common in hospitalized patients with COVID-19 and that the criteria for hospital admission, ICU admission, and mechanical ventilation differ between countries [14]. In some hospitals in the United States, more than 25% of hospitalized patients require ICU care, mostly due to acute respiratory failure. The NIH recommends close monitoring for worsening respiratory status for adults with COVID-19 who are receiving supplemental oxygen. These recommendations align with our findings in the non-ICU patient population.

Approximately half of the deaths (1335/2634, 50.7%) occurred at a non-ICU level of care despite admission to the appropriate care setting with normal staffing. Our analysis of patients who experienced at least one RRT/CA call at a non-ICU level of care revealed that 716/1112 (64.4%) required an escalation in their level of care. Of the RRT/CA patients, 664/1112 (59.7%) presented to the hospital with oxygen saturation levels greater than or equal to 90%. In addition, 687/1112 (61.8%) had no oxygen support. Of the RTT/CA patients, 1031/1112 (92.7%) were admitted to a non-ICU level of care with normal staffing levels, which was appropriate based on their care needs. At presentation to the ED, the oxygen saturation levels for these patients were significantly higher than those for patients admitted to the ICU. Before the RRT/CA call, the most recent oxygen saturation levels recorded for the non-ICU patients remained high, at ≥90% for 852/1112 (76.6%) of patients. Oxygen saturations were documented within two hours of the RRT/CA call in 454/1112 (40.9%) of patients in the RRT/CA cohort. When the RRT/CA was called, 479/1112 (43.1%) of patients had an oxygen saturation less than 80%, and 78.1% (868/1112) were on a nonrebreather mask or a nonrebreather mask with nasal cannula. These data imply a sudden, unexpected deterioration in respiratory status requiring an RRT/CA call in a large number of non-ICU patients.

Limitations

This study includes the following limitations. First, the study focuses on the demographic and clinical characteristics of in-hospital COVID-19 patients who died between March 13 and April 30, 2020; it does not provide a comparison group of similar patients who survived during the same time period. Second, data were obtained from the EHR and manually abstracted from medical records through retrospective review; however, some routine documentation was less detailed due to the volume of patients being treated. Third, race was documented as other/unknown in 685/2634 (26%) of patients; therefore, conclusions about race could not be drawn. Fourth, missing BMI data were included in the category of “unknown” BMI. Finally, the study does not recognize a specific trigger that can distinguish which non-ICU patients in the cohort should be monitored.

Conclusions

Patients admitted to a non-ICU level of care appear to suffer rapid clinical deterioration, often with the hallmark of a sudden decrease in oxygen saturation. This finding suggests that non-ICU patients could benefit from additional monitoring, such as continuous central oxygenation saturation. The availability of wireless patch monitoring should be considered along with other methods, such as carbon dioxide and cardiac monitoring. Although this approach does not ensure reduced mortality, the number of RRT/CA calls infers that this area warrants further study.

Acknowledgments

This work was supported by grants R24AG064191 from the National Institute on Aging of the National Institutes of Health and R01LM012836 from the National Library of Medicine of the National Institutes of Health. We thank the Northwell Health Institute for Clinical Excellence and Quality/Safety: Mary Cama, RN, MSN; Patricia Meo, RN, BSN; Amy Logeman, RN, BSN, CPPS; Maureen McCarthy, RN, BSN; Josephine Fernandez-Kapilevich, RN, BSN; Theresa A Droluk, RN, BSN; Jessica Martin, RN, BSN, RN-BC; Allison Carballo, RN, MBA; and Jimmy Diaz, BS. We also thank the Northwell Health Krasnoff Quality Management Institute: Alex Ma; Kahliik S Burrell; and Mark P Tursi, MBA. We would like to acknowledge the contributions of the Northwell Health COVID-19 Research Consortium, including Crystal R Herron, PhD, and Jennifer C Johnson, MS, for editorial support. We acknowledge and honor all our Northwell Health team members who consistently put themselves in harm’s way during the COVID-19 pandemic. We dedicate this article to them, as their vital contribution to knowledge about COVID-19 and sacrifices on the behalf of patients made it possible.

Authors' Contributions

MPJ had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. MPJ, SES, JSL, and KLN were responsible for the conception and design of the study. MPJ, SES, JSL, JJW, LS, MDG, and KLN were responsible for data acquisition, analysis, and interpretation. MPJ, SES, JSL, JJW, LS, and KLN were responsible for drafting the manuscript. MPJ, SES, JSL, JJW, LS, MDG, and KLN were responsible for critical revision of the manuscript for important intellectual content. JJW was responsible for the statistical analysis. MPJ, SES, JSL, JJW, LS, MDG, and KLN were responsible for administrative, technical, and material support. MPJ supervised the study.

Conflicts of Interest

None declared.

References

1. COVID-19 Dashboard. The Center for Systems Science and Engineering at Johns Hopkins University. 2020. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/ [accessed 2020-03-24]
2. COVID-19 Tracker. New York State Department of Health. 2020. URL: https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Map?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n#/views [accessed 2020-05-07]
3. CDC COVID-19 Response Team. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 03;69(13):382-386 [FREE Full text] [doi: 10.15585/mmwr.mm6913e2] [Medline: 32240123]
4. Wortham JM, Lee JT, Althomsons S, Latash J, Davidson A, Guerra K, et al. Characteristics of Persons Who Died with COVID-19 - United States, February 12-May 18, 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 17;69(28):923-929 [FREE Full text] [doi: 10.15585/mmwr.mm6928e1] [Medline: 32673298]
5. Wang C, Horby P, Hayden F, Gao G. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/S0140-6736(20)30185-9]
6. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
7. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3]
8. Goyal P, Choi J, Pinheiro L, Schenck EJ, Chen R, Jabri A, et al. Clinical Characteristics of Covid-19 in New York City. N Engl J Med 2020 Jun 11;382(24):2372-2374 [FREE Full text] [doi: 10.1056/NEJMc2010419] [Medline: 32302078]
9. Petrilli C, Jones S, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020 May 22;369:m1966. [doi: 10.1136/bmj.m1966] [Medline: 32444366]
10. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):458-464 [FREE Full text] [doi: 10.15585/mmwr.mm6915e3] [Medline: 32298251]
11. Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, El Burai Felix S, et al. Coronavirus Disease 2019 Case Surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):759-765. [doi: 10.15585/mmwr.mm6924e2] [Medline: 32555134]
12. Yancy CW. COVID-19 and African Americans. JAMA 2020 May 19;323(19):1891-1892. [doi: 10.1001/jama.2020.6548] [Medline: 32293639]
13. Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. N Engl J Med 2020 May 15:online. [doi: 10.1056/nejmcp2009575]
14. COVID-19 Treatment Guidelines. US National Institutes of Health. URL: https://www.covid19treatmentguidelines.nih.gov/introduction/ [accessed 2020-04-24]

Abbreviations

DNR: do not resuscitate
ED: emergency department
EHR: electronic health record
ICU: intensive care unit
LOS: length of stay
NIH: National Institutes of Health
OR: odds ratio
RRT/CA: rapid response team/cardiac arrest

Edited by G Eysenbach; submitted 16.08.20; peer-reviewed by D Hirschwerck; comments to author 04.09.20; revised version received 09.09.20; accepted 14.09.20; published 25.09.20.

Please cite as:
Jarrett M, Schultz S, Lyall J, Wang J, Stier L, De Geronimo M, Nelson K
Clinical Mortality in a Large COVID-19 Cohort: Observational Study
J Med Internet Res 2020;22(9):e23565
URL: http://www.jmir.org/2020/9/e23565/
doi:10.2196/23565
PMID:32930099

Original Paper

Understanding the Community Risk Perceptions of the COVID-19 Outbreak in South Korea: Infodemiology Study

Atina Husnayain[1,2], MPH; Eunha Shim[3], PhD; Anis Fuad[2], DEA; Emily Chia-Yu Su[1,4,5], PhD

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[2]Department of Biostatistics, Epidemiology, and Population Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

[3]Department of Mathematics, Soongsil University, Seoul, Republic of Korea

[4]Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan

[5]Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Emily Chia-Yu Su, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
172-1 Keelung Rd, Sec 2
Taipei, 106
Taiwan
Phone: 886 2 66382736 ext 1515
Email: emilysu@tmu.edu.tw

Abstract

Background: South Korea is among the best-performing countries in tackling the coronavirus pandemic by using mass drive-through testing, face mask use, and extensive social distancing. However, understanding the patterns of risk perception could also facilitate effective risk communication to minimize the impacts of disease spread during this crisis.

Objective: We attempt to explore patterns of community health risk perceptions of COVID-19 in South Korea using internet search data.

Methods: Google Trends (GT) and NAVER relative search volumes (RSVs) data were collected using COVID-19–related terms in the Korean language and were retrieved according to time, gender, age groups, types of device, and location. Online queries were compared to the number of daily new COVID-19 cases and tests reported in the Kaggle open-access data set for the time period of December 5, 2019, to May 31, 2020. Time-lag correlations calculated by Spearman rank correlation coefficients were employed to assess whether correlations between new COVID-19 cases and internet searches were affected by time. We also constructed a prediction model of new COVID-19 cases using the number of COVID-19 cases, tests, and GT and NAVER RSVs in lag periods (of 1-3 days). Single and multiple regressions were employed using backward elimination and a variance inflation factor of <5.

Results: The numbers of COVID-19–related queries in South Korea increased during local events including local transmission, approval of coronavirus test kits, implementation of coronavirus drive-through tests, a face mask shortage, and a widespread campaign for social distancing as well as during international events such as the announcement of a Public Health Emergency of International Concern by the World Health Organization. Online queries were also stronger in women (r=0.763-0.823; P<.001) and age groups ≤29 years (r=0.726-0.821; P<.001), 30-44 years (r=0.701-0.826; P<.001), and ≥50 years (r=0.706-0.725; P<.001). In terms of spatial distribution, internet search data were higher in affected areas. Moreover, greater correlations were found in mobile searches (r=0.704-0.804; P<.001) compared to those of desktop searches (r=0.705-0.717; P<.001), indicating changing behaviors in searching for online health information during the outbreak. These varied internet searches related to COVID-19 represented community health risk perceptions. In addition, as a country with a high number of coronavirus tests, results showed that adults perceived coronavirus test–related information as being more important than disease-related knowledge. Meanwhile, younger, and older age groups had different perceptions. Moreover, NAVER RSVs can potentially be used for health risk perception assessments and disease predictions. Adding COVID-19–related searches provided by NAVER could increase the performance of the model compared to that of the COVID-19 case–based model and potentially be used to predict epidemic curves.

Conclusions: The use of both GT and NAVER RSVs to explore patterns of community health risk perceptions could be beneficial for targeting risk communication from several perspectives, including time, population characteristics, and location.

(J Med Internet Res 2020;22(9):e19788) doi:10.2196/19788

KEYWORDS

Google Trends; risk; perception; communication; COVID-19; South Korea; outbreak; infodemiology

Introduction

The World Health Organization (WHO) declared the COVID-19 outbreak a pandemic on March 11, 2020 [1]. By May 31, 2020, the disease had infected 5,934,936 individuals worldwide [2] including 11,468 individuals in South Korea. The first COVID-19 case in South Korea was confirmed on January 20, 2020 [3]. Slow upturns in disease transmission were reported before February 19, 2020; the local clusters observed in Daegu led to daily increases in the number of new cases [4]. Numerous approaches were undertaken to prevent disease transmission, including coronavirus drive-through testing and social distancing [5,6]. Coronavirus drive-through tests were identified as a safe and efficient screening approach, with each test taking approximately 10 minutes, thus minimizing cross-infection among people being tested [6]. To date, the average number of daily new cases is lower by ten-fold or more compared to those during the peak of the epidemic (from February 19 to March 15, 2020) [3]. Consequently, South Korea is considered among the best-performing countries in tackling the pandemic.

On the contrary, adequate risk communication could also have helped minimize the impacts of disease transmission [7]. Thus, in the pandemic period, the WHO suggests regular risk communication by updating the public and stakeholders on any changes in the status of the pandemic [8]. This action might be challenging because proper risk communication needs a robust understanding of risk perceptions, which helps to identify what knowledge the public needs [7]. However, studies exploring risk perception are often conducted using survey methods or content analyses [7,9-11], which require more resources and longer time. In particular, when investigating an emerging disease, those approaches might be less affordable since the health system will be overburdened with the surge of health care use, thus resulting in more barriers to assessing community health risk perceptions.

Therefore, this study aims to explore patterns of community health risk perceptions toward COVID-19 in South Korea using internet search data. This study is part of infodemiological research that was first introduced in 1996 [12] and explores the distribution of information on the internet [13] for public health and policy about the ground situation in the population. Infodemiology commonly deals with disease-related topics as well as outbreaks and epidemics [14]. This approach can potentially be used since internet query data can be provided easily, promptly, [15], and in a cost-effective manner compared to survey methods [16], and it can potentially capture anomalous patterns in near real time [17].

In this analysis, we used COVID-19–related internet search data provided by Google Trends (GT) and NAVER to represent online queries from the world's largest search engine and Korean local search engine, which has a higher market share than Google in South Korea [18]. This study explores patterns of public health risk perceptions toward the ongoing outbreak from several different perspectives, including time, population characteristics, and location as used in epidemiological studies. We also constructed a prediction model of new COVID-19 cases using the number of COVID-19 cases, tests, and GT and NAVER relative search volumes (RSVs) in lag periods (of 1-3 days). Future studies are warranted to define the best lag period to perform effective risk communication in the early stages of a disease outbreak.

Methods

Data Sets

The daily numbers of new COVID-19 cases and coronavirus tests from January 20 to May 31, 2020, were collected from the Kaggle open-access data set by Kim and colleagues [3]. We used the Time.csv data set to retrieve the number of new daily COVID-19 cases and daily tests, and the TimeProvince.csv data set to collect cumulative coronavirus cases by region. Those data sets covered all cities in South Korea. In addition, internet search data related to COVID-19 were retrieved from the GT [19] and NAVER websites [20] in the same collocation. The information searched was collected 6 weeks earlier from December 5, 2019, to explore patterns before the occurrence of the first COVID-19 case in South Korea. Data were collected using COVID-19–related terms, including coronavirus (코로나 바이러스), coronavirus test (코로나 바이러스 테스트), Middle East respiratory syndrome (MERS; 메르 스), face mask (마스크), social distancing (사회적 거리두기), and Shinchoenji (신천지) in the Korean language, and data were retrieved according to time, gender, age groups, types of device, and location. These keywords were used to represent online information searches for COVID-19–related information, personal protective measures, and preventive approaches. Specific keywords for MERS (메르스) were used to assess whether there was an increase of information searches in the early stage of the outbreak using specific terms related to MERS as reported in previous research [21]. In addition, the Shinchoenji (신천지) keyword was also used to collect online information searches following a cluster in the Shinchoenji church and to define whether this cluster induced a surge of online information searches. For terms that were more than one word, quotes were used to increase the accuracy of data in both GT and NAVER as suggested in an earlier GT research framework [22]. The health category and web search option for GT queries were also used.

Online search data retrieved from GT and NAVER are presented as a relative number called the RSVs that ranges from 0 to 100.

The RSVs represent search requests made to those search engines. For GT, the RSVs for a specific term are normalized according to the corresponding time and location [23]. GT RSVs can be downloaded for different times and locations [19], while NAVER provides queries for various times, genders, ages, and types of device categories [20].

Statistical Analysis

Analyses of health risk perceptions toward COVID-19 were performed using data from January 20 to March 22, 2020. This time frame was selected since this study aims to explore patterns of internet searches representing health risk perceptions in the initial weeks of the outbreak. Data were analyzed in a single graphical form to explore trends in new COVID-19 cases, numbers of tests, and internet searches on a daily basis. Time-lag correlations calculated by Spearman rank correlation coefficients were employed to assess whether correlations of new COVID-19 cases with GT and NAVER RSVs were affected by time within 3 days of a lag or lead period. Statistical analyses were performed using Stata 13 (StataCorp), and strong correlations were defined as correlation coefficients $r>0.7$. Moreover, multilayer maps created using Tableau Public 2020 (Tableau Software, Inc) were generated to define the distributions of new COVID-19 cases and internet searches.

This study also undertakes the task of predicting new COVID-19 cases. Several predictors, including the number of COVID-19 cases, tests, and GT and NAVER RSVs in lag periods (of 1-3 days) were used to predict the target variable, which was the number of new COVID-19 cases. The prediction value was calculated using single and multiple linear regressions employing backward elimination and a variance inflation factor (VIF) of <5 in Stata 13. A lower VIF level was considered to minimize the presence of multicollinearity in the model, particularly in epidemiologic studies [24]. Models were constructed using the development data set (January 20 to March 22, 2020) as used in health risk perception analyses and validated using the future validation data set (March 23 to May 31, 2020). The root mean squared error (RMSE) was assessed for evaluating the models' performances, as well as Akaike information criterion (AIC) for selecting a correct model and Bayesian information criterion (BIC) for finding the best model for future predictions [25].

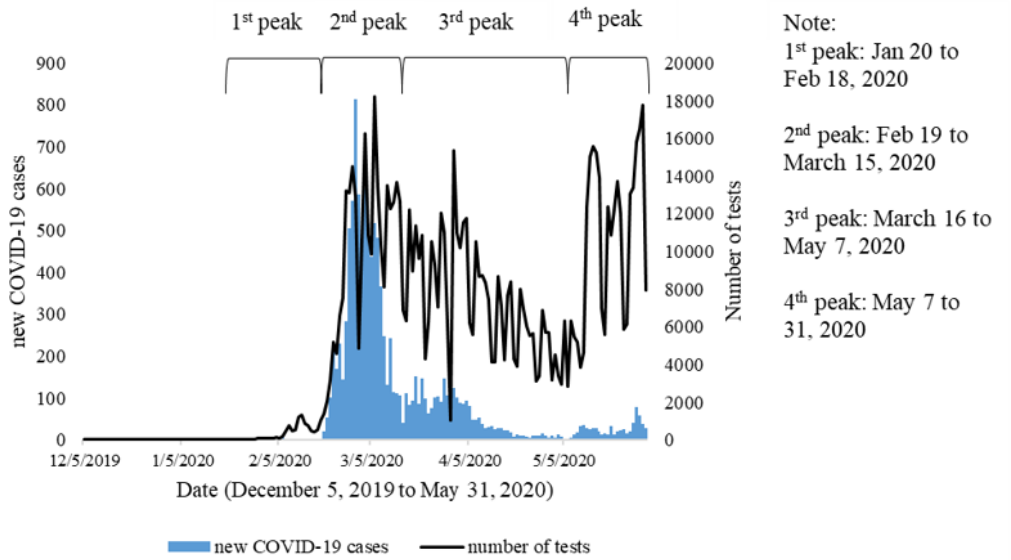
Results

Community health risk perceptions captured by GT and NAVER RSVs were divided into several parts including patterns by time, population characteristics, and location.

Trends in New COVID-19 Cases, Number of Tests, and Internet Searches on a Daily Basis

South Korea reported the first case of COVID-19 on January 20, 2020 (Figure 1), with four peaks of disease transmissions as of May 31, 2020. The first peak occurred until February 18, 2020. The average new cases increased to 311 per day and decreased to 50 cases per day since March 16, 2020. The fourth peak was observed on May 8, 2020, which corresponded with implementation of a new normal starting on May 6, 2020 [26]. Furthermore, as of May 31, 2020, South Korea had reported 11,468 cases of COVID-19. Large numbers of tests were also performed during the outbreak. South Korea performed 6848 tests on average per day from January 20 to May 31, 2020, and 910,822 tests in total, making South Korea one of the countries with the highest number of tests performed.

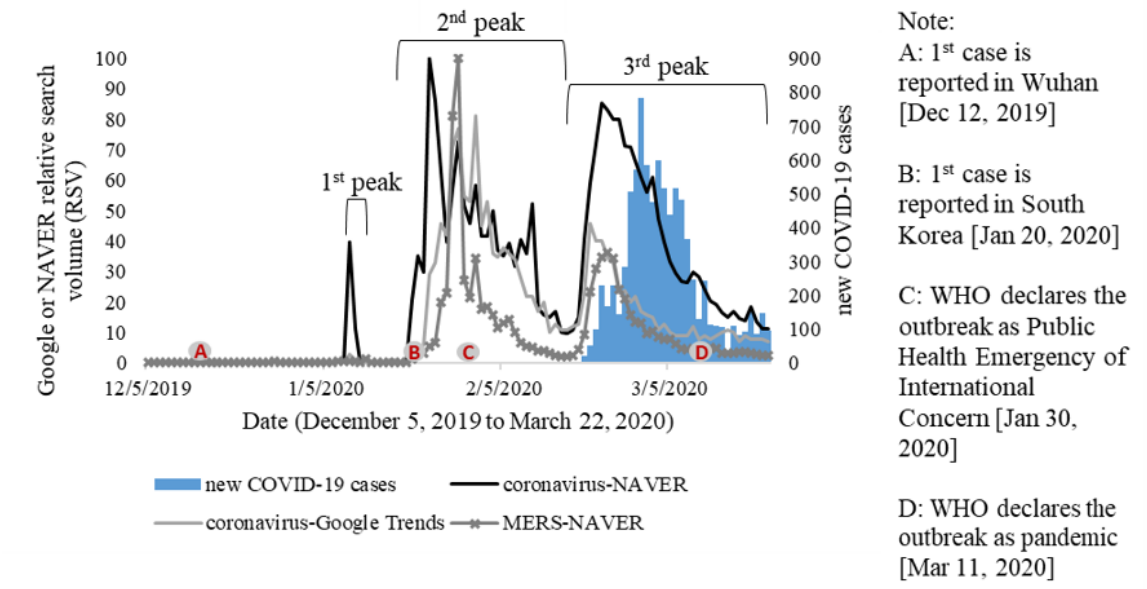
Figure 1. Time series of new COVID-19 cases and number of tests in South Korea.



During the outbreak, trends of information searches for coronavirus (코로나 바이러스) captured by GT and NAVER were similar (Figure 2). Three peaks of internet searches were observed in the second and fifth weeks of January and in the fourth week of February 2020. Coronavirus-related searches remained high for several days after the first COVID-19 case was reported in Wuhan on December 12, 2019, along with MERS (메르 스)–related queries, which were also elevated in the last two peaks. However, massive surges of information searches occurred along with the identification of the first COVID-19 case in South Korea on January 20 and with the WHO's declaration of the Public Health Emergency of

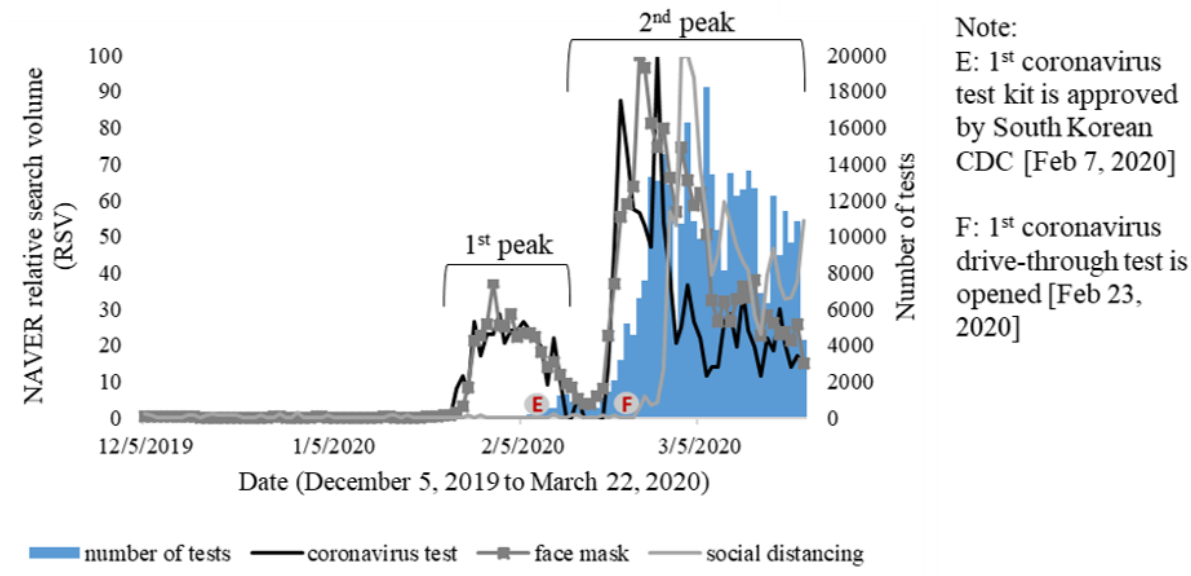
International Concern (PHEIC) on January 30, 2020. Compared to the daily data on new COVID-19 cases, information searches provided by GT and NAVER peaked 7-9 days earlier. The third peak of coronavirus searches possibly corresponded to the immense increase in the number of new COVID-19 cases due to local transmission. Searches gradually decreased even after the outbreak was declared a pandemic by the WHO on March 11, 2020 [1].

Figure 2. Time series of new COVID-19 cases and Google Trends and NAVER relative search volumes related to the coronavirus and MERS in South Korea. MERS: Middle East respiratory syndrome; WHO: World Health Organization.



Furthermore, coronavirus test–related (코로나 바이러스 테스트) searches were not captured in GT; hence, Figure 3 only illustrates NAVER RSVs related to coronavirus tests, face masks, and social distancing. Increases in internet searches were observed weeks after the COVID-19 cases were reported and before a coronavirus test kit was approved on February 7, 2020 [27]. The second wave of information searches was found in the third week of February 2020, which might have been caused by an increase in the number of new COVID-19 cases and the implementation of coronavirus drive-through tests on February 23, 2020 [6]. However, patterns of coronavirus test–related searches seemed more similar to trends of new COVID-19 cases compared to the daily numbers of tests.

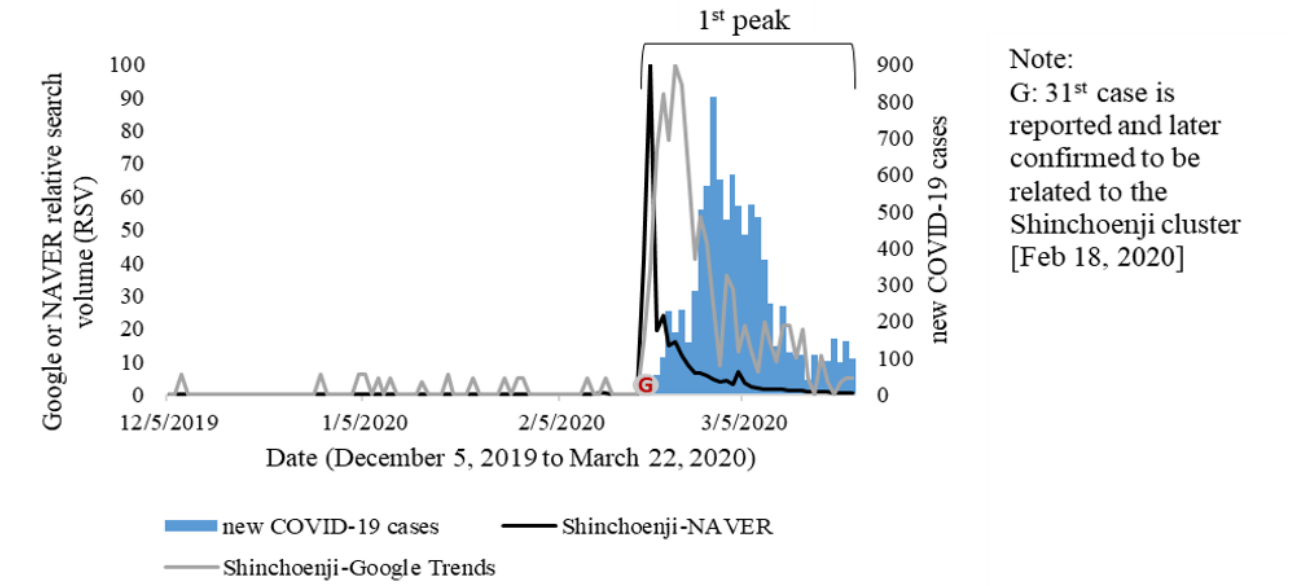
Figure 3. Time series of the daily number of coronavirus tests and NAVER relative search volumes related to coronavirus tests, face masks, and social distancing in South Korea. CDC: Centers for Disease Control and Prevention.



Similar patterns of online queries about coronavirus tests were also identified for face masks (마스크). From the perspective of personal protective measures, the number of face mask–related queries increased in the same period when people began to search for coronavirus tests and face mask shortages in early February [28] and gradually declined in late February, as a regular supply of face masks was provided by the federal government [29]. Moreover, the massive increase in locally

acquired cases also induced internet searches related to social distancing (사회적 거리두기) as one of the preventive approaches. Those searches reached a peak as a widespread campaign for social distancing was commended in the first week of March 2020 in South Korea [5]. In contrast, the number of Shinchoenji (신천지)–related searches increased as the Shinchoenji cluster was discovered on February 18, 2020 [30], and gradually decreased thereafter, even before the surge in new COVID-19 cases peaked on February 29, 2020 (Figure 4).

Figure 4. Time series of new COVID-19 cases, Google Trends, and NAVER relative search volumes related to the Shinchoenji cluster in South Korea.



Time-Lag Correlations Between new COVID-19 Cases and Internet Searches in Different Gender and age Groups

The results in Tables 1 and 2 demonstrated a moderate correlation (r=0.628) between new COVID-19 cases and GT RSVs related to coronavirus with a lag of 3 days. On the contrary, a strong correlation (r=0.718) of coronavirus information searches counting for both men and women with a lag of 3 days showed no differences for NAVER RSVs. However, the correlations varied across different age groups and lag periods. Strong correlations were observed with a lag of 3 days for all ages (r=0.729) and those aged ≤18 years (r=0.821), 19-24 years (r=0.784), 25-29 years (r=0.726), 50-54 years (r=0.706), and ≥50 years (r=0.725). Meanwhile, the weakest correlation was found in the age group of 35-39 years (r=0.622). The ≤18 years and 19-24 years age groups for NAVER RSVs had strong correlations in almost all lag and lead periods. Moreover, the strength of the correlations decreased in the lead period or a few days after the number of new COVID-19 cases increased for both GT and NAVER RSVs. Compared to NAVER RSVs, GT RSVs for coronavirus had weaker correlations with new COVID-19 cases.

Table 1. Time-lag correlation coefficients between new COVID-19 cases, Google Trends, and NAVER relative search volumes related to the coronavirus in South Korea.

Day	Google Trends	NAVER											
		Gender		Age groups (years)									
		Men	Women	Overall	≤18	19-24	25-29	30-34	35-39	40-44	45-49	50-54	≥55
–3 days													
r	0.628[a]	*0.718* [a,b]	*0.718* [a]	*0.729* [a]	*0.821* [a]	*0.784* [a]	*0.726* [a]	0.661[a]	0.622[a]	0.648[a]	0.685[a]	*0.706* [a]	*0.725* [a]
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
–2 days													
r	0.605	0.684	0.684	0.694	*0.805*	*0.759*	0.696	0.621	0.581	0.607	0.655	0.680	0.693
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
–1 day													
r	0.590	0.670	0.670	0.681	*0.812*	*0.759*	0.678	0.601	0.561	0.593	0.638	0.662	0.682
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
0 days													
r	0.576	0.654	0.654	0.663	*0.803*	*0.737*	0.659	0.578	0.538	0.565	0.606	0.634	0.655
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
1 day													
r	0.554	0.647	0.647	0.661	*0.794*	*0.736*	0.660	0.579	0.536	0.560	0.606	0.633	0.658
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
2 days													
r	0.505	0.591	0.591	0.606	*0.759*	0.688	0.600	0.513	0.477	0.508	0.554	0.580	0.606
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
3 days													
r	0.491	0.579	0.579	*0.597*	*0.749*	0.682	0.587	0.500	0.468	0.498	0.537	0.565	0.592
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

[a]Strongest correlation for each column.

[b]Italics represent a strong correlation with $r>0.7$.

Table 2. Time-lag correlation coefficients between new COVID-19 cases, Google Trends, and NAVER relative search volumes related to the coronavirus test in South Korea.

Day	Google Trends	NAVER											
		Gender		Age groups (years)									
		Men	Women	Overall	≤18	19-24	25-29	30-34	35-39	40-44	45-49	50-54	≥55
–3 days													
r	N/A[a]	*0.739* [b]	*0.769*	*0.770*	0.595[c]	0.681	0.654	*0.701*	*0.734*	0.696	0.624	0.612[c]	0.441
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
–2 days													
r	N/A	*0.769*	*0.790*	*0.797*	0.505	0.650	0.687	*0.752*	*0.786*	0.692	0.673[c]	0.581	0.445
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
–1 day													
r	N/A	*0.795* [b,c]	*0.799*	*0.824*	0.500	*0.725* [c]	0.645	*0.775*	*0.826* [c]	*0.704*	0.630	0.532	0.434
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
0 days													
r	N/A	*0.778*	*0.799*	*0.812*	0.542	*0.720*	0.653	*0.746*	*0.783*	*0.755* [c]	0.559	0.551	0.358
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
1 day													
r	N/A	*0.775*	*0.823* [c]	*0.828* [c]	0.508	0.682	0.688[c]	*0.786* [c]	*0.814*	*0.718*	0.586	0.557	0.450[c]
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
2 days													
r	N/A	*0.756*	*0.802*	*0.805*	0.549	0.620	0.623	*0.774*	*0.762*	*0.731*	0.586	0.537	0.433
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
3 days													
r	N/A	*0.744*	*0.763*	*0.781*	0.465	0.572	0.606	0.694	*0.756*	0.633	0.633	0.518	0.424
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

[a]N/A: not applicable.

[b]Italics represent strong correlations with r>0.7.

[c]Strongest correlation for each column.

Different patterns were noted in coronavirus test–related searches. No correlation could be calculated for GT RSVs due to the insufficient number of queries recorded. Strong correlations were found with a lag of 1 day for men (*r*=0.795) and a lead of 1 day for women (*r*=0.823) for NAVER RSVs, as well as for all age groups with a lead of 1 day (*r*=0.828). Moreover, weak to strong correlations were reported in different age groups. The 19-24 years age group had a strong correlation (*r*=0.725) with a lag of 1 day, followed by the 30-34 years age group (*r*=0.786 with a lead of 1 day), 35-39 years age group (*r*=0.826 with a lag of 1 day), and 40-44 years age group (*r*=0.755 with a lag of 0 days).

Trends in Online Information Searches Based on the Type of Device Used for Accessing the Internet

Figures 5 and 6 show trends of online information searches for coronavirus and coronavirus tests using mobile devices and desktops. Mobile search queries for coronavirus were higher in all peaks of information searches. For coronavirus test–related searches, mobile searches seemed to be more frequent and stable than those of desktop searches in all peaks.

Figure 5. Time series of new COVID-19 cases and NAVER relative search volumes related to the coronavirus in South Korea.

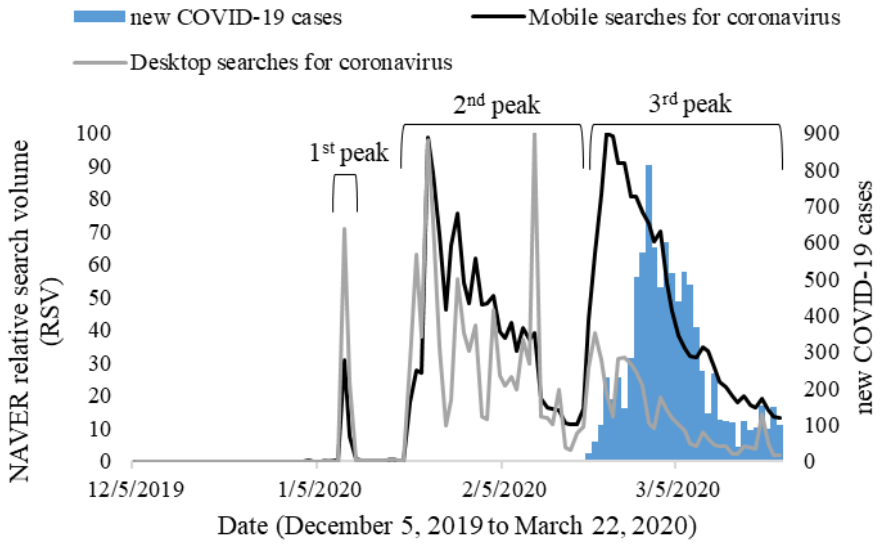
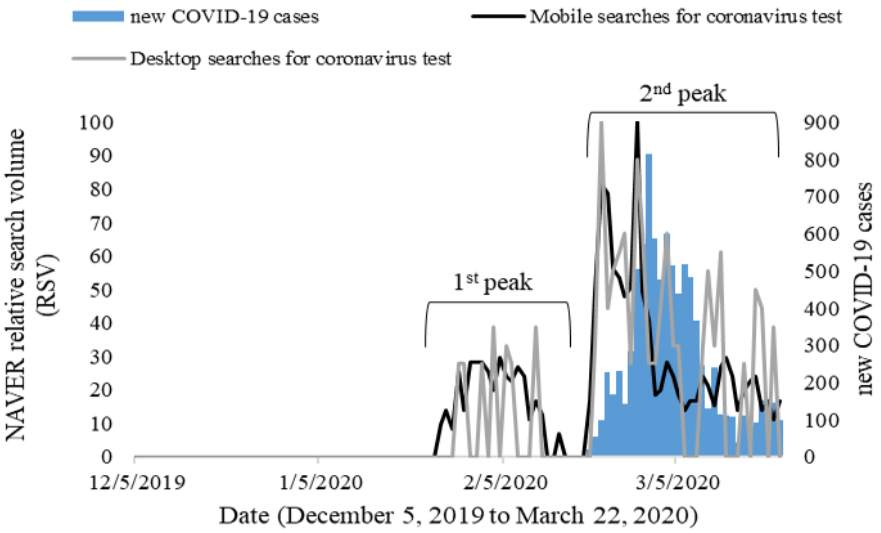


Figure 6. Time series of new COVID-19 cases and NAVER relative search volumes related to the coronavirus test in South Korea.



Spearman rank correlation coefficients in Table 3 demonstrated strong correlations for the overall data set (mobile and desktop searches) of coronavirus searches with a lag of 3 days (r=0.729), as well as mobile searches (r=0.761). Interestingly, mobile searches had stronger correlation coefficients for all lag and lead periods than did overall searches. However, weak to moderate correlations (r=0.417-0.546) were observed for coronavirus-related searches through desktop devices. For coronavirus test online searches, strong correlations (r=0.770-0.828) were reported for all lag and lead days. Still, mobile searches were observed to have a stronger correlation coefficient than desktop searches. The strongest correlations were found with a lag of 0 days for mobile searches (r=0.804) and with a lag of 1 day for desktop searches (r=0.717).

Table 3. Time-lag correlation coefficients between new COVID-19 cases and NAVER relative search volumes related to the coronavirus and coronavirus test in South Korea.

Day	Coronavirus searches (type of device)			Coronavirus test searches (type of device)		
	Overall	Mobile	Desktop	Overall	Mobile	Desktop
–3 days						
r	*0.729* [a,b]	*0.761* [a]	0.546[a]	*0.770*	*0.756*	0.677
P value	<.001	<.001	<.001	<.001	<.001	<.001
–2 days						
r	0.694	*0.726*	0.534	*0.797*	*0.787*	0.657
P value	<.001	<.001	<.001	<.001	<.001	<.001
–1 day						
r	0.681	*0.720*	0.497	*0.824*	*0.799*	*0.717* [a]
P value	<.001	<.001	<.001	<.001	<.001	<.001
0 days						
r	0.663	*0.704*	0.461	*0.812*	*0.804* [a]	0.638
P value	<.001	<.001	<.001	<.001	<.001	<.001
1 day						
r	0.661	0.692	0.475	*0.828* [a]	*0.804*	*0.705*
P value	<.001	<.001	<.001	<.001	<.001	<.001
2 days						
r	0.606	0.650	0.417	*0.805*	*0.788*	0.654
P value	<.001	<.001	<.001	<.001	<.001	<.001
3 days						
r	0.597	0.633	0.423	*0.781*	*0.761*	0.626
P value	<.001	<.001	<.001	<.001	<.001	<.001

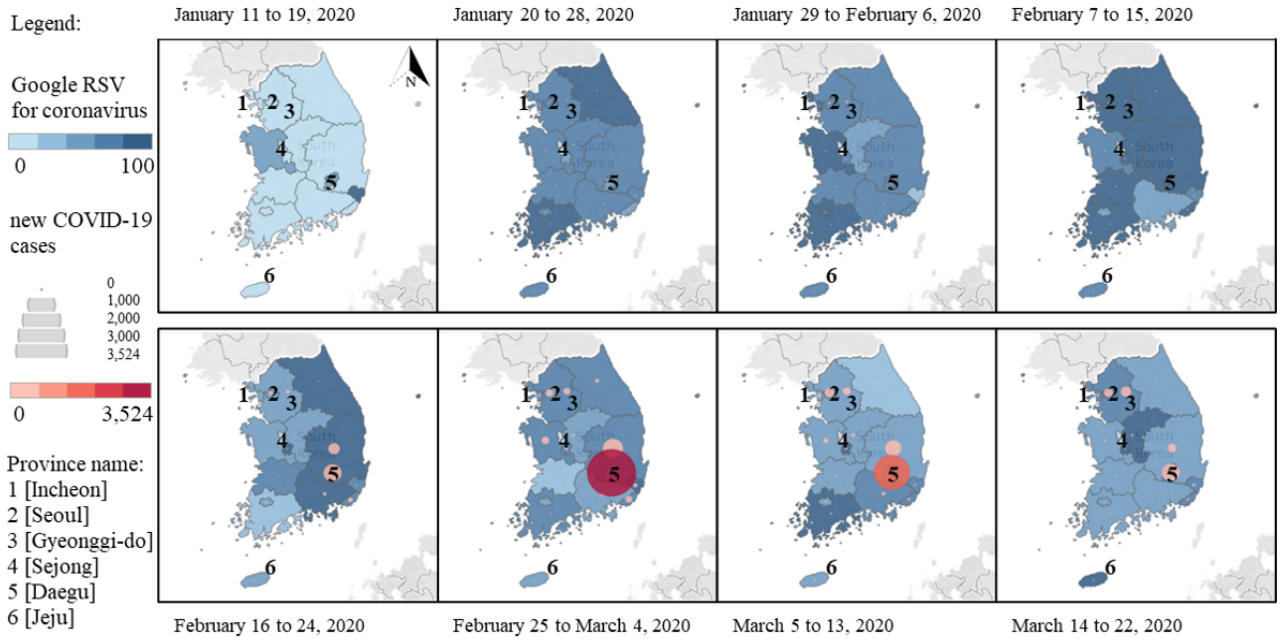
[a]Strongest correlation for each column.

[b]Italics represent a strong correlation with $r>0.7$.

Distributions of new COVID-19 Cases and Internet Searches

Spatial distributions of new COVID-19 cases and GT RSVs are illustrated in Figure 7. Results showed that 9 days before confirmed cases were reported in South Korea, the numbers of GT RSVs related to the coronavirus captured in Gyeonggi-do, Seoul, Chungcheongnam-do, Daegu, and Ulsan Provinces increased. Thereafter, the aforementioned provinces reported COVID-19–confirmed cases. During the early weeks of disease transmission (as of February 15, 2020), COVID-19 had spread in Seoul, Incheon, Gwangju, Gyeonggi-do, and Jeollabuk-do (Figure 7). Similar patterns were also captured for GT RSVs, which seemed to be elevated in those periods in the western part of South Korea where confirmed cases were reported.

Figure 7. Distribution of new COVID-19 cases and Google Trends RSVs in South Korea. RSV: relative search volume.



Furthermore, a surge in new COVID-19 cases began on February 19, 2020. GT RSVs gradually increased during that period in the eastern part of South Korea, including Daegu, the epicenter of local transmission. Daegu contributed 71.79% of confirmed cases or 262.14 cases per 100,000 population as of March 22, 2020 [31], and had a higher estimated death rate than the national rate [32]. Interestingly, increases in the number of online searches were observed a week before those massively expanding cases in provinces surrounding Daegu. The large numbers of locally acquired cases were reported from February 25 to March 4, 2020, and swiftly declined in mid-March. When the number of new cases decreased, the number of internet searches in the western part of South Korea began to increase, which indicated an elevation in the number of COVID-19 cases in the latter part of the study period.

Predicting new COVID-19 Cases

Three different models for predicting new COVID-19 cases were established in this study (Table 4). New COVID-19 cases with a lag of 1 day, number of COVID-19 tests with lags of 2 days and 1 day, GT coronavirus searches with a lag of 1 day, and NAVER coronavirus searches with a lag of 3 days were selected as important predictors for the models. Model 1 showed high performance, which indicates that this model represented 89% of new COVID-19 cases in contrast with model 2, which only represented 35% of cases as shown in the adjusted r^2 values. By combining those two models (a case-based model and internet search data–based model), the model's performance seemed to have slightly increased to nearly 90%, resulting in the lowest RMSE as observed in model 3.

Table 4. Prediction model of new COVID-19 cases in South Korea.

Models and predictors	Coef[a] (95% CI)	*P* value for *F* test	Adjusted r^2	RMSE[b]	AIC[c]	BIC[d]
Model 1 (predictors included new COVID-19 cases and number of COVID-19 tests)		<.001	0.891	54.348	1851.326	1864.03
New COVID-19 cases lag 1 day	0.942 (0.883 to 1.001)					
Number of tests lag 2 days	–0.004 (–0.007 to –0.001)					
Number of tests lag 1 day	0.004 (0.001 to 0.007)					
Cons[e]	3.957 (–5.415 to 13.329)					
Model 2 (predictors included GT[f] and NAVER RSVs[g] related to coronavirus)		<.001	0.354	133.802	2153.293	2162.805
GT RSVs lag 1 day	–0.964 (–1.604 to –0.324)					
NAVER RSVs lag 3 days	3.583 (2.859 to 4.308)					
Cons	28.920 (4.338 to 53.503)					
Model 3 (predictors included new COVID-19 cases, number of tests, and GT and NAVER RSVs related to coronavirus)		<.001	0.895	53.177	1835.169	1851.022
New COVID-19 cases lag 1 day	0.880 (0.809 to 0.951)					
Number of tests lag 2 days	–0.004 (–0.006 to –0.001)					
Number of tests lag 1 day	0.004 (0.002 to 0.007)					
NAVER RSVs lag 3 days	0.536 (0.177 to 0.894)					
Cons	–4.334 (–15.136 to 6.467)					

[a]Coef: coefficient.

[b]RMSE: root mean squared error.

[c]AIC: Akaike information criterion.

[d]BIC: Bayesian information criterion.

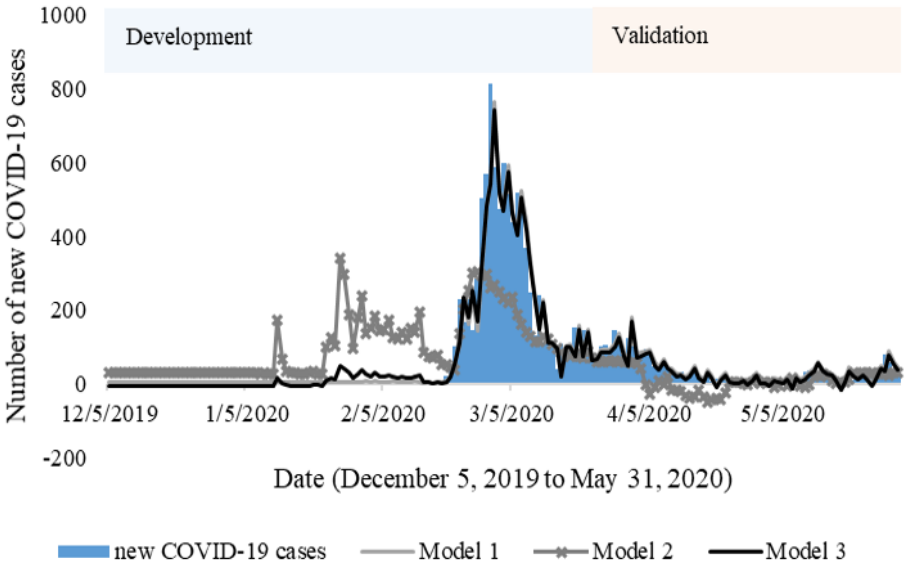
[e]Cons: constant.

[f]GT: Google Trends.

[g]RSV: relative search volume.

Models were then plotted in Figure 8 for both the development and validation sets. Model 3 performed better compared to the two other models in the development set as assessed by the value of the adjusted r^2 as well as RMSE, AIC, and BIC. In the validation set, this model also performed well, and this was indicated by the RMSE decreasing to 18.320.

Figure 8. Prediction of new COVID-19 cases in South Korea.



Discussion

Public Health Risk Perceptions

Risk perception is defined as a person's subjective judgment toward the likelihood of negative occurrences including diseases or illnesses [33]. In terms of disease outbreaks, understanding community health risk perceptions are needed in the early phase of an outbreak, particularly in the case of an emerging disease. This is because in the initial period, there will be limited treatments, few resources, and delays in active interventions [34]. Therefore, exploring the perception of risk is a necessary step in managing the risk of an outbreak. Since a robust public risk perception assessment could help in divining effective risk communication, this step should be taken immediately to reduce the impact of the COVID-19 outbreak. Consequently, it is more affordable to conduct the community health risk perception assessment using internet search data, since it can be provided more easily, promptly, and cost-effectively compared to survey methods [16] and can potentially capture anomalous patterns in real time [17]. With the widespread use of the internet and mobile devices, internet search data can be more accurate in representing the community health risk perceptions [35], as information-seeking intentions are directly affected by risk perceptions [9].

Principal Results

In this study, we found various correlations, which ranged from weak to strong, among GT and NAVER RSVs, new COVID-19 cases, and the number of tests. Previous studies also reported strong correlations between GT and NAVER RSVs compared to surveillance data [16,36]. Therefore, increased searches for COVID-19–related information might represent community health risk perceptions during local and international events. NAVER RSVs, as a local search engine that has the largest market share in South Korea (57.31% for all search categories in 2020 as of June 14) [18], seemed to be more sensitive to local issues such as coronavirus tests as shown in Figure 3. A similar result was also reported in a previous study that demonstrated that Baidu (in China) has better predictive performance for disease prediction than GT RSVs [36]. These findings suggest that NAVER RSVs could also potentially complement the use of GT RSVs, which are excessively used in the fields of infodemiology.

Patterns of community risk perceptions retrieved from information searches in this analysis were explained by examining different aspects: time, gender, age groups, types of device used for accessing the internet, and spatial distributions. Patterns according to time revealed that the number of online queries related to COVID-19 increased during local events including local transmission, approval of coronavirus test kits, implementation of coronavirus drive-through tests, a face mask shortage, a widespread campaign for social distancing, and transmission of the Shinchoenji cluster, as well as during international events such as the announcement of the PHEIC. Yet, South Korea was also one of the countries affected by the MERS epidemic [37]. That experience might have also contributed to the increased number of searches for coronavirus information even though cases had not yet been detected until then. Moreover, MERS-related searches also remained high during the study period. These findings indicated that public health risk perceptions increased following both local and international crises. Hence, risk communication should promptly be conducted, considering that health risk perceptions might change over time as the outbreak progresses.

Patterns according to time also revealed decreased numbers of GT and NAVER RSVs in the middle of the epidemic curve, which might have been caused by the extensive availability of online news and health expert reports during that period [38]. It might also have been provoked by decreased risk perceptions as the epidemic progressed [7]. Thus, using internet query data to analyze community risk perceptions could be useful in the early stage of an outbreak.

Moreover, patterns categorized by different age groups revealed that younger (≤29 years) and older age groups (≥50 years) had strong correlations of internet searches for coronavirus information with new COVID-19 cases. This finding demonstrated the high-risk perceptions of those age groups, even 3 days before an increase in the number of new COVID-19 cases locally. High-risk perceptions in younger age groups might have been induced by massive internet access for acquiring information and high numbers of confirmed cases in that age group (33.24%) in South Korea [31,39]. Meanwhile, perceived vulnerability might be common in older age groups, since an older age is one of the prominent risk factors for COVID-19 mortality [40], and 98.08% of fatal cases in South Korea occurred in older adults [31]. Additionally, a previous study showed that the older age group had higher risk perceptions [7].

In contrast, the age group of 30-49 years only showed weak to moderate correlations even 3 days before the event. This might have been due to the lower percentage of confirmed cases (23.94%) in that age group compared to that in the younger age group (≤29 years), which could also have influenced health risk perceptions. Meanwhile, online queries concerning coronavirus tests showed high-risk perceptions in the 35-44 years age group. These findings illustrate that adults perceived the coronavirus test–related information to be more important than disease-related knowledge. It might also have been influenced by the massive number of coronavirus tests conducted so far. Meanwhile, younger (aged ≤29 years) and older age groups (aged ≥50 years) had a different perception, thereby making infection-related information an essential search. In terms of gender, both men and women perceived the coronavirus as having similar levels of risk, but risk perception for coronavirus tests was higher among women. This result is similar to that reported in a previous study, which showed a higher risk perception in the women's group [7]. Hence, health risk communication should target both men and women as well as vulnerable age groups.

As to device use, patterns demonstrated that mobile device searches had stronger correlations with COVID-19–related searches compared to desktop queries. Strong correlations for mobile device searches were even observed 3 days before the outbreak. However, desktop searches showed a strong correlation with a lag of 1 day, which was 2 days later compared to mobile searches. This finding implies that high-risk

perceptions stimulated an enormous number of mobile searches during the outbreak period. Identical results were also illustrated in a previous study by Shin and colleagues [16]. The widespread use of mobile devices in the digital era [35] has promoted changes in behavior from desktop to mobile device users. Therefore, the government should ensure that risk communication can be easily accessed through mobile platforms for rapid dissemination. Research findings also demonstrated that the spatial distributions of internet searches were higher in locations with new COVID-19 cases. This finding was similar to that in previous studies, which indicated that individuals in affected areas have higher risk perceptions [7,11].

Later in the analysis, we also addressed the prediction of new COVID-19 cases using three different models. Results showed that adding COVID-19–related searches provided by NAVER could increase the performance of the model compared to that of the COVID-19 case–based model. This result resembled an earlier study [17], which also found that a model's performance increased with use of internet search data from local search engines. Furthermore, in the validation set, this model performed better, which might have been caused by a longer period for querying NAVER data; therefore, trends could be adjusted better and affect the model's performance in the validation set. Hence, considering NAVER RSVs data for case prediction could be important, employing the same data set to better understand health risk perceptions is also of importance, particularly in the early stage of an outbreak.

Briefly, this study provides a depiction of community health risk perceptions toward COVID-19 in South Korea, which tended to be higher in the period of local and international events, also for women, certain age groups, and people in affected areas. During the outbreak, people were more likely to access the internet through mobile devices, which are potential channels where health risk communication can be effectively and densely disseminated. Moreover, NAVER RSVs can potentially be used for health risk perception assessments and disease prediction. This method demonstrated an easy and low-cost approach for estimating health risk perceptions during a pandemic. Since providing a rapid risk perception assessment is needed in the early stage of an outbreak, combining GT and NAVER RSVs could be beneficial for targeting risk communication in terms of time, population characteristics, and location. GT RSVs alone only revealed patterns according to time and location [41]. However, this study only explored the positive risk perceptions toward COVID-19 rather than negative risk perceptions such as psychological impacts. As multiple studies also reported increases in incidence of anxiety, depression, anger, insomnia, distress, and suicidality during the initial phase of the epidemic [42], exploring the negative risk perceptions of the COVID-19 pandemic would be important for future works.

Limitations

As online search queries might change over time, identifying the best lag time for conducting risk communication is challenging. However, using either GT or NAVER RSVs allowed flexibility in defining the time range of data queries. Thus, we can collect adequate retrospective data sets for identifying the best lag time. In addition, this analysis might be limited to specific time frames and included only two popular search engines and certain keywords, as well as was limited for positive risk perceptions. Therefore, further research that considers those aspects to improve results of the risk perception analysis is required.

Conclusions

Community health risk perceptions toward the COVID-19 outbreak in South Korea observed from GT and NAVER RSVs increased during local and international events and were higher in women, certain age groups, and in affected areas. Although NAVER RSVs tended to be more sensitive in terms of local issues, integrating GT and NAVER RSVs could potentially provide varied search patterns in terms of time, population characteristics, and location. Moreover, online searches also identified important variables in predicting epidemic curves in the initial stage of an outbreak.

Acknowledgments

This study was funded in part by the Ministry of Science and Technology (MOST) in Taiwan (grant no MOST108-2221-E-038-018 and MOST109-2221-E-038-018) and the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan (grant no DP2-108-21121-01-A-01-04) to ECYS. This work also was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT; no 2018R1C1B6001723) to ES. The sponsors had no role in the research design or contents of the manuscript for publication. Moreover, we gratefully thank Jihoo Kim and colleagues for providing open-access data on COVID-19–related information in South Korea on Kaggle.

Authors' Contributions

AH designed the study, performed the experiments, analyzed the data, and drafted and revised the manuscript. ES contributed analytical suggestions and revised the manuscript. AF analyzed the data and revised the manuscript. ECYS conceived the study, designed the experiments, and revised the manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-03-30]
2. Coronavirus disease (COVID-19): situation report – 132. World Health Organization. 2020 May 31. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200531-covid-19-sitrep-132.pdf?sfvrsn=d9c2eaef_2 [accessed 2020-06-15]
3. Kim JH, Lee JK, Kim MH, Jang SJ, Ryoo SH, In YJ, et al. Data science for COVID-19 (DS4C). Kaggle. 2020 Jan 20. URL: https://www.kaggle.com/kimjihoo/coronavirusdataset?fbclid=IwAR2e4aTLe0x1OWp45eeyf0szA9pQpoFI6pM2KPHiXSNxHIAQDP_yz4D1qIM#Weather.csv [accessed 2020-06-21]
4. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. Int J Infect Dis 2020 Apr;93:339-344 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.031] [Medline: 32198088]
5. Gallo W. South Korea tries 'social distancing' to prevent coronavirus spread. VOA News. 2020 Mar 06. URL: https://learningenglish.voanews.com/a/south-korea-tries-social-distancing-to-prevent-coronavirus-spread/5316633.html [accessed 2020-03-30]
6. Kwon KT, Ko JH, Shin H, Sung M, Kim JY. Drive-through screening center for COVID-19: a safe and efficient screening system against massive community outbreak. J Korean Med Sci 2020 Mar 23;35(11):e123 [FREE Full text] [doi: 10.3346/jkms.2020.35.e123] [Medline: 32193904]
7. Jang WM, Kim U, Jang DH, Jung H, Cho S, Eun SJ, et al. Influence of trust on two different risk perceptions as an affective and cognitive dimension during Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea: serial cross-sectional surveys. BMJ Open 2020 Mar 04;10(3):e033026. [doi: 10.1136/bmjopen-2019-033026] [Medline: 32139484]
8. What is phase 6? World Health Organization. 2009. URL: https://www.who.int/csr/disease/swineflu/frequently_asked_questions/levels_pandemic_alert/en/ [accessed 2020-04-06]
9. Hubner AY, Hovick SR. Understanding risk information seeking and processing during an infectious disease outbreak: the case of Zika virus. Risk Anal 2020 Jun;40(6):1212-1225. [doi: 10.1111/risa.13456] [Medline: 32067244]
10. Lohiniva A, Sane J, Sibenberg K, Puumalainen T, Salminen M. Understanding coronavirus disease (COVID-19) risk perceptions among the public to enhance risk communication efforts: a practical approach for outbreaks, Finland, February 2020. Euro Surveill 2020 Apr;25(13):2000317 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.13.2000317] [Medline: 32265008]
11. Yang JZ. Whose risk? Why did the U.S. public ignore information about the Ebola outbreak? Risk Anal 2019 Aug;39(8):1708-1722. [doi: 10.1111/risa.13282] [Medline: 30726559]
12. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
13. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
14. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
15. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
16. Shin S, Kim T, Seo D, Sohn CH, Kim S, Ryoo SM, et al. Correlation between national influenza surveillance data and search queries from mobile devices and desktops in South Korea. PLoS One 2016;11(7):e0158539 [FREE Full text] [doi: 10.1371/journal.pone.0158539] [Medline: 27391028]
17. Zhang Y, Yakob L, Bonsall MB, Hu W. Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local internet query data. Sci Rep 2019 Mar 01;9(1):3262. [doi: 10.1038/s41598-019-39871-2] [Medline: 30824756]
18. Market share of search engine in South Korea. Internet Trend. 2020 Jun 15. URL: http://www.internettrend.co.kr/trendForward.tsp [accessed 2020-06-15]
19. Google Trends.: Google URL: https://trends.google.com/ [accessed 2020-06-15]
20. NAVER. URL: https://datalab.naver.com/ [accessed 2020-06-15]
21. Strzelecki A, Rizun M. Infodemiological study using Google Trends on coronavirus epidemic in Wuhan, China. Int J Onl Eng 2020 Apr 08;16(04):139. [doi: 10.3991/ijoe.v16i04.13531]
22. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
23. FAQ about Google Trends data. Google. URL: https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052
24. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology (Sunnyvale) 2016 Apr;6(2):227 [FREE Full text] [doi: 10.4172/2161-1165.1000227] [Medline: 27274911]

25. Chakrabarti A, Ghosh JK. AIC, BIC and recent advances in model selection. In: Bandyopadhyay PS, Forster MR, editors. Philosophy of Statistics. Amsterdam: Elsevier; 2011:583-605.
26. Maresca T. With COVID-19 cases in decline, South Korea opens up to a new normal. United Press International. 2020 May 06. URL: https://www.upi.com/Top_News/World-News/2020/05/06/With-COVID-19-cases-in-decline-South-Korea-opens-up-to-a-new-normal/4181588747240/ [accessed 2020-06-15]
27. Normile D. Coronavirus cases have dropped sharply in South Korea. What's the secret to its success? Science.: American Association for the Advancement of Science; 2020 Mar 17. URL: https://www.sciencemag.org/news/2020/03/coronavirus-cases-have-dropped-sharply-south-korea-whats-secret-its-success# [accessed 2020-03-30]
28. Arin K. Still not enough face masks to go around. The Korea Herald. 2020 Feb 29. URL: http://www.koreaherald.com/view.php?ud=20200229000149 [accessed 2020-04-16]
29. Greenhalgh T, Howard J. Masks for all? The science says yes. fast.ai. 2020 Apr 13. URL: https://www.fast.ai/2020/04/13/masks-summary/ [accessed 2020-04-24]
30. Hancocks P, Seo Y. How novel coronavirus spread through the Shincheonji religious group in South Korea. CNN. 2020 Feb 28. URL: https://edition.cnn.com/2020/02/26/asia/shincheonji-south-korea-hnk-intl/index.html [accessed 2020-06-15]
31. Press release: the updates on COVID-19 in Korea as of 22 March. Korean Center for Disease Control and Prevention. 2020 Mar 22. URL: https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030
32. Shim E, Mizumoto K, Choi W, Chowell G. Estimating the risk of COVID-19 death during the course of the outbreak in Korea, February-May 2020. J Clin Med 2020 May 29;9(6):1641 [FREE Full text] [doi: 10.3390/jcm9061641] [Medline: 32485871]
33. Paek HJ, Hove T. Risk perceptions and risk characteristics. In: Oxford Research Encyclopedia of Communication. Oxford: Oxford University Press; 2017.
34. World Health Organization outbreak communication planning guide. World Health Organization. 2008. URL: https://www.who.int/ihr/elibrary/WHOOutbreakCommsPlanngGuide.pdf?ua=1 [accessed 2020-04-06]
35. Liang B, Scammon DL. Incidence of online health information search: a useful proxy for public health risk perception. J Med Internet Res 2013 Jun 17;15(6):e114 [FREE Full text] [doi: 10.2196/jmir.2401] [Medline: 23773974]
36. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
37. Kim H. South Korea learned its successful COVID-19 strategy from a previous coronavirus outbreak: MERS. Bulletin of the Atomic Scientists. 2020 Mar 20. URL: https://thebulletin.org/2020/03/south-korea-learned-its-successful-covid-19-strategy-from-a-previous-coronavirus-outbreak-mers/ [accessed 2020-04-19]
38. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, et al. Use of unstructured event-based reports for global infectious disease surveillance. Emerg Infect Dis 2009 May;15(5):689-695. [doi: 10.3201/eid1505.081114] [Medline: 19402953]
39. 2018 Korea internet white paper. Korea Internet and Security Agency. URL: https://www.kisa.or.kr/eng/usefulreport/whitePaper_List.jsp [accessed 2020-06-15]
40. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
41. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: a case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
42. Sher L. COVID-19, anxiety, sleep disturbances and suicide. Sleep Med 2020 Jun;70:124 [FREE Full text] [doi: 10.1016/j.sleep.2020.04.019] [Medline: 32408252]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
GT: Google Trends
MERS: Middle East respiratory syndrome
MOE: Ministry of Education
MOST: Ministry of Science and Technology
NRF: National Research Foundation of Korea
PHEIC: Public Health Emergency of International Concern
RMSE: root mean squared error
RSV: relative search volume
VIF: variance inflation factor
WHO: World Health Organization

Edited by G Eysenbach; submitted 02.05.20; peer-reviewed by SY Shin, A Mavragani; comments to author 12.06.20; revised version received 01.07.20; accepted 14.09.20; published 29.09.20.

Please cite as:
Husnayain A, Shim E, Fuad A, Su ECY
Understanding the Community Risk Perceptions of the COVID-19 Outbreak in South Korea: Infodemiology Study
J Med Internet Res 2020;22(9):e19788
URL: http://www.jmir.org/2020/9/e19788/
doi:10.2196/19788
PMID:32931446

Original Paper

Self-Reported Compliance With Personal Preventive Measures Among Chinese Factory Workers at the Beginning of Work Resumption Following the COVID-19 Outbreak: Cross-Sectional Survey Study

Yihang Pan[1,2*], MD; Yuan Fang[3*], PhD; Meiqi Xin[4], PhD; Willa Dong[5], MSPH; Liemin Zhou[6], MD; Qinghua Hou[6], MD; Fanping Li[7], MD; Gang Sun[8], MD; Zilong Zheng[1], MD; Jinqiu Yuan[1,2,9*], PhD; Zixin Wang[4*], PhD; Yulong He[10*], MD

[1]Big Data Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[2]Precision Medicine Center, Scientific Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[3]Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong, China (Hong Kong)

[4]JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[5]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]Department of Neurology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[7]Department of Endocrinology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[8]Department of Clinical Nutrition, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[9]Clinical Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[10]Center for Digestive Disease, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Yulong He, MD
Center for Digestive Disease
The Seventh Affiliated Hospital
Sun Yat-sen University
No. 628 Zhenyuan Road
Guangming (New) District
Shenzhen, 518107
China
Phone: 86 755 81206988
Email: heyulong@mail.sysu.edu.cn

Abstract

Background: Maintaining compliance with personal preventive measures is important to achieve a balance of COVID-19 pandemic control and work resumption.

Objective: The aim of this study was to investigate self-reported compliance with four personal measures to prevent COVID-19 among a sample of factory workers in Shenzhen, China, at the beginning of work resumption in China following the COVID-19 outbreak. These preventive measures included consistent wearing of face masks in public spaces (the workplace and other public settings); sanitizing hands using soap, liquid soap, or alcohol-based hand sanitizer after returning from public spaces or touching public installations and equipment; avoiding social and meal gatherings; and avoiding crowded places.

Methods: The participants were adult factory workers who had resumed work in Shenzhen, China. A stratified two-stage cluster sampling design was used. We randomly selected 14 factories that had resumed work. All full-time employees aged ≥18 years who had resumed work in these factories were invited to complete a web-based survey. Out of 4158 workers who had resumed work in these factories, 3035 (73.0%) completed the web-based survey from March 1 to 14, 2020. Multilevel logistic regression models were fitted.

Results: Among the 3035 participants, 2938 (96.8%) and 2996 (98.7%) reported always wearing a face mask in the workplace and in other public settings, respectively, in the past month. However, frequencies of self-reported sanitizing hands (2152/3035,

70.9%), avoiding social and meal gatherings (2225/3035, 73.3%), and avoiding crowded places (1997/3035, 65.8%) were relatively low. At the individual level, knowledge about COVID-19 (adjusted odds ratios [AORs] from 1.16, CI 1.10-1.24, to 1.29, CI 1.21-1.37), perceived risk (AORs from 0.58, CI 0.50-0.68, to 0.85, CI 0.72-0.99) and severity (AOR 1.05, CI 1.01-1.09, and AOR 1.07, CI 1.03-1.11) of COVID-19, perceived effectiveness of preventive measures by the individual (AORs from 1.05, CI 1.00-1.10, to 1.09, CI 1.04-1.13), organization (AOR 1.30, CI 1.20-1.41), and government (AORs from 1.14, CI 1.04-1.25, to 1.21, CI 1.02-1.42), perceived preparedness for a potential outbreak after work resumption (AORs from 1.10, CI 1.00-1.21, to 1.50, CI 1.36-1.64), and depressive symptoms (AORs from 0.93, CI 0.91-0.94, to 0.96, CI 0.92-0.99) were associated with self-reported compliance with at least one personal preventive measure. At the interpersonal level, exposure to COVID-19–specific information through official media channels (AOR 1.08, CI 1.04-1.11) and face-to-face communication (AOR 0.90, CI 0.83-0.98) were associated with self-reported sanitizing of hands. The number of preventive measures implemented in the workplace was positively associated with self-reported compliance with all four preventive measures (AORs from 1.30, CI 1.08-1.57, to 1.63, CI 1.45-1.84).

Conclusions: Measures are needed to strengthen hand hygiene and physical distancing among factory workers to reduce transmission following work resumption. Future programs in workplaces should address these factors at multiple levels.

(J Med Internet Res 2020;22(9):e22457) doi:10.2196/22457

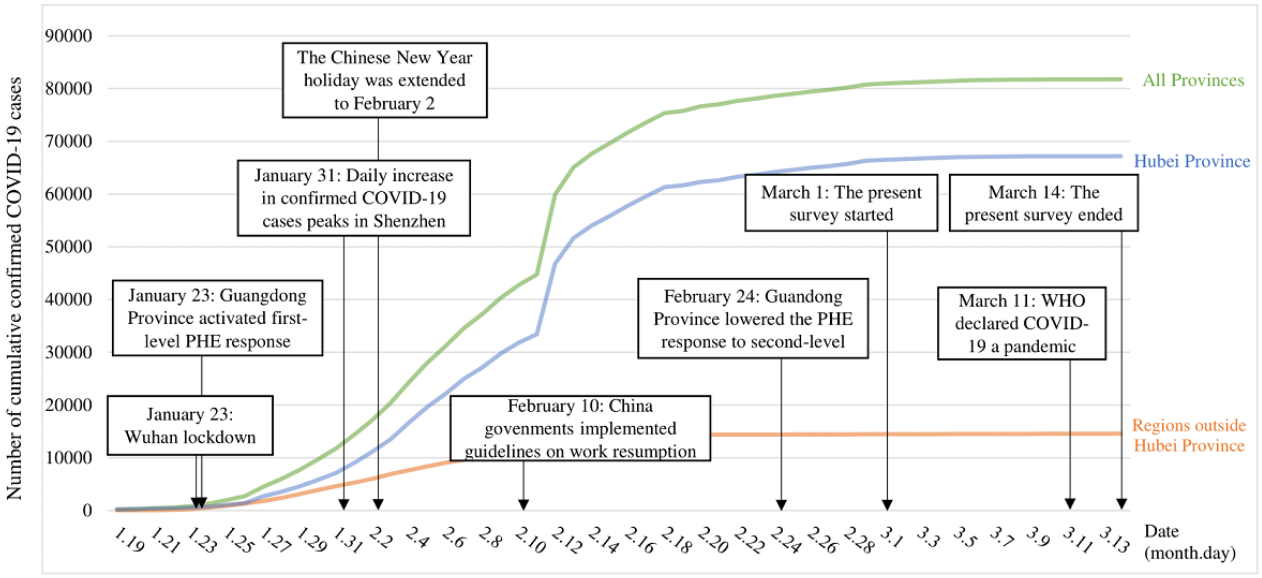
KEYWORDS

COVID-19; work resumption, factory workers; facemask wearing; hand hygiene; physical distancing; prevention; cross-sectional; online; survey; compliance

Introduction

As of July 1, 2020, 10,357,662 cases of COVID-19 and 508,055 deaths from the disease have been reported worldwide [1]. China has reported 85,232 confirmed COVID-19 cases and 4648 deaths [1]. To curb the epidemic, the Chinese government formally requested that enterprises not resume work prior to February 10, 2020, with the exception of those involved in providing basic and essential services [2,3]. These strict control measures were shown to be effective but were likely detrimental to the economy [1], as China reported a 6.8% decline in its first quarter gross domestic product in 2020 compared to the previous year [4] (Figure 1).

Figure 1. Background of the present survey, including the trend of cumulative confirmed COVID-19 cases in mainland China and critical responses to COVID-19 in Shenzhen, a city in Guangdong Province. PHE: public health emergency; WHO: World Health Organization.



In China, full work resumption is imminent. Starting on February 10, 2020, the Chinese government implemented guidelines to ensure that enterprises were adequately prepared for work resumption. Each enterprise was required to establish a comprehensive contingency plan, appoint a designated coordinator, monitor the health status of all employees and their travel history, and ensure the supply of all necessary preventive equipment [2,3]. Local governments are assessing these preparations and granting official permission for work resumption [2,3]. To scale up work resumption, official permission from local governments was no longer required as of February 20, 2020, in some Chinese cities (eg, Shenzhen) [2]. There are concerns that an increase in public contact after work resumption may result in a second wave of the COVID-19 pandemic in China [5].

Maintaining compliance with personal preventive measures plays an important role in achieving the balance between pandemic control and work resumption. Universal use of face masks [6], hand hygiene [7], and physical distancing (eg, avoiding social and meal gatherings and avoiding crowded

places) [8] are strongly advocated by the World Health Organization (WHO) and have been implemented worldwide [9,10]. The effectiveness of these personal preventive measures is crucially dependent on compliance by the public [11]. Studies conducted in China, Australia, and Thailand consistently supported that achieving very high compliance (80%-95%) with personal preventive measures has been important to control the COVID-19 pandemic in these countries [5,9,12].

Understanding factors associated with compliance with personal preventive measures is important to develop effective interventions. As interventions addressing factors at multiple levels are more likely to be successful in changing behavior, we used the socio-ecological model as the conceptual framework of our study [13]. This model considers determinants of health behaviors at the individual, interpersonal, and social-structural levels. Prior research on COVID-19 and other pandemics suggests the applicability of the socio-ecological model to inform behavioral change interventions in China. At the individual level, being knowledgeable about COVID-19 was associated with higher adoption of personal preventive measures among Hong Kong Chinese residents in the early phase of the pandemic [14]. Perceptions related to COVID-19 may also affect compliance with these personal preventive measures. For example, risk perception, perceived severity of the disease, perceived effectiveness of the preventive measures, and perceived preparedness of health systems and governments were associated with adoption of personal preventive measures during the severe acute respiratory syndrome (SARS) and H1N1 pandemics in China [15-18]. In addition, mental health status may be a particularly salient individual-level factor, as early studies in China have documented high levels of psychological problems (eg, stress, panic, depression, and anxiety) triggered by the COVID-19 pandemic [19-21]. Mental health problems have been associated with lower adoption of personal preventive measures during the COVID-19 pandemic [19]. At the interpersonal level, the heightened level of governmental alerts was accompanied by widespread coverage of COVID-19–related information across different media, including television, newspapers, and social media [14]. Additionally, different media channels may have varying effects on compliance with personal preventive measures. During the Middle East respiratory syndrome (MERS) outbreak, increased exposure to MERS-specific information through social media and interpersonal communication was associated with higher adoption of personal preventive measures. However, the association between exposure to information disseminated through traditional media (eg, television and newspapers) and personal preventive measures was nonsignificant [22,23]. At the social-structural level, implementation of organizational preventive measures during work resumption may differ across factories, which may also affect compliance with personal preventive measures.

To the best of our knowledge, no study has investigated self-reported compliance with personal preventive measures and associated factors among workers who resumed work during the COVID-19 pandemic. To address these gaps, this study investigated self-reported compliance with four personal preventive measures among a sample of factory workers in Shenzhen, China. We examined the effects of sociodemographic factors, individual-level factors (knowledge, perception, and depressive symptoms), interpersonal-level factors (exposure to COVID-19–specific information through different media), and social-structural–level factors (preventive measures implemented by the factories).

Methods

Study Design

We conducted a closed cross-sectional web-based survey of 3035 factory workers in Shenzhen, China from March 1 to 14, 2020. Of the 13 million residents in Shenzhen in 2018, 65.1% were internal migrants and 34.3% were factory workers [24].

Participants and Data Collection

By March 1, 2020, 100 factories in Shenzhen had resumed work. A stratified two-stage cluster sampling design was used to recruit the study participants. First, 14 factories were randomly selected by the research team. Of these 14 factories, 10 (71%) manufactured electronic devices, 2 (14%) manufactured watches, 1 (7%) manufactured beverages, and 1 (7%) manufactured biotechnology products. All full-time employees aged ≥18 years who had resumed work in these factories were invited to complete a web-based survey.

We developed a web-based questionnaire using Questionnaire Star, a commonly used web-based survey platform in China, and the link to the questionnaire could be shared using the WeChat social media platform. In addition to national guidelines, the Shenzhen government requested that each factory establish WeChat groups including all employees as part of the preparation for work resumption [2,3]. A designated coordinator responsible for COVID-19 control in each factory facilitated the data collection. This coordinator posted the study information and the link to access the web-based self-administered questionnaire in the WeChat group, and they invited all eligible workers who had resumed work to participate. The coordinator also sent out reminders in the WeChat groups biweekly during the recruitment period. These designated coordinators did not participate in the actual survey. The coordinators and participants were asked not to disseminate the link to access the survey to people outside the 14 selected factories. Before starting the web-based survey, the participants read a statement indicating that participation was voluntary, refusal to participate would have no effect on them, the survey would not collect personal contacts or identifying information, and the data would be kept strictly confidential and would only be used for research purposes. Web-based informed consent was obtained. Each individual WeChat account was allowed to access the web-based questionnaire once to avoid duplicate responses. The survey contained 93 items (approximately 15 items per page for 6 pages) and required approximately 20 minutes to complete. The Questionnaire Star tool performed completeness checks before the questionnaire was submitted. Participants were able to review and change their responses using a Back button. An electronic coupon for ¥10 (US $1.3) was sent to participants upon completion. All data were stored in the Questionnaire Star server and protected by a password. Only the corresponding authors had access to the database.

Ethics approval was obtained from the Seventh Affiliated Hospital, Sun Yat-sen University (reference: KY-2020-005-001).

Measures

Design of the Questionnaire

A panel consisting of two public health researchers, a health psychologist, two clinicians, a senior factory manager, and a factory worker was formed to develop the questionnaire used in the current study. The questionnaire was pilot-tested among 10 factory workers to assess its clarity and readability. These 10 workers did not participate in the actual survey. Based on the participants' comments, the panel revised and finalized the questionnaire.

Self-Reported Compliance With Personal Preventive Measures in the Past Month

Participants were asked to report the frequency at which they wore face masks in the workplace and in other public settings (public places or transportation) in the past month (response categories: every time, often, sometimes, never). A composite variable was created representing self-reported consistent wearing of a face mask in public places (referring to participants who reported always using a face mask both in the workplace and in other public settings). The participants were also asked what types of face mask they used and whether they reused their face masks. The participants also reported the frequency at which they sanitized their hands using soaps, liquid soaps, and alcohol-based sanitizers after returning from public spaces or touching public installations and equipment (eg, handrails, escalator control panels, or door knobs; response categories: every time, often, sometimes, never), and whether they avoided social meals and gatherings with people who do not live together or avoided crowded places in the past month.

Background Characteristics

Participants were asked to report sociodemographic characteristics such as age, gender, internal migrant status, highest education level, relationship status, monthly personal income, status as frontline workers or management staff, and the type of factory they worked in.

Individual-Level Variables

To assess the participants' knowledge related to transmission routes of COVID-19, a composite indicator variable was constructed by counting the number of correct responses to five knowledge items related to COVID-19 transmission routes (ranging from 0 to 5).

To assess their perceptions related to COVID-19, four scales were constructed for this study: (1) the 4-item Perceived Severity Scale, (2) the 4-item Perceived Effectiveness of Individual Preventive Measures Scale, (3) the 2-item Perceived Effectiveness of Governmental Preventive Measures Scale, and (4) the 2-item Perceived Preparedness Scale (preparedness of the health system and workplace). The response categories for these scales were 1=disagree/ineffective, 2=neutral, and 3=agree/effective. The Cronbach alpha values of these four scales ranged from .70 to .92, and single factors were identified by exploratory factor analysis (EFA) that explained 77.3% to 80.9% of the total variance. In addition, a single item was used to measure the participants' perceived risk of contracting COVID-19 in the next three months (response categories: 1=low, 2=moderate, 3=high), and another item measured the perceived effectiveness of preventive measures implemented by the factories (response categories: 1=very ineffective, 2=ineffective, 3=neutral, 4=effective, 5=very effective).

Depressive symptoms were measured by a validated Chinese version of the Patient Health Questionnaire-9 (PHQ-9) [25]. The Cronbach alpha of the PHQ-9 was .90; one factor was identified by EFA that explained 54.7% of the total variance.

Interpersonal-Level Variables

Three items were used to assess the daily average time (hours) of exposure to COVID-19–specific information through official media sources (television, newspapers, and official web-based media such as news apps or blogs and social media accounts of governmental organizations). The Exposure Through Official Media Channels Scale was formed by summing the individual item scores. The Cronbach alpha of the Exposure Through Official Media Channels Scale was .71; one factor was identified by EFA that explained 63.4% of the total variance. In addition, two single items measured the daily average time of exposure to COVID-19–specific information through unofficial media channels (individual blogs and social media accounts) and direct interpersonal communication. The response categories for the aforementioned items were 1=almost none, 2=less than 1 hour, 3=1-2 hours, 4=3-4 hours, and 5=>4 hours.

Social-Structural–Level Variables

Both the designated coordinators responsible for COVID-19 control within the sampled factories and the study participants were asked to report whether their factory had implemented seven preventive measures advocated by the Shenzhen government [2,3]. A composite indicator variable was constructed by counting the number of preventive measures implemented by the factory (ranging from 0 to 7). The items and scales measuring individual-level, interpersonal-level, and social-structural–level variables are shown in Multimedia Appendix 1.

Sample Size Planning

The target sample size was 3000. Given a statistical power of .80 and an alpha value of .05 and assuming the self-reported level of compliance with a personal preventive measure in the reference group (without a facilitating condition) to be 30%-80%, the sample size could detect a smallest odds ratio (OR) of 1.23 between people with and without the facilitating conditions (PASS 11.0, NCSS LLC). Assuming the response rate was 60%, it was necessary to invite 5000 workers to participate in the survey. The median number of workers who had resumed working in factories by the end of February 2020 was approximately 350. Therefore, the research team selected 14 factories for the study.

Statistical Analysis

Self-reported consistent face mask wearing in public spaces, sanitizing hands every time after returning from public spaces or touching public installations or equipment, avoiding social

and meal gatherings with people who do not live together, and avoiding crowded places were the dependent variables. Multilevel logistic regression models (level 1: factories; level 2: individual participants) were fit to analyze the factors associated with the dependent variables. Random intercept models were used to allow the intercept of the regression model to vary across factories, which could account for intracorrelated nested data. Multilevel logistic regression models are commonly used in studies using cluster sampling methods [26]. Univariate two-level logistic models were first used to assess the significance of the association between each of the background characteristics and the dependent variables. Background characteristics with $P<.05$ in the univariate analysis were adjusted in the multivariate two-level logistic regression models. In addition, principal component analysis with varimax rotation was used to perform EFA [27]. SPSS version 23.0 for Windows (IBM Corporation) was used for the data analysis, and $P<.05$ was considered to be statistically significant.

Results

Background Characteristics

Of 4158 workers (between 90 and 835 across different factories) who had resumed work in the selected factories on March 1, 2020, 3035 completed the web-based survey (between 56 and 635 participants across different factories); the overall response rate was 73.0%. Over half the 3035 participants were aged ≤30 years (1552, 51.1%), male (1612, 53.1%), internal migrants (2956, 97.4%), married (1812, 59.7%), had not received tertiary education (2004, 66%), had a monthly income lower than ¥5000 (US $714) (1538, 50.8%), were frontline workers (1847, 60.9%), and were manufacturing electronic devices (2353, 77.5%) (Table 1).

Table 1. Background characteristics of the participants (N=3035), n (%).

Characteristic	Value
Age (years)	
18-25	653 (21.5)
26-30	899 (29.6)
31-40	1195 (39.4)
>40	288 (9.5)
Gender	
Male	1612 (53.1)
Female	1423 (46.9)
Internal migrant	
Yes	2956 (97.4)
No	79 (2.6)
Relationship status	
Single	878 (28.9)
Have a stable boyfriend or girlfriend	345 (11.4)
Married	1812 (59.7)
Highest education level attained	
Junior high school or below	1163 (38.3)
Senior high school or equivalent	841 (27.7)
College or university	895 (29.5)
Postgraduate	136 (4.5)
Monthly personal income (¥)[a]	
<3000	175 (5.9)
3000-4999	1363 (44.9)
5000-6999	763 (25.1)
7000-9999	327 (10.8)
≥10,000	403 (13.3)
Type of work	
Frontline worker	1847 (60.9)
Manager	1188 (39.1)
Type of factory worked in	
Electronic device manufacturing	2353 (77.5)
Watchmaking	307 (10.1)
Beverage manufacturing	191 (6.3)
Biotechnology product manufacturing	184 (6.1)

[a]1 ¥=US $0.14 on March 1, 2020.

Self-Reported Compliance With Personal Preventive Measures in the Past Month

In the past month, 2938/3035 participants (96.8%) reported always wearing a face mask in the workplace, and 2996/3035 participants (98.7%) reported always wearing a face mask in other public settings. More than 95% of participants (2904/3035, 95.7%) reported consistently wearing a face mask in any public place. Nonsurgical grade respirators were most commonly used by participants (2073/3035, 68.3%), and 601/3035 (19.8%) reused face masks. Self-reported sanitizing of hands (2152/3035, 70.9%), avoiding social and meal gatherings (2225/3035, 73.3%) and avoiding crowded places (1997/3035, 65.8%) were less common (Table 2).

Table 3, Table 4, and Table 5 show the responses to the survey items measuring the individual-, interpersonal-, and social-structural–level variables, respectively.

Table 2. Self-reported compliance with personal preventive measures related to COVID-19 (N=3035), n (%).

Measure and responses	Value
Frequency of face mask wearing in the workplace	
Every time	2996 (98.7)
Often	33 (1.1)
Sometimes	3 (0.1)
Never	3 (0.1)
Frequency of face mask wearing in public places other than the workplace or on public transportation	
Every time	2938 (96.8)
Often	91 (3.0)
Sometimes	3 (0.1)
Never	3 (0.1)
Consistent face mask wearing in any public space	
No	131 (4.3)
Yes	2904 (95.7)
Type of face mask worn	
Surgical mask	1360 (44.8)
Nonsurgical grade respirator	2073 (68.3)
N-95 mask	801 (26.4)
Cloth mask	161 (5.3)
Reuse of face masks	
No	2434 (80.2)
Yes	601 (19.8)
Frequency of hand sanitation (using soap, liquid soap, or alcohol-based sanitizer) after returning from public spaces or touching public installations	
Every time	2152 (70.9)
Often	419 (16.8)
Sometimes	243 (8.0)
Never	131 (4.3)
avoiding social and meal gatherings with other people who do not live together	
No	810 (26.7)
Yes	2225 (73.3)
Avoiding crowded places	
No	1056 (34.8)
Yes	1997 (65.8)

Table 3. Responses to survey items measuring individual-level variables (N=3035).

Variable	Value
Knowledge about transmission route of COVID-19	
Knowledge about transmission route of COVID-19 (answered Yes), n (%)	
Contact with droplets	2871 (94.6)
Touching contaminated objects	2707 (89.2)
Direct contact with wildlife	2625 (86.5)
Contact with feces	2364 (77.9)
Contact with asymptomatic patients	2319 (76.4)
Correct responses to COVID-19 transmission route questions	
Number of correct responses to COVID-19 transmission route questions, mean (SD)	4.2 (1.3)
0 correct responses, n (%)	131 (4.3)
1 correct response, n (%)	49 (1.6)
2 correct responses, n (%)	94 (3.1)
3 correct responses, n (%)	264 (8.7)
4 correct responses, n (%)	634 (20.9)
5 correct responses, n (%)	1863 (61.4)
Perceptions related to COVID-19	
Perceived risk of contracting COVID-19 (answered High), n (%)	36 (1.2)
Perceived risk of contracting COVID-19, mean (SD)	1.3 (0.5)
Perceived consequences of COVID-19 (answered Agree), n (%)	
Permanent bodily damage to infected people	1226 (40.4)
High mortality rate of infected people	1687 (55.6)
Lack of effective treatment	1687 (55.6)
Lack of effective vaccines for prevention	1772 (58.4)
Perceived Severity Scale[a] score, mean (SD)	9.1 (2.1)
Perceived effectiveness of individual-level preventive measures (answered Effective), n (%)	
Wearing face masks	2407 (79.3)
Sanitizing hands frequently	2464 (81.2)
Household disinfection	2331 (76.8)
Avoiding gatherings	2722 (89.7)
Perceived Effectiveness of Individual Preventive Measures Scale[b] score, mean (SD)	11.1 (1.8)
Perceived effectiveness of preventive measures taken by the factory (answered Effective or Very effective), n (%)	2525 (83.2)
Perceived effectiveness of preventive measures taken by the factory, mean score (SD)	4.2 (1.0)
Perceived effectiveness of governmental preventive measures (answered Effective), n (%)	
Closure of public spaces (eg, restaurants, theaters)	2610 (86.0)
Restricting people coming in and out of Shenzhen	2583 (85.1)
Perceived Effectiveness of Governmental Preventive Measures Scale[c] score, mean (SD)	5.6 (0.9)
Perceived organizational preparedness for COVID-19 outbreak after work resumption (answered Agree), n (%)	
The factory in which you are working is well prepared for a COVID-19 outbreak after work resumption	2586 (85.2)
The medical system in Shenzhen is well prepared for a COVID-19 outbreak after work resumption	2297 (75.7)
Perceived Preparedness Scale[d] score, mean (SD)	5.6 (0.8)
Mental health status	

Variable	Value
PHQ-9[e] score, mean (SD)	2.1 (4.0)
Probable depression (PHQ-9 score ≥10), n (%)	170 (5.6)

[a]Perceived Severity Scale: 4 items, Cronbach α=0.70; 1 factor was identified by exploratory factor analysis explaining 77.3% of the total variance.

[b]Perceived Effectiveness of Individual-Level Preventive Measures Scale: 4 items, Cronbach α=.92; 1 factor was identified by exploratory factor analysis explaining 80.9% of the total variance.

[c]Perceived Effectiveness of Structural-Level Preventive Measures Scale: 2 items, Cronbach α=.85.

[d]Perceived Organizational Preparedness Scale: 2 items, Cronbach α=.76.

[e]PHQ-9: Patient Health Questionnaire-9, 9 items, Cronbach α=.90; 1 factor was identified by exploratory factor analysis explaining 54.7% of the total variance.

Table 4. Responses to items measuring interpersonal-level variables (N=3035).

Variable	Value
Daily average time of exposure to COVID-19–related information through different official media channels, n (%)	
Television	
Almost no exposure	613 (20.2)
<1 hour	1408 (46.4)
1-2 hours	607 (20.0)
3-4 hours	146 (4.8)
>4 hours	258 (8.5)
Newspapers	
Almost no exposure	1627 (53.6)
<1 hour	907 (29.9)
1-2 hours	294 (9.7)
3-4 hours	79 (2.6)
>4 hours	127 (4.2)
Official web-based media (news apps, blogs of governmental organizations)	
Almost no exposure	134 (4.4)
<1 hour	1263 (41.6)
1-2 hours	911 (30.0)
3-4 hours	258 (8.5)
>4 hours	469 (15.5)
Exposure Through Official Media Channels Scale[a] score, mean (SD)	7.0 (2.6)
Daily average time of exposure to COVID-19–related information through unofficial media channels (eg, personal blogs)	
Hours of exposure, mean (SD)	2.4 (1.1)
Almost no exposure, n (%)	543 (17.9)
<1 hour, n (%)	1436 (47.3)
1-2 hours, n (%)	571 (18.8)
3-4 hours, n (%)	185 (6.1)
>4 hours, n (%)	300 (9.9)
Daily average time of exposure to COVID-19–related information through face-to-face communication	
Hours of exposure, mean (SD)	1.9 (1.0)
Almost no exposure, n (%)	1269 (41.8)
Less than 1 hour, n (%)	1260 (41.5)
1-2 hours, n (%)	310 (10.2)
3-4 hours, n (%)	76 (2.5)
>4 hours, n (%)	121 (4.0)

[a]Exposure Through Official Media Channels Scale, 3 items, Cronbach α=.71; 1 factor was identified by exploratory factor analysis explaining 63.4% of the total variance.

Table 5. Responses to items measuring social-structural–level variables (n=3035), n (%).

Preventive measures implemented by the factory	Factory workers (answered Yes)	People responsible for COVID-19 control (answered Yes)
Mandatory 14-day quarantine for employees returning from high-risk areas	2901 (95.6)	14 (100.0)
Prohibiting nonemployees from entering the workplace	2664 (87.8)	14 (100.0)
Taking body temperature and requiring hand sanitation for all employees entering the workplace	2980 (98.2)	14 (100.0)
Providing face masks to all employees	2999 (98.8)	14 (100.0)
Requiring employees to wear face masks in the workplace	3023 (99.6)	14 (100.0)
Frequent workplace disinfection	2986 (98.4)	14 (100.0)
Setting up partitions in factory canteens	2838 (93.5)	14 (100.0)

Factors Associated With Self-Reported Compliance With Personal Preventive Measures in the Past Month

In the univariate multilevel logistic regression analysis, age, gender, education level, monthly personal income, status as frontline workers or management staff, and type of factory the participants were working in were significantly associated with self-reported compliance with one or more personal preventive measures (Table 6).

Table 6. Associations between background characteristics and self-reported compliance with different personal preventive measures.

Characteristic	Wearing a face mask consistently in any public space		Sanitizing hands every time after returning from public spaces or touching installations		Avoiding social and meal gatherings with people who do not live together		Avoiding crowded places	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)								
18-25	Reference	N/A[b]	Reference	N/A	Reference	N/A	Reference	N/A
26-30	1.31 (0.77-2.20)	.32	1.17 (0.94-1.46)	.17	1.10 (0.87-1.39)	.42	1.16 (0.93-1.44)	.19
31-40	1.29 (0.78-2.12)	.33	1.22 (0.98-1.52)	.07	1.23 (0.98-1.54)	.08	1.27 (1.02-1.57)	.03
>40	0.51 (0.29-0.91)	.02	1.34 (0.95-1.88)	.09	1.04 (0.75-1.44)	.81	1.18 (0.86-1.60)	.30
Gender								
Male	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Female	0.83 (0.58-1.19)	.31	1.20 (1.01-1.41)	.04	0.71 (0.60-0.84)	<.001	0.73 (0.62-0.85)	<.001
Internal migrants								
Yes	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
No	0.86 (0.33-2.26)	.76	1.55 (0.87-2.79)	.14	1.29 (0.72-2.31)	.39	1.40 (0.81-2.43)	.23
Relationship status								
Single	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Having a stable boyfriend or girlfriend	1.28 (0.66-2.48)	.46	1.04 (0.80-1.36)	.77	1.11 (0.83-1.48)	.50	1.13 (0.85-1.48)	.40
Married	1.10 (0.73-1.64)	.66	1.30 (1.08-1.57)	.005	1.16 (0.95-1.40)	.14	1.11 (0.93-1.33)	.26
Highest education level attained								
Junior high or below	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Senior high or equivalent	2.47 (1.53-4.01)	<.001	1.12 (0.91-1.38)	.29	1.64 (1.35-2.00)	<.001	1.77 (1.47-2.13)	<.001
College or university	2.80 (1.64-4.77)	<.001	0.94 (0.75-1.18)	.59	3.38 (2.66-4.29)	<.001	4.63 (3.70-5.80)	<.001
Postgraduate	3.69 (1.07-12.71)	.04	1.19 (0.78-1.82)	.42	28.58 (8.94-91.36)	<.001	11.50 (6.04-21.87)	<.001
Monthly personal income (¥)[c]								
<3000	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
3000-4999	0.84 (0.41-1.73)	.64	1.16 (0.83-1.64)	.39	1.27 (0.92-1.75)	.15	1.40 (1.02-1.93)	.04
5000-6999	1.82 (0.81-4.07)	.15	1.25 (0.87-1.79)	.24	1.71 (1.21-2.42)	.002	2.16 (1.54-3.03)	<.001
7000-9999	3.58 (1.21-10.59)	.02	0.89 (0.60-1.34)	.59	3.84 (2.46-5.99)	<.001	4.62 (3.05-7.02)	<.001

Characteristic		Wearing a face mask consistently in any public space		Sanitizing hands every time after returning from public spaces or touching installations		Avoiding social and meal gatherings with people who do not live together		Avoiding crowded places	
		OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
	≥10,000	2.04 (0.81-5.10)	.13	1.32 (0.89-1.97)	.17	7.36 (4.54-11.92)	<.001	8.26 (5.32-12.82)	<.001
Type of work									
	Frontline worker	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
	Manager	1.69 (1.13-2.52)	.01	1.04 (0.88-1.23)	.66	2.37 (1.96-2.86)	<.001	2.65 (2.22-3.15)	<.001
Type of factory									
	Electronic device manufacturing	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
	Watchmaking	2.06 (0.90-4.76)	.09	1.94 (1.44-2.61)	<.001	0.70 (0.55-0.91)	.006	0.63 (0.50-0.80)	<.001
	Beverage manufacturing	0.45 (0.26-0.78)	.005	0.94 (0.69-1.29)	.70	0.94 (0.67-1.30)	.70	0.76 (0.56-1.02)	.07
	Biotechnology product manufacturing	0.40 (0.24-0.69)	.001	2.15 (1.45-3.17)	<.001	1.85 (1.24-2.76)	<.001	2.30 (1.57-3.37)	<.001

[a]OR: odds ratio; crude ORs obtained from two-level logistic regression models (level 1: factories, level 2: individual participants).

[b]N/A: not applicable.

[c]1 ¥=US $0.15.

After adjusting for these significant background characteristics, knowledge about transmission routes of COVID-19 (adjusted odds ratios [AORs] from 1.16, CI 1.10-1.24, to 1.29, CI 1.21-1.37), perceived risk of contracting COVID-19 (AORs from 0.58, CI 0.50-0.68, to 0.85, CI 0.72-0.99), perceived effectiveness of individual (AORs from 1.05, CI 1.00-1.10, to 1.09, CI 1.04-1.13) and governmental (AORs from 1.14, CI 1.04-1.25, to 1.21, CI 1.02-1.42) preventive measures, and the number of preventive measures implemented by the factory (AORs from 1.30, CI 1.08-1.57, to 1.63, CI 1.45-1.84) were associated with self-reported compliance with all four personal preventive measures. Perceived preparedness for a potential outbreak after work resumption was associated with self-reported compliance with all personal preventive measures (AORs from 1.10, CI 1.00-1.21, to 1.50, CI 1.36-1.64), with the exception of consistent wearing of a face mask. Depressive symptoms were associated with consistent wearing of a facemask and self-reported sanitizing of hands (AORs of 0.96, CI 0.92-0.99, and 0.93, CI 0.91-0.94). Perceived severity of COVID-19 was associated with higher self-reported compliance with two physical distancing measures (AORs of 1.05, CI 1.01-1.09, and 1.07, CI 1.03-1.11) but not with consistent face mask wearing or sanitizing hands. In addition, the perceived effectiveness of preventive measures implemented by the factory (AOR 1.30, CI 1.20-1.41), and exposure to COVID-19–specific information through official media channels (AOR 1.08, CI 1.04-1.11) and face-to-face communication (AOR 0.90, CI 0.83-0.98) were associated with self-reported sanitizing of hands but not with other personal preventive measures (Table 7).

Table 7. Factors associated with self-reported compliance with different personal preventive measures.

Factor	Consistent face mask wearing in any public spaces		Sanitizing hands every time after returning from public spaces or touching installations		Avoiding social/meal gathering with people who do not live together		Avoiding crowded places	
	AOR[a] (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value
Individual-level variables								
Knowledge and perception								
Knowledge about transmission routes of COVID-19	1.21 (1.08-1.36)	.001	1.16 (1.10-1.24)	<.001	1.18 (1.11-1.26)	<.001	1.29 (1.21-1.37)	<.001
Perceived risk of contracting COVID-19	0.71 (0.50-0.99)	.045	0.58 (0.50-0.68)	<.001	0.85 (0.72-0.99)	.047	0.81 (0.69-0.95)	.01
Perceived severity of COVID-19	1.03 (0.95-1.12)	.46	1.03 (0.99-1.07)	.09	1.05 (1.01-1.09)	.04	1.07 (1.03-1.11)	.001
Perceived effectiveness of individual preventive measures	1.08 (1.00-1.18)	.048	1.09 (1.04-1.13)	<.001	1.06 (1.01-1.11)	.01	1.05 (1.00-1.10)	.03
Perceived effectiveness of preventive measures taken by the factories	1.00 (0.83-1.20)	.97	1.30 (1.20-1.41)	<.001	1.01 (0.92-1.10)	.87	1.00 (0.92-1.09)	.98
Perceived effectiveness of governmental preventive measures	1.21 (1.02-1.42)	.03	1.14 (1.04-1.24)	.003	1.15 (1.05-1.26)	.004	1.14 (1.04-1.25)	.004
Perceived organizational preparedness	0.92 (0.72-1.16)	.47	1.50 (1.36-1.64)	<.001	1.12 (1.02-1.24)	.03	1.10 (1.00-1.21)	.049
Mental health status								
PHQ-9	0.96 (0.92-0.99)	.02	0.93 (0.91-0.94)	<.001	1.01 (0.99-1.03)	.43	1.00 (0.98-1.02)	.66
Interpersonal-level variables								
Exposure through official media channels	1.02 (0.96-1.10)	.51	1.08 (1.04-1.11)	<.001	1.00 (0.97-1.03)	.89	1.00 (0.97-1.03)	.80
Exposure through unofficial media channels	1.03 (0.88-1.21)	.70	1.07 (0.99-1.15)	.13	0.99 (0.92-1.07)	.77	0.99 (0.92-1.06)	.72
Exposure through face-to-face communication	1.12 (0.92-1.37)	.25	0.90 (0.83-0.98)	.003	1.00 (0.92-1.09)	.99	1.02 (0.94-1.10)	.70
Social-structural–level variable								
Number of preventive measures implemented by the factory	1.30 (1.08-1.57)	.006	1.63 (1.45-1.84)	<.001	1.34 (1.19-1.51)	<.001	1.47 (1.30-1.66)	<.001

[a]AOR: adjusted odds ratio; background characteristics with $P<.05$ in the univariate analysis were adjusted in the multivariate two-level logistic regression models (level 1: factories, level 2: individual participants).

Discussion

Principal Findings

A recent study suggested that physical distancing and population behavioral changes that have a less disruptive economic impact than total lockdown can be effective in controlling COVID-19 [28]. Our study showed that both factories and workers in Shenzhen were well prepared for work resumption. The prevalence of consistent face mask wearing surged from 60% in the early phase of the COVID-19 outbreak (February 2020) [21] to over 95% in our study. Consistent face mask wearing is especially important in workplaces such as factories, where physical distancing cannot be guaranteed. It is also encouraging to see that all the sampled factories proactively implemented all preventive measures advocated by the government [2,3]. These efforts by factories and workers may contribute to effective COVID-19 control after work resumption in China [1].

However, this study highlighted issues related to personal preventive measures that should be addressed by future

interventions. First, many workers used non–surgical-grade respirators or even cloth masks, and approximately 20% (601/3035, 19.8%) had reused face masks in the past month. This is understandable, as surgical-grade masks, which provide the highest level of protection against COVID-19, were in limited supply in the early phase of the COVID-19 outbreak in China. To address the supply issue, China has rapidly increased its face mask production capacity. Second, there is a need to improve adherence to hand hygiene and physical distancing measures. Despite WHO recommendations on hand hygiene [7], only 2152 of the 3035 study participants (70.9%) always sanitized their hands. There are some possible explanations for the relatively low adoption of this preventive measure. The importance of hand hygiene may have been less emphasized than consistent face mask wearing in China during the outbreak. Moreover, there may be a lack of appropriate places for workers to sanitize their hands. Only 70% of the factory workers avoided social meals and gatherings or crowded places in the past month. Most Chinese cities enforced community lockdown in the early phase of the outbreak. Some voluntary physical distancing measures will be relaxed when this lockdown is lifted. Without strengthening of preventive measures, local infection is likely to occur.

Our findings provide empirical insights to inform intervention development and suggest the need to tailor interventions to specific groups. Male factory workers were less likely to sanitize their hands frequently but were more likely to comply with physical distancing measures. Promotion efforts should account for gender differences. More attention should be given to workers with lower education levels, as they showed lower compliance with consistent face mask wearing and physical distancing measures compared to workers with higher levels of education. Health communication messages should be straightforward and written at appropriate literacy levels. Management staff performed better in complying with personal preventive measures than frontline workers. These results may be due to the fact that unlike management staff, who primarily work in offices, frontline workers may face barriers to compliance related to their duties and working environment. It is important for factories to identify and address these barriers and enable workers to take necessary precautions. Moreover, the level of self-reported compliance with personal preventive measures varied across different types of factories. Different compositions of workers may explain some of these differences. For example, compared to electronic device manufacturers, workers in watchmaking factories reported higher compliance with hand hygiene but poorer compliance with physical distancing. This difference may be due to the higher proportion of female workers in watchmaking factories (over 70% in this study) compared to that in electronic device manufacturing facilities (approximately 50%). Therefore, interventions should be tailored to different types of factories. Interventions targeting watchmaking factories should focus on physical distancing, while those targeting beverage producers and biotechnology product manufacturers should emphasize consistent face mask wearing.

At the social-structural level, the preventive measures implemented by the sampled factories played important roles in COVID-19 prevention, as knowledge of more preventive measures implemented by the factories was positively associated with compliance with all four personal preventive measures. Some of these measures directly increase access to face masks and facilitate physical distancing (eg, establishing partitions in factory canteens). Moreover, factories can cultivate widely shared organizational norms to facilitate behavioral changes among the workers when implementing these preventive measures [29]. Factories should disseminate these measures to all workers and monitor the implementation of these preventive measures regularly during the pandemic.

Consistent with findings of previous studies, knowledge and perceptions related to COVID-19 had a strong influence on compliance with personal preventive measures [15-18]. Most workers were knowledgeable about transmission routes of COVID-19. New findings such as the risk of transmission among asymptomatic patients or possible fecal-oral transmission should be disseminated to the workers. Compared to results of other studies during the early phase of the outbreak, fewer participants perceived a high risk of contracting COVID-19, probably due to the initial control of the pandemic in China [14]. One possible explanation for the observed negative association between risk perception and compliance may be that people who were not able to comply with these behaviors would perceive higher risk.

Increasing the knowledge and perceived severity of COVID-19 and disseminating the efficacy of individual and governmental preventive measures may be useful strategies in future programs. To enhance compliance with these preventive measures, governments and factories should make their preparedness plans transparent to factory workers. The significant association between perceived effectiveness of preventive measures implemented by the factories and hand sanitation appears to support our speculation that facilities for sanitizing hands in the workplace are an important determinant. Strategically placing hand sanitizer in high-traffic locations throughout the workplace should be considered. Noncompliance with personal preventive measures may be used as a negative coping response to depressive symptoms [30]. Providing psychological support to workers during work resumption is also useful to enhance their compliance with personal preventive measures.

We also found that exposure to different types of media had differing effects on compliance with personal preventive measures, as our results showed that media exposure only influenced hand hygiene. Moreover, exposure through official media channels had a positive impact on hand hygiene, while exposure through unofficial media channels and face-to-face communication had no impact or even a negative impact on the same behavior. Previous studies suggested that the more people read newspapers and watched television reports about MERS, the more knowledge they acquired about the disease and its prevention strategies [22,23]. Compared to official media channels, which mainly report information verified by expert sources, unofficial web-based media channels and face-to-face communication can disseminate not only knowledge but also false or unverified information during a crisis. The null effects of exposure through unofficial web-based media channels may have resulted from conflicting content. The consequences of

misinformation can be long-lasting and should not be underestimated in health crisis management [31]. We speculate that hand sanitation was impacted not only by individual perceptions but also by peers' practices. Because hand sanitation was not highly prevalent, factory workers may have discouraged others from engaging in this behavior during face-to-face communication. Future studies should verify our speculation with a robust examination.

Our study was one of the first studies targeting factory workers at the beginning of work resumption during the COVID-19 pandemic. We used the socio-ecological model as a theoretical framework and examined potential associated factors at multiple levels. This study provides evidence to inform multilevel programs to strengthen compliance with personal preventive measures among factory workers. Currently, many countries are in the early stage of work resumption and are attempting to achieve a balance of economic reactivation and COVID-19 pandemic control; our findings have some reference value for these countries.

Limitations

This study has some limitations. First, policies and guidelines related to COVID-19 control are being updated rapidly in response to the quickly changing pandemic. These changes in national policies and guidelines have strong influences on self-reported compliance with personal preventive measures. For example, the National Health Commission of China updated the requirement to wear a face mask in the workplace on March 18, 2020, stating that face mask wearing is required in the workplace only when people are in close contact with others (<1 meter). Therefore, our findings are most applicable to the early phase of the COVID-19 outbreak, when strict measures were enforced, and have limited implication for the current situation in China. However, the risk of a second wave of COVID-19 still exists in China. In the case of another wave, some strict control measures are likely to be implemented again. Our findings could inform effective interventions facilitating the implementation of these strict control measures. Second, we only included factory workers in one Chinese city. Generalization should be made cautiously to individuals working in other types of enterprises or to other geographic locations in China. Third, because this study was anonymous and participants' personal contacts and identifying information were not collected, we were not able to collect information from workers who refused to participate in the study. Factory workers who refused to complete the survey may have different characteristics from the participants. Selection bias existed. Our response rate was relatively high (73.0%) compared to other web-based surveys on similar topics [19,20]. Fourth, the data were self-reported, and verification was not feasible. Recall bias may have occurred. Participants may have also overreported their compliance with personal preventive measures due to social desirability. Fifth, most items and scales used in this study were self-constructed based on those used in previous studies on SARS and H1N1 in China [15-18]. The internal reliability of these scales was acceptable; however, they may require external validation. Sixth, we arbitrarily chose the cutoffs for different age groups. Moreover, some behavioral factors that may influence personal preventive behaviors during the COVID-19 pandemic were not considered in this study, such as previous experiences with pandemics, concerns related to personal protective equipment supply, resource constraints, and comfort of adopting these preventive measures [32]. National guidelines emphasize that maintaining good ventilation in the workplace is an essential strategy for COVID-19 control [33]. Failure to consider ventilation in the workplace was another limitation of this study. Furthermore, causality could not be established, as this was a cross-sectional study.

Conclusions

Factory workers in China self-reported a very high level of compliance with consistent face mask wearing at the beginning of work resumption. However, compliance with hand hygiene and physical distancing measures should be strengthened. Strategically placing hand sanitizer in the workplace should be considered. Future studies should address multilevel factors associated with these preventive measures. Our findings have some reference value for other countries that are in the early stage of work resumption.

Acknowledgments

This study was funded by the Startup Fund of 100 Top Talents Program, Sun Yat-sen University (grant number 392012) and the National Key Research and Development Program (grant number 2018YFA0902801).

Authors' Contributions

YP and YF contributed equally as first authors. JY, ZW, and YH contributed equally as corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey items measuring individual-, interpersonal-, and social-structural–level variables in both English and Chinese.
[DOCX File , 21 KB - jmir_v22i9e22457_app1.docx]

References

1. Coronavirus disease (COVID-2019) situation reports. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-09-17]
2. A Letter to Enterprises in Shenzhen. Office of Shenzhen Headquarters for Prevention and Control of Pneumonia Epidemic Caused by Novel Coronavirus. 2020 Feb 02. URL: http://fao.sz.gov.cn/ydmh/xxgk/zyxw/202002/t20200205_18998951.htm [accessed 2020-09-17]
3. The State Council's joint prevention and control mechanism issued a notice calling for the effective strengthening of scientific epidemic prevention and control and orderly resumption of work and production of enterprises. Webpage in Chinese. Government of the People's Republic of China. 2020 Feb 09. URL: http://www.gov.cn/xinwen/2020-02/09/content_5476550.htm [accessed 2020-09-17]
4. 2020 first quarter GDP. Chinese National Bureau of Statistics. 2020 Apr 18. URL: http://www.stats.gov.cn/tjsj/zxfb/202004/t20200417_1739602.html [accessed 2020-09-17]
5. Zhang L, Shen M, Ma X, Su S, Gong W, Wang J, et al. What is required to prevent a second major outbreak of the novel coronavirus SARS-CoV-2 upon lifting the metropolitan-wide quarantine of Wuhan city, China. medRxiv 2020 Mar 30:preprint. [doi: 10.1101/2020.03.24.20042374]
6. Esposito S, Principi N, Leung CC, Migliori GB. Universal use of face masks for success against COVID-19: evidence and implications for prevention policies. Eur Respir J 2020 Jun;55(6):2001260. [doi: 10.1183/13993003.01260-2020] [Medline: 32350103]
7. Recommendations to Member States to improve hand hygiene practices to help prevent the transmission of the COVID-19 virus. World Health Organization. 2020 Apr 01. URL: https://www.who.int/publications-detail/recommendations-to-member-states-to-improve-hand-hygiene-practices-to-help-prevent-the-transmission-of-the-covid-19-virus [accessed 2020-09-17]
8. Coronavirus disease (COVID-19) advice for the public. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-09-17]
9. Wipatayotin A. 70% Thais stick to social distancing. Bangkok Post. 2020 Apr 02. URL: https://www.bangkokpost.com/thailand/general/1891290/70-thais-stick-to-social-distancing [accessed 2020-09-17]
10. Briscese G, Lacetera N, Macis M, Tonin M. Compliance with COVID-19 social-distancing measures in Italy: the role of expectation and duration. National Bureau of Economic Research. 2020 Mar. URL: https://www.nber.org/papers/w26916 [accessed 2020-09-17]
11. Maharaj S, Kleczkowski A. Controlling epidemic spread by social distancing: do it well or not at all. BMC Public Health 2012 Aug 20;12:679. [doi: 10.1186/1471-2458-12-679] [Medline: 22905965]
12. Chang S, Harding N, Zachreson C, Cliff O, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. arXiv. 2020 Mar 23. URL: https://arxiv.org/abs/2003.10218 [accessed 2020-09-17]
13. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
14. Kwok KO, Li KK, Chan HHH, Yi YY, Tang A, Wei WI, et al. Community Responses during Early Phase of COVID-19 Epidemic, Hong Kong. Emerg Infect Dis 2020 Jul;26(7):1575-1579 [FREE Full text] [doi: 10.3201/eid2607.200500] [Medline: 32298227]
15. Park J, Cheong H, Son D, Kim S, Ha C. Perceptions and behaviors related to hand hygiene for the prevention of H1N1 influenza transmission among Korean university students during the peak pandemic period. BMC Infect Dis 2010 Jul 28;10:222 [FREE Full text] [doi: 10.1186/1471-2334-10-222] [Medline: 20663229]
16. Lau JT, Griffiths S, Choi KC, Tsui HY. Avoidance behaviors and negative psychological responses in the general population in the initial stage of the H1N1 pandemic in Hong Kong. BMC Infect Dis 2010 May 28;10:139 [FREE Full text] [doi: 10.1186/1471-2334-10-139] [Medline: 20509887]
17. Lau JTF, Yang X, Tsui H, Pang E, Kim JH. SARS preventive and risk behaviours of Hong Kong air travellers. Epidemiol Infect 2004 Aug;132(4):727-736. [doi: 10.1017/s0950268804002225] [Medline: 15310175]
18. Lau JTF, Yang X, Tsui HY, Pang E. SARS related preventive and risk behaviours practised by Hong Kong-mainland China cross border travellers during the outbreak of the SARS epidemic in Hong Kong. J Epidemiol Community Health 2004 Dec;58(12):988-996. [doi: 10.1136/jech.2003.017483] [Medline: 15547057]
19. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
20. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psych 2020 Mar 06;33(2):e100213. [doi: 10.1136/gpsych-2020-100213]
21. Qian M, Wu Q, Wu P, Hou Z, Liang Y, Cowling BJ, et al. Psychological responses, behavioral changes and public perceptions during the early phase of the COVID-19 outbreak in China: a population based cross-sectional survey. medRxiv 2020 Feb 20:preprint. [doi: 10.1101/2020.02.18.20024448]

22. Seo M. Amplifying Panic and Facilitating Prevention: Multifaceted Effects of Traditional and Social Media Use During the 2015 MERS Crisis in South Korea. Journal Mass Commun Q 2019 Jul 26:1077699019857693. [doi: 10.1177/1077699019857693]
23. Ludolph R, Schulz P, Chen L. Investigating the Effects of Mass Media Exposure on the Uptake of Preventive Measures by Hong Kong Residents during the 2015 MERS Outbreak: The Mediating Role of Interpersonal Communication and the Perception of Concern. J Health Commun 2018;23(1):1-8. [doi: 10.1080/10810730.2017.1388455] [Medline: 29261430]
24. Shenzhen Statistics Yearbook (2019). Statistics Bureau of Shenzhen Municipality. 2020 Aug 04. URL: http://tjj.sz.gov.cn/attachment/0/695/695422/7971762.pdf [accessed 2020-09-21]
25. Wang W, Bian Q, Zhao Y, Li X, Wang W, Du J, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2014;36(5):539-544. [doi: 10.1016/j.genhosppsych.2014.05.021] [Medline: 25023953]
26. Han J, Jia P, Huang Y, Gao B, Yu B, Yang S, et al. Association between social capital and mental health among older people living with HIV: the Sichuan Older HIV-Infected Cohort Study (SOHICS). BMC Public Health 2020 Apr 28;20(1):581 [FREE Full text] [doi: 10.1186/s12889-020-08705-6] [Medline: 32345273]
27. Joliffe I, Morgan B. Principal component analysis and exploratory factor analysis. Stat Methods Med Res 1992;1(1):69-95. [doi: 10.1177/096228029200100105] [Medline: 1341653]
28. Cowling B, Ali S, Ng T, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020 May;5(5):e279-e288. [doi: 10.1016/s2468-2667(20)30090-6]
29. Stewart GL, Courtright SH, Barrick MR. Peer-based control in self-managing teams: linking rational and normative influence with individual and group performance. J Appl Psychol 2012 Mar;97(2):435-447. [doi: 10.1037/a0025303] [Medline: 21895352]
30. Folkman S, Chesney MA, Pollack L, Phillips C. Stress, coping, and high-risk sexual behavior. Health Psychology 1992;11(4):218-222. [doi: 10.1037/0278-6133.11.4.218]
31. Countering misinformation about COVID-19. World Health Organization. 2020 May 11. URL: https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19 [accessed 2020-09-17]
32. Face masks and coverings for the general public: behavioral knowledge, effectiveness of cloth covering and public messaging. The Royal Society. 2020 Jun 26. URL: https://royalsociety.org/-/media/policy/projects/set-c/set-c-facemasks.pdf?la=en-GB&hash=A22A87CB28F7D6AD9BD93BBCBFC2BB24 [accessed 2020-08-21]
33. Technical plans to control COVID-19 in workplace and public spaces. Webpage in Chinese. Government of the People's Republic of China. 2020 Feb 25. URL: http://www.gov.cn/xinwen/2020-02/25/content_5483078.htm [accessed 2020-08-21]

Abbreviations

AOR: adjusted odds ratio
EFA: exploratory factor analysis
MERS: Middle East respiratory syndrome
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 12.07.20; peer-reviewed by J Xu, H Li Oi- Yee, P Wicks; comments to author 15.08.20; revised version received 28.08.20; accepted 03.09.20; published 29.09.20.

Please cite as:
Pan Y, Fang Y, Xin M, Dong W, Zhou L, Hou Q, Li F, Sun G, Zheng Z, Yuan J, Wang Z, He Y
Self-Reported Compliance With Personal Preventive Measures Among Chinese Factory Workers at the Beginning of Work Resumption Following the COVID-19 Outbreak: Cross-Sectional Survey Study
J Med Internet Res 2020;22(9):e22457
URL: http://www.jmir.org/2020/9/e22457/
doi:10.2196/22457
PMID:32924947

Original Paper

VA Video Connect for Clinical Care in Older Adults in a Rural State During the COVID-19 Pandemic: Cross-Sectional Study

Kalpana P Padala[1,2], MD; Kerrie B Wilson[1], MS; C Heath Gauss[1,2], MS; Jessica D Stovall[1], MD; Prasad R Padala[1,2], MD

[1]Central Arkansas Veterans Healthcare System, North Little Rock, AR, United States

[2]University of Arkansas for Medical Sciences, Little Rock, AR, United States

Corresponding Author:
Kalpana P Padala, MD
Central Arkansas Veterans Healthcare System
2200 Fort Roots Drive
North Little Rock, AR, 72114
United States
Phone: 1 501 257 2044
Email: kalpana.padala@va.gov

Abstract

Background: The COVID-19 pandemic has accelerated the need for telehealth at home. Although the Department of Veterans Affairs is a leading provider of telehealth, disparities may exist in reaching older veterans living in rural areas. VA Video Connect (VVC) is a video conferencing app that enables veterans to connect with their health care provider via a secure and private session.

Objective: The aim of this study was to examine the capability and willingness of older veterans to participate in a VVC visit during the COVID-19 pandemic.

Methods: A cross-sectional study was conducted on older veterans (N=118) at the Central Arkansas Veterans Healthcare System. Participants were interviewed over the phone and responses to the following items were recorded: availability of internet, email, and an electronic device with a camera; veterans' willingness to complete an appointment via a VVC visit; and availability of assistance from a caregiver for those who were unable to participate in a VVC visit alone.

Results: Participants' mean age was 72.6 (SD 8.3) years, 92% (n=108) were male, 69% (n=81) were Caucasian, 30% (n=35) were African Americans, and 36% (n=42) lived in a rural location. The majority reported having access to the internet (n=93, 77%) and email service (n=83, 70%), but only 56% (n=67) had a camera-equipped device. Overall, 53% (n=63) were willing and capable of participating in a VVC visit. The availability of internet access was significantly lower in rural compared to nonrural participants (*P*=.045) and in those with or less than a high school education compared to those who pursued higher education (*P*=.02). Willingness to participate in the VVC visit was significantly lower in rural compared to nonrural participants (*P*=.03). Of the participants who reported they were able and willing to partake in a VVC visit (n=54), 65% (n=35) opted for VVC and 35% (n=19) preferred a phone visit. In total, 77% (n=27) of the scheduled VVC visits were successful.

Conclusions: Despite advances in technology, and willingness on the part of health care systems, there are some lingering issues with capability and willingness to participate in video telehealth visits, particularly among older adults residing in rural areas.

(J Med Internet Res 2020;22(9):e21561) doi:10.2196/21561

KEYWORDS

VA Video Connect; older adults; rural; COVID-19; veterans; telehealth; elderly; disparity; veteran affairs; capabillity; cross-sectional

Introduction

The COVID-19 pandemic has exposed many technological, ideological, and policy shortcomings in the telehealth transition. To ease policy burdens, the Centers for Medicare & Medicaid Services (CMS) has now expanded telehealth services for all Medicare beneficiaries while making virtual care reimbursable at the same rate as in-person visits, at least during the pandemic [1]. Governmental agencies have issued an emergency waiver suspending the requirement for compliance with the Health Insurance Portability and Accountability Act (HIPAA) and have noted that popular applications for video chats, such as Zoom video conference, Apple FaceTime, and Facebook Messenger video chat, which are not HIPAA compliant, may be used if

necessary [2]. All major medical associations have urged health care providers to implement telehealth systems [3].

The Department of Veterans Affairs (VA) is a leader in telehealth and has been seeing an uptake in telehealth across the nation [4]. In 2018 alone, VA provided more than 2.29 million telehealth episodes of care through clinical video telehealth (CVT) services [5] to roughly 12% of the veteran population. Veterans can visit a community-based outpatient clinic (ie, VA satellite clinics often located in rural areas) close to their home and are connected to providers in the main medical centers via CVT. Of the veterans who received CVT, 45% lived in rural areas and may have otherwise had limited access to VA health care. However, less than 1% received care through a telehealth modality in their home or through other non-VA locations [6]. Additionally, the COVID-19 pandemic has encouraged staying home to reduce the potential exposure to infection, making CVTs less desirable. In 2018, VA unveiled VA Video Connect (VVC), a video conferencing app for veterans and VA providers as part of its *Anywhere to Anywhere* initiative [7]. VVC is a secure and private session that allows veterans to have real-time access to their VA care team from their home using the camera on their smartphone, computer, or tablet, and an email address to connect to staff. Although fairly well received, veterans in some studies expressed concerns about errors in their care, perceived providers paid less attention to them, and stated they were less comfortable speaking up and asking questions [8]. It is important to determine if there is a subgroup of veterans that is underserved by VVC.

Currently, there is a clear and collective national will to adapt technology to provide safe medical care, but it is very important to address the potential for creating a digital divide. Older veterans residing in rural areas may be particularly vulnerable to the digital divide. Rural residents account for a quarter of US veteran population and a third of the VA caseload [9,10]. Likewise, about 47% of veterans are over the age of 65 years and are much more likely to be living in rural areas compared with their younger counterparts [6,10]. Rural older veterans face the difficult choice of risking iatrogenic COVID-19 exposure during a clinician visit or postponing needed care [11]. We hypothesized that internet availability would be lower among rural older veterans compared to nonrural older veterans. Hence, the primary objective of this study was to examine the availability of internet access and the willingness of older adults in a rural state to participate in a VVC visit during the COVID-19 pandemic. The secondary objective was to examine the characteristics of veterans that make them less likely to be able or willing to participate in such visits.

Methods

Sampling and Recruitment

A cross-sectional study was conducted over a 4-week period in the early part of the COVID-19 pandemic on older veterans who had appointments at the 6 clinics (5 geriatric clinics and 1 memory disorders clinic) of the Central Arkansas Veterans Healthcare System. All veterans scheduled for an appointment in these clinics during the study duration were invited to participate in the study.

The Institutional Review Board deemed this research to be exempt from review.

Procedures

The study was introduced by the clinician and completed by a research assistant after obtaining verbal permission. All interviews were conducted over the phone. The questionnaire was developed by clinical researchers who routinely see patients and conduct clinical research in the target population. Some of the questions were adapted from previously published work by the team [12]. Participants were not compensated for their participation.

Variables and Data Source

Demographic data, education, living situation, zip code, and comorbidities of the veterans were abstracted from the patient's electronic medical records, which currently includes ICD-10 (International Statistical Classification of Diseases and Related Health Problems–10th revision) codes. Rurality was determined by searching the Office of Rural Health's database by zip code.

A *My HealtheVet* account is a patient portal operated by VA through which veterans are able to see their laboratory results and communicate with their providers as needed. Whether or not veterans' *My HealtheVet* accounts were linked to their electronic medical records was recorded.

Information Collected During the Interview

During the interview, veterans were asked about their willingness to have their visit conducted via VVC instead of an in-person or CVT visit. As the VVC visit requires a participant to have an email account to receive the VVC link, veterans were asked if they had an electronic device with email access. If so, the type of device (eg, smartphone, tablet, or computer) and whether any of the devices had a camera attached or built into it were asked. They were also asked if they had internet access at home to complete the VVC visit and if they had ever used their *My HealtheVet* account. If veterans reported no to all the above questions, they were asked if they had a caregiver available to help them set up a VVC call. For those who reported caregiver availability, their relationship was noted.

Information Collected Upon Completion of VVC Visits

All participants who were willing to and capable of using VVC appointments were asked if they had an upcoming appointment in the next 4 weeks. If they had an upcoming appointment, their preference for VVC vs phone call was recorded. A chart review was performed to ascertain if the VVC appointment was successfully completed and to capture any reasons noted in the chart for not completing a scheduled VVC appointment.

Statistical Analysis

For demographic variables, descriptive statistics were obtained, including means, standard deviations, and percentages, as applicable. A veteran was considered capable of participating in a VVC visit if he or she had internet and email access and a device with a camera. To explore whether there was an association between internet access and willingness to participate in a VVC visit and rurality, a chi-square test or Fisher exact test was performed, as appropriate. A significance level of 5% was

used. Exploratory analyses were also performed to test an association between internet access and educational status, *My HealtheVet* portal use or clinic visit type, and VVC willingness and/or capability using chi-square tests. The statistical analyses were conducted using SAS Enterprise Guide 7.15 (SAS Institute).

Results

Demographic and Descriptive Statistics

A total of 118 older adults participated in the cross-sectional interview during the COVID-19 pandemic stay-home period. All participants approached for the interview agreed to participate. Participants' mean age was 72.6 (SD 8.3) years, 92% (n=108) were male, 68.6% (n=81) were Caucasian, and 29.7% (n=35) were African American (Table 1). The majority of the participants lived at home (n=116, 98%) and had a high school education/General Educational Development (GED) or some college/associate degree (n=79, 69%). Participants living in rural locations accounted for 36% (n=42) of the sample. In total, 40% (n=47) had their *My HealtheVet* account linked to their VA electronic medical records; of those, 81% (n=38) used their account. The most common chronic conditions included hypertension (n=96, 81%) and hyperlipidemia (n=54, 46%), followed by depression (n=38, 32%) and posttraumatic stress disorder (n=29, 25%); 26% (n=30) had either mild cognitive impairment or dementia (Table 1).

Table 1. Demographics and descriptive statistics (N=118).

Variable	Value
Age (years)	
Mean (SD)	72.6 (8.3)
Age category, n (%)	
≤64 years	16 (14)
65-74 years	62 (52)
≥75 years	40 (34)
Gender (male), n (%)	108 (92)
Race, n (%)	
Caucasian	81 (69)
African American	35 (30)
Native Hawaiian/Pacific Islander	2 (1)
Education, n (%)	
Less than high school	1 (1)
High school/GED[a]	31 (27)
Some college/associate degree	48 (42)
College/master's/doctoral degree	34 (30)
Living situation, n (%)	
Home	116 (98)
Assisted living	0 (0)
Nursing home	2 (2)
Rural	42 (36)
***My HealtheVet* linked to CPRS[b] (yes) , n (%)**	47 (40)
Used *My HealtheVet* account (n=47)	38 (81)
Comorbidities, n (%)	
Dementia	14 (12)
Mild cognitive impairment	16 (14)
Depression	38 (32)
Posttraumatic stress disorder	29 (25)
Hypertension	96 (81)
Hyperlipidemia	54 (46)
Sensorineural hearing loss	24 (20)
Macular degeneration	3 (3)
Legally blind	0 (0)

[a]GED: General Educational Development.

[b]CPRS: Computerized Patient Record System.

Baseline Characteristics of the Participants

In total, 93 out of 118 participants (77%) had internet access, and 83 (70%) participants had email access; however, only 67 (56%) participants had a device with a camera (Table 2). The majority used a desktop computer (44/83, 53%) or a smartphone (28/83, 34%) to access their email. Among those with a device, 67 out of 83 participants (81%) had a device with a camera. Of the 118 participants, 66 (56%) were capable of participating in a VVC visit and 69 (58%) expressed willingness. Approximately half the veterans (n=63, 53%) were both capable and willing to participate in an appointment via VVC. Almost all of those capable were willing to participate in the appointment via VVC (63/66, 95%). Among those not capable, 6 out of 52 participants (12%) were willing to participate in the appointment via VVC.

Table 2. Details on capability and willingness to participate in a VA Video Connect (VVC) appointment (N=118).

Variable	Participants, n (%)
Availability of internet access	91 (77)
Availability of email access	83 (70)
Availability of a device with a camera	67 (56)
Capable of participating in VVC appointments[a]	66 (56)
Willingness to participate in VVC appointments	69 (58)
Capable and willing to participate in VVC appointments	63 (53)
Type of device available (n=83)	
Smartphone	28 (34)
Tablet	6 (7)
Desktop computer	44 (53)
All above devices	5 (6)
Availability of a caregiver to help set up the VVC appointment (n=40)	23 (58)
Relationship of the caregiver (n=23)	
Spouse	13 (57)
Sibling	2 (9)
Child	7 (30)
Other	1 (4)

[a]Older adults with availability of internet access, email, and a device with camera were considered capable to do the VVC appointments.

Information Collected During the Interview

Primary Analyses

The proportion of participants who had access to the internet was significantly lower in rural veterans compared to nonrural ones (*P*=.045). The proportion of rural participants who were willing to participate in a clinic visit via VVC was significantly lower compared to nonrural participants (*P*=.03). The differences in availability of various resources and capability and willingness among rural and urban veterans are depicted in Table 3.

Table 3. Differences in the availability of various resources and capability and willingness among rural and urban veterans.

Variable	Participants		
	Rural, n (%)	Urban, n (%)	*P* value
Internet	28 (67)	63 (83)	.04[a]
Email	26 (62)	57 (75)	.14
Device with a camera	20 (47)	47 (82)	.55
Capable[b]	20 (48)	46 (61)	.18
Willing	19 (45)	50 (66)	.03[a]
Capable and willing[c]	18 (43)	45 (59)	.09

[a]*P* value <.05 is clinically significant.

[b]Capable: a veteran was considered capable of participating in a VVC visit if he or she had internet access and also had a smartphone, tablet, or computer that had both email access and a camera.

[c]Capable and willing: a veteran capable and willing to participate in a VVC visit.

Secondary and Exploratory Analyses

The proportion of participants who had internet access was significantly lower in those with high school or less education compared to those with more than high school education (*P*=.02). The proportion of participants who had access to the internet and email service was significantly lower for those who had an appointment at the memory disorders clinic (*P*=.01) compared to those who had an appointment at the other geriatric clinics (*P*=.01). The proportion of veterans who had their *My HealtheVet* account linked to their electronic medical records expressed significantly higher capability in participating in a VVC visit (*P*=.001), willingness to participate in a VVC visit (*P*=.001), and capability and willingness to participate in a VVC

visit (*P*=.003), compared to those who did not have their *My HealtheVet* account linked to their electronic medical records.

Attempts and Completion of VVC Visits

Of the 63 VVC-capable and willing veterans, 54 had upcoming visits in the aforementioned clinics within 4 weeks from the time of their interviews. Of these 54 veterans, 19 (35%) preferred their appointment over the phone and 35 (65%) preferred a VVC visit. A total of 35 VVC appointments were scheduled during the study period. Of these, 27 (77%) were successfully completed. Of the completed VVC appointments, 13 (48%) received assistance from a relative or caregiver for the appointment. In total, 6 veterans could not complete their VVC appointments; 4 had internet connectivity issues and 2 were no-shows for their appointments.

Discussion

Principal Findings

This study describes the feasibility of conducting VVC visits in a rural state. The findings from this study are encouraging, with 77% of older adults having access to the internet and 70% having email service. These findings are similar to other studies that found increasing internet use in older veterans and willingness to participate in video programs for health [11]. It is troubling, however, to see that the historically described disparities in internet and email access among rural veterans is still prevalent, albeit at a lower frequency. Education appears to have a stronger influence on internet access, although willingness to participate in a VVC visit was more strongly influenced by rurality than education. These findings suggest that older veterans with a high school or less education residing in rural area—the very consumers that need the VVC services the most—may require extra attention. The Federal Communications Commission has taken key steps to bring broadband to rural America, such as the $20-billion investment in the *Rural Digital Opportunity Fund* and $1.4-billion allocation for the *Connect America Fund Phase II Reverse Auction* to expand broadband to more than 700,000 rural locations in 45 states [13]. Future measures are needed to make internet services affordable to rural residents.

The biggest hurdle for VVC capability in this sample was the availability of a camera-enabled device; only 56% of the participants had this. It is interesting to note that almost all of those who were capable of participating in a VVC appointment were willing to do so (95%). This suggests that providing video-enabled devices could improve uptake of VVCs. VA has been giving tablets to select groups of veterans for telehealth since 2016. Tablet recipients (N=604) reported that the care they received via video consultations was equivalent to in-person visits [14]. However, in the same study, 20% of the participants did not use the tablets, and one third, those with technological difficulties and/or multiple comorbidities, preferred in-person visits over remote visits [14]. However, the survey respondents in this study were younger (mean age 56.0, SE 0.20 years) than those of the current study [14]. In our study, only 12% of those that did not have VVC capability were willing to engage in VVCs, suggesting that providing tablets, although an important first step, may not solve problems for all veterans. Nevertheless, such selected patients could be targeted for tablet provision after ensuring that an internet connection is in place.

The participants that had an appointment at the memory disorders clinic had significantly lower access to internet or email services compared to those enrolled in other clinics. This difference could be due to a cohort effect and may be ameliorated by the presence of caregivers. Home-based video programs for dementia care are pioneered by groups such as Dr Lauren Moo's team at the Bedford Geriatric Research, Education and Clinical Center, a center of excellence within VA focused on enhancing geriatric care [15]. The availability of such home-based video programs for dementia care has expanded exponentially across the nation [15]. With VA's growing emphasis on caregiver services, telehealth options for veterans with dementia could grow exponentially.

The strong association between *My HealtheVet* portal use and VVC willingness and capability is very encouraging as the use of this portal by veterans could be a proxy for VVC capability and/or willingness. This finding, if proven in larger studies, may help assess generalizability; a clinician could then simply check if a veteran is currently using the portal before making the decision to offer a VVC visit.

The results of this study should be considered in the context of several limitations. The study population consisted of US veterans, of mostly male gender, and from a single geographic area, and thus, the patient-reported responses in our results may not be generalizable to other population groups, women, or other areas that use different telehealth equipment, workflows, and processes of care. Despite these limitations, our study can be characterized by many strengths, including the systematic collection of actionable information from a group of veterans who are older and at high risk of being left out in the current digital transformation.

Conclusion

Despite advances in technology, and willingness on the part of health care systems, there are some lingering issues relating to capability and willingness to participate in video telehealth visits particularly among older adults residing in rural areas. Policy efforts to bring broadband internet access to rural residents are the necessary first steps in addressing the digital divide. Targeted educational efforts are needed to train those residing in rural areas. The distribution of video-capable devices needs to be augmented by the provision of high-speed internet and age-friendly technical assistance.

Acknowledgments

KPP was supported by the Department of Veterans Affairs RR&D merit grant I01 RX002961 and the National Institute of Health. PRP was supported by the Department of Veterans Affairs RR&D merit grant 1I01RX002638 and HSR&D grant C19 20-398 and the National Institute of Health.

The sponsor had no role in the design, methods, subject recruitment, data collections, analysis, or preparation of the paper.

Authors' Contributions

All authors have contributed significantly to the paper and approved the final version. Detailed contributions are as follows—KPP: conceptualization, recruiting participants and conducting the study, interpreting data analysis, and manuscript preparation; KBW: recruiting participants and conducting the study, interpreting data analysis, and critical review of the manuscript; CHG: data analysis and critical review of the manuscript; JDS: critical review of the manuscript; PRP: conceptualization, interpreting data analysis, and critical review of the manuscript.

Conflicts of Interest

None declared.

References

1. President Trump Expands Telehealth Benefits for Medicare Beneficiaries During COVID-19 Outbreak. Centers for Medicare and Medicaid Services. 2020 May 17. URL: https://www.cms.gov/newsroom/press-releases/president-trump-expands-telehealth-benefits-medicare-beneficiaries-during-covid-19-outbreak [accessed 2020-06-05]
2. Wright J, Caudill R. Remote Treatment Delivery in Response to the COVID-19 Pandemic. Psychother Psychosom 2020;89(3):130-132 [FREE Full text] [doi: 10.1159/000507376] [Medline: 32213775]
3. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
4. Shura RD, Brearly TW, Tupler LA. Telehealth in Response to the COVID-19 Pandemic in Rural Veteran and Military Beneficiaries. J Rural Health 2020 May 13:e [FREE Full text] [doi: 10.1111/jrh.12454] [Medline: 32402128]
5. Adams SV, Mader MJ, Bollinger MJ, Wong ES, Hudson TJ, Littman AJ. Utilization of Interactive Clinical Video Telemedicine by Rural and Urban Veterans in the Veterans Health Administration Health Care System. J Rural Health 2019 Jun;35(3):308-318. [doi: 10.1111/jrh.12343] [Medline: 30600557]
6. Department of Veterans Affairs FY 2018 - 2024 Strategic Plan. US Department of Veteran Affairs. 2019 May 31. URL: https://www.va.gov/oei/docs/VA2018-2024strategicPlan.pdf [accessed 2020-06-05]
7. VA expands telehealth by allowing health care providers to treat patients across state lines (press release). US Department of Veterans Affairs. 2018 May 11. URL: https://tinyurl.com/yxvsbpho [accessed 2020-06-05]
8. Gordon HS, Solanki P, Bokhour BG, Gopal RK. "I'm Not Feeling Like I'm Part of the Conversation" Patients' Perspectives on Communicating in Clinical Video Telehealth Visits. J Gen Intern Med 2020 Jun 3;35(6):1751-1758. [doi: 10.1007/s11606-020-05673-w] [Medline: 32016705]
9. Teich J, Ali MM, Lynch S, Mutter R. Utilization of Mental Health Services by Veterans Living in Rural Areas. J Rural Health 2017 Jun;33(3):297-304. [doi: 10.1111/jrh.12221] [Medline: 27701791]
10. Rural veterans health care atlas, 2nd edition, FY-2015. Health VOoR. 2015. URL: https://www.ruralhealth.va.gov/docs/atlas/CHAPTER_01_Overview_Location_Medical_Facilities.pdf [accessed 2020-06-05]
11. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/nejmp2005835]
12. Padala PR, Jendro AM, Padala KP. Conducting Clinical Research During the COVID-19 Pandemic: Investigator and Participant Perspectives. JMIR Public Health Surveill 2020 Apr 06;6(2):e18887 [FREE Full text] [doi: 10.2196/18887] [Medline: 32250281]
13. Bridging The Digital Divide For All Americans. Federal Communication Commission. URL: https://tinyurl.com/y3dvslau [accessed 2020-06-05]
14. Slightam C, Gregory AJ, Hu J, Jacobs J, Gurmessa T, Kimerling R, et al. Patient Perceptions of Video Visits Using Veterans Affairs Telehealth Tablets: Survey Study. J Med Internet Res 2020 Apr 15;22(4):e15682 [FREE Full text] [doi: 10.2196/15682] [Medline: 32293573]
15. Moo LR, Jafri Z, Morin PJ. Home-Based Video Telehealth for Veterans With Dementia. Fed Pract 2014 Dec;31(12):36-38 [FREE Full text] [Medline: 29398887]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
HIPAA: Health Insurance Portability and Accountability Act

CVT: clinical video telehealth
VA: Department of Veterans Affairs
VVC: VA Video Connect

Edited by G Eysenbach, G Fagherazzi; submitted 17.06.20; peer-reviewed by H Lum, B Yarns; comments to author 13.07.20; revised version received 10.08.20; accepted 14.09.20; published 30.09.20.

Please cite as:
Padala KP, Wilson KB, Gauss CH, Stovall JD, Padala PR
VA Video Connect for Clinical Care in Older Adults in a Rural State During the COVID-19 Pandemic: Cross-Sectional Study
J Med Internet Res 2020;22(9):e21561
URL: http://www.jmir.org/2020/9/e21561/
doi:10.2196/21561
PMID:32936773

Original Paper

Effects of the COVID-19 Pandemic on Obsessive-Compulsive Symptoms Among University Students: Prospective Cohort Survey Study

Guangjun Ji[1], MD; Wenjun Wei[1], MD; Kai-Chen Yue[2], MD; Heng Li[1], PhD; Li-Jing Shi[2], BSN; Jian-Dong Ma[2], MD; Chen-Yang He[1], MD; Sheng-Sheng Zhou[2], MD; Zongya Zhao[1], PhD; Tao Lou[2], MD; Jie Cheng[2], MD; Shi-Chang Yang[2], MD, PhD; Xian-Zhang Hu[2], MD, PhD

[1]Xinxiang Medical University, Xinxiang, China

[2]The Second Affiliated Hospital of Xinxiang Medical University, Xinxiang City, China

Corresponding Author:
Xian-Zhang Hu, MD, PhD
The Second Affiliated Hospital of Xinxiang Medical University
No 388, Jianshe Road
Muye District
Xinxiang City, 453002
China
Phone: 86 13938744850
Email: huxianzhang@xxmu.edu.cn

Abstract

Background: The COVID-19 pandemic is associated with common mental health problems. However, evidence for the association between fear of COVID-19 and obsessive-compulsive disorder (OCD) is limited.

Objective: This study aimed to examine if fear of negative events affects Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) scores in the context of a COVID-19–fear-invoking environment.

Methods: All participants were medical university students and voluntarily completed three surveys via smartphone or computer. Survey 1 was conducted on February 8, 2020, following a 2-week-long quarantine period without classes; survey 2 was conducted on March 25, 2020, when participants had been taking online courses for 2 weeks; and survey 3 was conducted on April 28, 2020, when no new cases had been reported for 2 weeks. The surveys comprised the Y-BOCS and the Zung Self-Rating Anxiety Scale (SAS); additional items included questions on demographics (age, gender, only child vs siblings, enrollment year, major), knowledge of COVID-19, and level of fear pertaining to COVID-19.

Results: In survey 1, 11.3% of participants (1519/13,478) scored ≥16 on the Y-BOCS (defined as possible OCD). In surveys 2 and 3, 3.6% (305/8162) and 3.5% (305/8511) of participants had scores indicative of possible OCD, respectively. The Y-BOCS score, anxiety level, quarantine level, and intensity of fear were significantly lower at surveys 2 and 3 than at survey 1 (P<.001 for all). Compared to those with a lower Y-BOCS score (<16), participants with possible OCD expressed greater intensity of fear and had higher SAS standard scores (P<.001). The regression linear analysis indicated that intensity of fear was positively correlated to the rate of possible OCD and the average total scores for the Y-BOCS in each survey (P<.001 for all). Multiple regressions showed that those with a higher intensity of fear, a higher anxiety level, of male gender, with sibling(s), and majoring in a nonmedicine discipline had a greater chance of having a higher Y-BOCS score in all surveys. These results were redemonstrated in the 5827 participants who completed both surveys 1 and 2 and in the 4006 participants who completed all three surveys. Furthermore, in matched participants, the Y-BOCS score was negatively correlated to changes in intensity of fear (r=0.74 for survey 2, P<.001; r=0.63 for survey 3, P=.006).

Conclusions: Our findings indicate that fear of COVID-19 was associated with a greater Y-BOCS score, suggesting that an environment (COVID-19 pandemic) × psychology (fear and/or anxiety) interaction might be involved in OCD and that a fear of negative events might play a role in the etiology of OCD.

(J Med Internet Res 2020;22(9):e21915) doi:10.2196/21915

KEYWORDS

COVID-19; fear; anxiety; obsessive-compulsive disorder; OCD; Yale-Brown Obsessive-Compulsive Scale; university student; mental health

Introduction

At the end of 2019, an outbreak of COVID-19, caused by the novel coronavirus SARS-CoV-2 was reported in Wuhan, China [1]. With the growing number of cases and deaths, fear and uncertainty have spread around the globe as COVID-19 continues to capture the world's attention. Since the end of January 2020, many provinces in China, including Henan Province, implemented quarantine measures, which may further instill fear of the virus in communities. At the same time, public education on disease prevention and environmental hygiene was emphasized across the country. Information outlets, such as the internet, television, radio, newspapers, cellphones, and social media (eg, WeChat), were used to disseminate advice on how to prevent infection (eg, stay at home, wear face masks, wash hands frequently, and/or sanitize hands). In addition, the rapid transmission of COVID-19, its approximately 2% fatality rate, lack of effective treatments and vaccines, and mass quarantine measures are associated with common mental health problems (eg, fear, anxiety, depression, and sleep problems) in subpopulations, including COVID-19 patients, those with close contact to infectees, the public, and health care professionals [2,3]. A study including 1210 respondents from 194 cities in China found that 54% of respondents rated the psychological impact of the COVID-19 outbreak as moderate or severe; 29% reported moderate-to-severe anxiety symptoms; and 17% reported moderate-to-severe depressive symptoms [4]. However, there is a lack of research on the effects of the COVID-19 pandemic on specific mental disorders, such as obsessive-compulsive disorder (OCD).

OCD is a chronic and debilitating mental disorder and tends to be treatment refractory. It is characterized by unwanted intrusive thoughts (obsessions) and repetitive compulsive behaviors or mental rituals (compulsions). Individuals perform compulsions in response to the distress associated with the content of the obsessions. Often of early onset (ie, before 18 years of age), OCD impacts 2% to 3% of the US population [5] and affects individuals throughout their lives, leading to a diminished quality of life for patients and their families, reduced productivity, and high health care costs. OCD accounts for 2.2% of total years lived with disability, which is approximately the same percentage as schizophrenia [6].

OCD has been reported to have several specific clinical characteristics. First, individuals manifest OCD symptoms only under certain situations that usually invoke a fear of negative events. Second, more than 90% of the general population has experienced intrusive thoughts [7]. Third, the more effort put into controlling the obsession, the more frequently and intensely it intruded the patient's mind [8]. Forth, compulsions can make intrusive thoughts become more frequent, repetitive, and disturbing [9]. Fifth, the performance of repetitive behaviors (eg, handwashing or checking) are generally related to a fear of negative events, such as fear of contamination or fear of a house catching on fire. Finally, in an OCD symptom–induced situation, the fear of negative events, obsessions, and compulsions can be considered as stressors and their effects on individuals can be neutralized when appropriate coping strategies are used [10-12]. This evidence suggests that fear of negative events is involved in symptom development and maintenance. For example, individuals with OCD repetitively check—for instance, the door or stove—due to a fear of loss of property, and spend a long time on handwashing due to a fear of contamination [13,14]. In addition, worry, disgust, and "just-not-right" sentiments can be involved in the onset of OCD [15,16]. Recently, a study in pediatric OCD found that individuals with OCD exhibit greater fear acquisition and impaired inhibitory learning compared to healthy controls [17]. However, there is lack of prospective studies on the relationship between the onset of OCD and a fear of negative events [18].

In this prospective study, we conducted surveys on students of the Xinxiang Medical University (XXMU) at three timepoints during the COVID-19 pandemic. We primarily sought to investigate whether fear of COVID-19 affects the prevalence of possible OCD based on a score of ≥16 on the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS). We also aimed to investigate the predictors for possible OCD. We hypothesized that fear of COVID-19 infection would be correlated to Y-BOCS scores.

Methods

Participants and Procedure

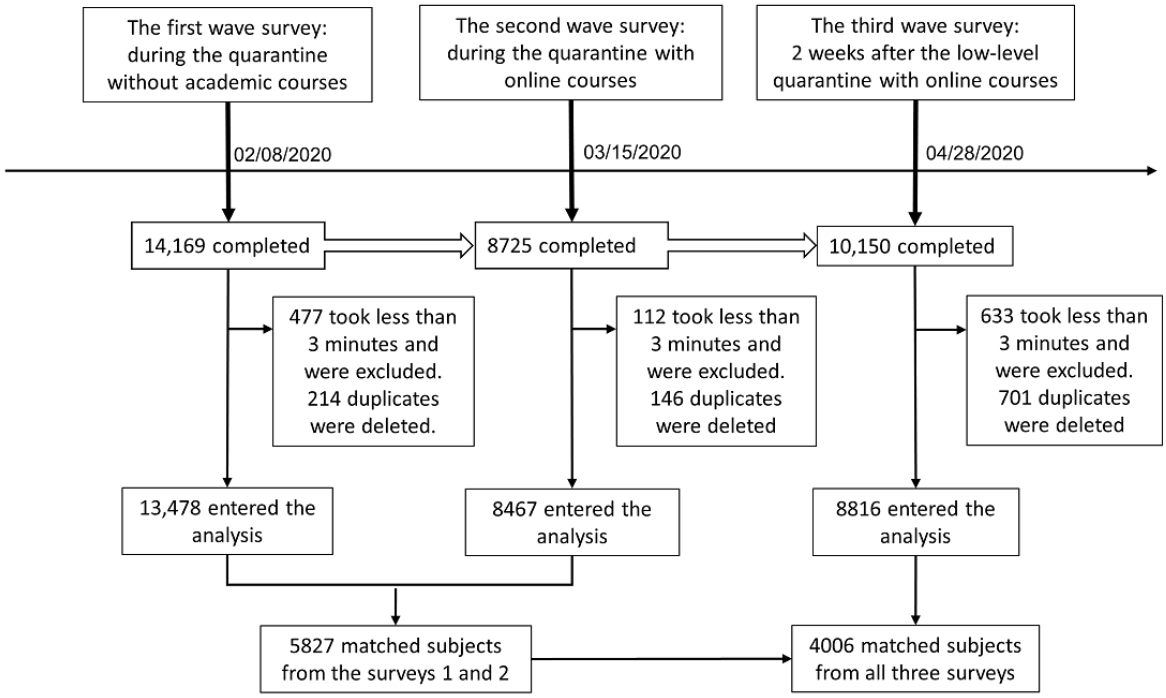
In this prospective cohort study, we surveyed college students at XXMU, including medical and nonmedical students. All participants voluntarily completed the survey via smartphone or computer at three timepoints.

The initial survey (survey 1) was distributed on February 8, 2020, when the participants were on winter break under a high level quarantine due to increases in COVID-19 cases reported. Survey 2 was distributed on March 15, 2020, when the participants had been taking online courses for 2 weeks under a moderate level of quarantine. Survey 3 was distributed on April 30, 2020, when the participants were still taking academic courses at home under a low quarantine level with no new cases reported.

For survey 1, we received 14,691 completed questionnaires. Among them, the 477 students who completed the survey in less than 3 minutes were excluded, and 214 duplications were deleted. For survey 2, we received 8725 completed questionnaires. Among them, the 112 students who completed the survey in less than 3 minutes were excluded, and 146 duplications were deleted. For survey 3, we received 10,150 completed questionnaires. Among them, the 633 students who completed the survey in less than 3 minutes were excluded, and 701 duplications were deleted. In order to track the identity of individuals who completed all three surveys, we used nicknames, age, gender, address (city), grade, and major as ID

variables; there were 5827 ID-matched participants from surveys 1 and 2 and 4006 ID-matched participants from all three surveys, after excluding one ID duplicate in each survey and those who completed the surveys in 3 minutes (Figure 1). In the surveys, quarantine level in different areas where participants lived were announced by the government and were designated as low (score of 1), medium (score of 2), and high (score of 3), which reflected the severity of the COVID-19 pandemic in that area.

Figure 1. Flow diagram of surveys among university students.



The survey protocol was approved by the Committee on Human Research at XXMU. Since the study involved internet technologies, this ensured compliance with the principles of voluntary participation. All participants provided electronic informed consent.

Assessment

We used a battery of questionnaires in our surveys, including questions on basic information (age, gender, only child vs sibling[s], enrollment year, major), knowledge on COVID-19 (0 for “do not know” to 3 for “very knowledgeable”), level of fear (0 for “no fear” to 9 for “extreme fear”), as well as the Y-BOCS and the SAS. The Y-BOCS is an undisputed gold standard to evaluate the severity of OCD symptoms [19,20]. It is the most widely used semistructural scale in both clinical and research settings. It consists of a comprehensive symptom checklist to identify the specific type and content of obsessive and compulsive symptoms in addition to a 10-item rating scale. After inquiring about what types of obsessions and compulsions the patient experiences using a standard checklist, individuals are asked to identify their main symptoms (obsessions and compulsions) and respond to a series of questions. The scale is divided into two subscales that separately measure obsessions and compulsions. For each subscale, five aspects of obsessive and compulsive pathology are each rated on a scale ranging from 0 (no symptoms) to 4 (extreme symptoms): time spent, degree of interference, distress, resistance (greater resistance is assigned lower scores), and perceived control over the symptom. Subscale scores are summed to yield a Y-BOCS total score. There is a moderate correlation in consistency and discrepancy between self-reported and clinician-rated Y-BOCS scores; the highest correlation is observed for the compulsion subscale and patients tend to rate symptoms lower than clinicians [21,22]. Since many studies use a Y-BOCS score ≥16 as an inclusion criteria for OCD, we defined that score as “possible OCD” in this study. We used the SAS to quantify a participant's level of anxiety [23]. The SAS is a 20-item self-report assessment device built to measure anxiety levels, based on scoring in 4 groups of manifestations: cognitive, autonomic, motor, and central nervous system symptoms. Responding to each item, a person should indicate how much each statement applies to him or her in the past 1 or 2 weeks. Each question is scored on a scale of 1 to 4 (1: “a little of the time,” 2: “some of the time,” 3: “a good part of the time,” 4: “most of the time”). Of the 20 items, five items are negatively worded to avoid the problem of set response (ie, careless responding). The total raw scores range from 20 to 80. The raw score is used as an anxiety-severity score. The total score of 20-44 indicates normal range, 45-59 indicates mild-to-moderate anxiety, 60-74 indicates marked-to-severe anxiety, and >75 indicates extreme anxiety.

Statistical Analyses

The microdata from the three surveys of the same population were analyzed using SPSS, version 24 (IBM Corp). Demographic characteristics of participants were tabulated using means and standard deviations for continuous variables and frequency distributions for categorical variables. The repeated measure analyses (Wilks lambda) were performed for the 5827

participants who were matched using ID variables from the first two surveys and for the 4006 matched participants from all three surveys, to examine the changes in Y-BOCS score, anxiety levels (SAS score), and intensity of fear of COVID-19. For all eligible participants from all surveys, analysis of variance (ANOVA), chi-square tests, and regression analyses were performed to examine demographic characteristics and Y-BOCS score predictors.

Results

There were 13,478 participants in survey 1, 8467 participants in survey 2, and 8816 participants in survey 3 who were included in the analysis. The participants were aged between 17-50 years (mean 21.3, SD 2.5 years for survey 1; mean 21.2, SD 2.3 years for survey 2; mean 20.9, SD 2.0 years for survey 3; this mean age was lower than those of surveys 1 and 2, $P<.001$). In total, 664 (4.9%) participants in survey 1, 274 (3.2%) participants in survey 2, and 199 (2.6%) participants in survey 3 were aged ≥26 years old. The proportion of the participants who majored in clinical medicine was higher in survey 1, compared to those in surveys 2 and 3 ($P<.001$). There were 5827 participants who had at least five out of six ID variables matched in surveys 1 and 2, and 4006 participants were matched across all three surveys. The gender composition ratio and the rate of having one or more sibling(s) were not significantly different across survey participants. The distribution of participants in terms of enrollment year (2015-2019) was different between the three surveys ($\chi^2_4=151.6$, $P<.001$) (Table 1).

Table 1. Demographic characteristics and questionnaire score.

Characteristic	Survey 1	Survey 2	Survey 3	F or χ^2	*P* value
Age (year), mean (SD)	21.3 (2.5)	21.2 (2.3)	20.9 (2.0)	F=85.6	<.001
Intensity of fear, mean (SD)	7.8 (2.0)	6.7 (2.2)	6.5 (2.2)	F=1147.9	<.001
Y-BOCS[a] score, mean (SD)	7.9 (5.7)	4.8 (5.1)	4.5 (5.1)	F=1366.6	<.001
SAS[b] standard score, mean (SD)	36.9 (7.9)	36.1 (8.2)	36.2 (8.1)	F=34.2	<.001
Quarantine level, mean (SD)	2.44 (0.7)	1.26 (0.5)	1.00 (0.0)	F=27,129.4	<.001
Age (years), n (%)				χ^2=113.6	<.001
<26	12,814 (95.1)	8193 (96.7)	8617 (97.4)		
≥26	664 (4.9)	274 (3.3)	199 (2.6)		
Gender, n (%)				χ^2=1.8	.41
Male	4662 (34.6)	2991 (35.3)	3113 (35.3)		
Female	8816 (65.4)	5476 (64.7)	5703 (54.7)		
Major, n (%)				χ^2=227.3	<.001
Clinical	8549 (63.4)	4576 (54.0)	5259 (59.7)		
Basic medical	3428 (26.4)	2902 (34.3)	2467 (28.0)		
Nonmedical	1501 (11.1)	989 (11.7)	1090 (12.3)		
Sibling(s), n (%)				χ^2=1.5	.48
No	2452 (18.2)	1495 (17.7)	1557 (17.7)		
Yes	11,026 (81.8)	6972 (82.3)	7259 (82.3)		
Year of enrollment, n (%)				χ^2=360.9	<.001
2015	1319 (10.0)	941 (11.1)	665 (7.5)		
2016	2343 (17.7)	1863 (22.0)	1403 (15.9)		
2017	2997 (22.6)	1511 (17.8)	2013 (22.8)		
2018	3274 (24.7)	2169 (25.6)	2474 (28.1)		
2019	3017 (22.8)	1883 (22.2)	2206 (25.0)		
Other	291 (2.2)	92 (1.3)	55 (0.6)		
Y-BOCS score[c], n (%)				χ^2=704.5	<.001
≥16	1519 (11.3)	305 (3.6)	305 (3.5)		
<16	11,959 (88.7)	8162 (96.4)	8511 (96.5)		

[a]Y-BOCS: Yale-Brown Obsessive-Compulsive Scale.

[b]SAS: Zung Self-Rating Anxiety Scale.

[c]The odds ratio for the Y-BOCS was 2.4 (95% CI 2.2-2.7).

In survey 1, 11.3% (n=1519) of participants had a Y-BOCS score ≥16 (possible OCD); this was significantly higher than the 3.6% (n=305) observed in survey 2 (χ^2_1=401.2, odds ratio [OR] 2.4, 95% CI 2.2-2.7, *P*<.001) and 3.5% (n=305) in survey 3 (χ^2_1=431.9, OR 3.5, 95% CI 3.1-4.0, *P*<.001). Compared to the baseline, the self-reported intensity of fear of COVID-19, Y-BOCS score, SAS standard score, and quarantine level (1=low, 2=medium, and 3=high) were significantly reduced among surveys 2 and 3 participants (*P*<.001 for all) (Table 1).

In the 5827 matched participants from surveys 1 and 2 and the 4006 matched participants from all three surveys, the repeated measure analysis (Wilks lambda) showed that the Y-BOCS score, the SAS standard score, the intensity of fear of COVID-19, and the quarantine level of surveys 2 and 3 decreased significantly from baseline (*P*<.001 for all). The Y-BOCS score, intensity of fear of COVID-19, and quarantine level were lower in survey 3 than in survey 2, while the SAS standard score in survey 3 was higher than that in survey 2 (*P*<.001 for all) (Table 2).

Table 2. Repeated measure analysis (Wilks lambda) in matched samples between surveys 1 and 2 (n=5827) and across all three surveys (n=4006).

Variable	Survey 1	Survey 2	Survey 3	F (*df*)	*P* value
Surveys 1 and 2					
Y-BOCS[a] score, mean (SD)	8.0 (5.6)	4.7 (4.9)	—[b]	1858.6 (1)	<.001
SAS[c] score, mean (SD)	36.6 (7.6)	35.7 (7.9)	—	81.2 (1)	<.001
Intensity of fear, mean (SD)	7.8 (2.0)	6.6 (2.3)	—	1357.9 (1)	<.001
Quarantine level, mean (SD)	2.5 (0.7)	1.3 (0.5)	—	21,371.4 (1)	<.001
All three surveys					
Y-BOCS score, mean (SD)	7.9 (5.5)	4.7 (4.9)	4.3 (4.9)	823.8 (2)	<.001
SAS score, mean (SD)	36.3 (7.4)	35.2 (7.6)	35.6 (7.9)	41.2 (2)	<.001
Intensity of fear, mean (SD)	7.7 (2.0)	6.6 (2.2)	6.4 (2.2)	707.2 (2)	<.001
Quarantine level, mean (SD)	2.5 (0.7)	1.3 (0.4)	1.0 (0.03)	9627.4 (2)	<.001

[a]Y-BOCS: Yale-Brown Obsessive-Compulsive Scale.

[b]Not applicable.

[c]SAS: Zung Self-Rating Anxiety Scale.

To further analyze the characteristic of participants, the two-way ANOVA analysis using two independent variables of the surveys and Y-BOCS score (dichotomously grouped into "possible OCD" with a score ≥16 and <16) was performed. Significant differences of Y-BOCS score, intensity of fear, SAS standard score, and quarantine level were found among the groups (*P*<.001), in which no statistical difference in Y-BOCS score was found between the participants with possible OCD across all three surveys. In surveys 2 and 3, no difference was found in quarantine level between the participants with possible OCD and those with a Y-BOCS score <16 (Table 3). In addition, the chi-square test was applied to test the distribution of possible OCD in the survey participants. The prevalence of possible OCD in males was higher than that in female across all surveys. Taking age into account, the rates of possible OCD in males aged <26 years were higher than those in females (*P*=.001, *P*=.002, and *P*<.001 for surveys 1, 2, and 3, respectively), while the rates of possible OCD were not significantly different between males and females aged ≥26 years across all three surveys. The distribution of possible OCD was significantly different in terms of intensity of fear (*P*≤.001). The rate of possible OCD in participants who had sibling(s) was higher than that in those who had no sibling(s) (χ^2_1=11.2, *P*=.001) in survey 1, but no difference was found in surveys 2 and 3. The distribution of possible OCD in terms of the year of enrollment was different in surveys 2 and 3 (*P*=.03 and *P*=.02; Table 3).

Table 3. Comparison of participants with higher Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) scores (≥16) to those with lower Y-BOCS scores (<16) across all three surveys.

Variable	Survey 1		Survey 2		Survey 3	
	Y-BOCS score <16	Y-BOCS score ≥16	Y-BOCS score <16	Y-BOCS score ≥16	Y-BOCS score <16	Y-BOCS score ≥16
Age (years), mean (SD)	21.3 (2.5)	21.4 (2.4)	21.2 (2.3)	21.2 (2.1	20.9 (2.0)	21.0 (1.9)
Y-BOCS score, mean (SD)	6.5 (4.3)	19.0 (3.2)[a]	4.3 (4.3)	19.0) (3.73[a]	4.0 (4.2)	19.3 (2.9)[a]
Intensity of fear, mean (SD)	7.7 (2.0)	8.7 (1.7)[a]	6.6 (2.3)	7.3 (2.2)[a]	6.5 (2.2)	7.3 (2.0)[a]
SAS standard score, mean (SD)	36.0 (7.1)	44.6 (9.4)[a]	35.6 (7.5)	50.1 (12.6)[a]	35.7 (7.5)	50.5 (10.6)[a]
Quarantine level, mean (SD)	2.4 (0.7)	2.5 (0.6)[b]	1.3 (0.5)	1.3 (0.4)	1.0 (0.0)	1.0 (0.0)
Gender, n (%)						
Male	4081 (87.5)	581 (12.5)[b]	2860 (95.6)	131 (4.4)[b]	2974 (95.5)	139 (4.5)[a]
Female	7878 (89.4)	938 (10.6)	5302 (97.3)	174 (2.7)	5537 (97.1)	166 (2.9)
Age, n (%)						
<26						
Male	3865 (87.4)	556 (12.6)[b]	2772 (95.6)	129 (4.4)[b]	2911 (95.5)	137 (4.5)[a]
Female	7499 (89.3)	894 (10.7)	5127 (96.9)	165 (3.1)	5407 (97.1)	162 (2.9)
≥26						
Male	216 (89.6)	25 (10.4)	91 (97.8)	2 (2.2)	63 (96.9)	2 (3.1)
Female	379 (89.6)	44 (10.4)	172 (95.0)	9 (5.0)	130 (97.0)	4 (3.0)
Major, n (%)						
Clinical	7652 (89.5)	897 (10.5)[b]	4425 (96.7)	151 (3.3)[b]	5098 (96.9)	161 (3.1)[c]
Basic medical	3002 (87.6)	426 (12.4)	2801 (96.5)	101 (3.5)	2367 (95.9)	100 (4.1)
Nonmedical	1305 (86.9)	196 (13.1)	936 (94.6)	53 (5.4)	1046 (96.0)	44 (4.0)
Have sibling(s), n (%)						
No	2223 (95.6)	229 (4.4)[b]	1449 (96.9)	46 (3.1)	1501 (96.4)	56 (3.6)
Yes	9736 (97.3)	1290 (2.7)	6713 (96.3)	259 (3.7)	7010 (96.6)	249 (3.4)
Year of enrollment, n (%)						
2015	1170 (88.7)	149 (11.3)	914 (97.1)	27 (2.9)[c]	645 (97.0)	20(3.0)[c]
2016	2076 (88.6)	267 (11.4)	1788 (96.0)	75 (4.0)	1348 (96.1)	55 (3.9)
2017	2633 (87.9)	364 (12.1)	1452 (96.1)	59 (3.9)	1923 (95.5)	90 (4.5)
2018	2914 (89.0)	360 (11.0)	2103 (97.0)	66 (3.0)	2407 (97.3)	67 (2.7)
2019	2708 (89.8)	309 (10.2)	1815 (96.4)	68 (3.6)	2135 (96.8)	71 (3.2)
Intensity of fear, n (%)						
0	96 (97.0)	3 (3.0)[a]	180 (98.4)	3 (1.6)[b]	185 (99.5)	1 (0.5)[a]
1	100 (98.0)	2 (2.0)	223 (98.2)	4 (1.7)	218 (99.5)	1 (0.5)
2	216 (97.3)	6 (2.7)	382 (97.7)	9 (2.3)	471 (98.5)	7 (1.5)
3	293 (97.7)	7 (2.3)	446 (97.4)	12 (2.6)	544 (97.5)	14 (2.5)
4	980 (95.1)	51 (4.9)	1189 (97.1)	35 (2.9)	1278 (97.6)	31 (2.4)
5	1921 (93.5)	134 (6.5)	1607 (96.6)	55 (3.3)	1758 (96.8)	59 (3.2)
6	1781 (92.0)	154 (8.0)	1179 (97.0)	37 (3.0)	1295 (96.2)	51 (3.8)
7	2088 (90.0)	231 (10.0)	1116 (95.0)	59 (5.0)	1138 (96.3)	44 (3.7)
8	1086 (87.7)	153 (12.3)	513 (96.7)	18 (3.4)	462 (93.7)	31 (6.3)

Variable	Survey 1		Survey 2		Survey 3	
	Y-BOCS score <16	Y-BOCS score ≥16	Y-BOCS score <16	Y-BOCS score ≥16	Y-BOCS score <16	Y-BOCS score ≥16
9	3398 (81.4)	778 (18.6)	1327 (94.8)	73 (5.2)	1162 (94.6)	66 (5.4)
Quarantine level, n (%)						
Low	1102 (90.6)	114 (9.4)[b]	6073 (96.4)	228 (3.6)	8501 (96.5)	305 (3.5)
Medium	4885 (89.6)	564 (10.4)	2032 (96.4)	76 (3.6)	10 (100.0)	0 (0)
High	6272 (88.2)	841 (11.8)	57 (98.3)	1 (1.7)	0 (0)	0 (0)

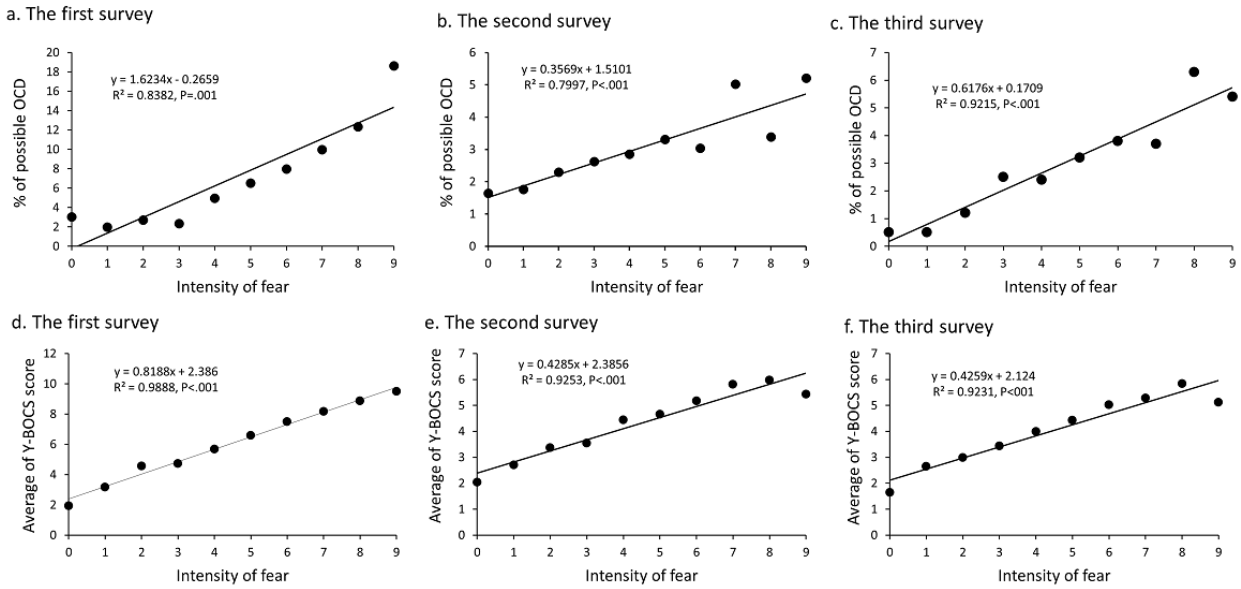
[a]*P*<.001.

[b]*P*<.01.

[c]*P*<.05.

The regression linear analysis indicated that intensity of fear was significantly correlated to the proportions of possible OCD and the average total scores for the Y-BOCS. The correlation coefficient between intensity of fear and rate of participants with possible OCD was 0.92 for survey 1, 0.89 for survey 2, and 0.96 for survey 3 (*P*<.001 for all). The correlation coefficient between intensity of fear and average Y-BOCS score was 0.99 for survey 1, 0.96 for survey 2, and 0.96 for survey 3 (*P*<.001 for all) (Figure 2).

Figure 2. Correlations of the intensity of fear with rates of possible obsessive-compulsive disorder (OCD) and Yale-Brown Obsessive-Compulsive Scale scores.



In the 5827 matched participants from surveys 1 and 2 and the 4006 matched participants from all three surveys, regression analyses indicated that the changes in the intensity of fear (Δfear = fear score from survey 1 – fear score from survey 2 or survey 3) were negatively correlated to the average Y-BOCS score (*P*<.001) (Figure 3).

In order to test the factors that potentially served as predictors for the Y-BOCS score, multiple linear stepwise regressions were conducted, and five variables (intensity of fear, SAS standard score, gender, having sibling[s], and major [1: clinical medicine, 2: basic medical science, and 3: nonmedical major]) entered the equation for all three surveys. Knowledge on COVID-19 entered the equations for surveys 1 and 2 and was negatively correlated to the Y-BOCS score. The quarantine level entered the equation for survey 1 only. Educational level and year of enrollment were excluded from all equations (Table 4). The R^2 of the regression equation was 0.23 for survey 1, 0.23 for survey 2, and 0.26 for survey 3.

Figure 3. Correlation between changes in the intensity of fear and Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) scores in matched participants.

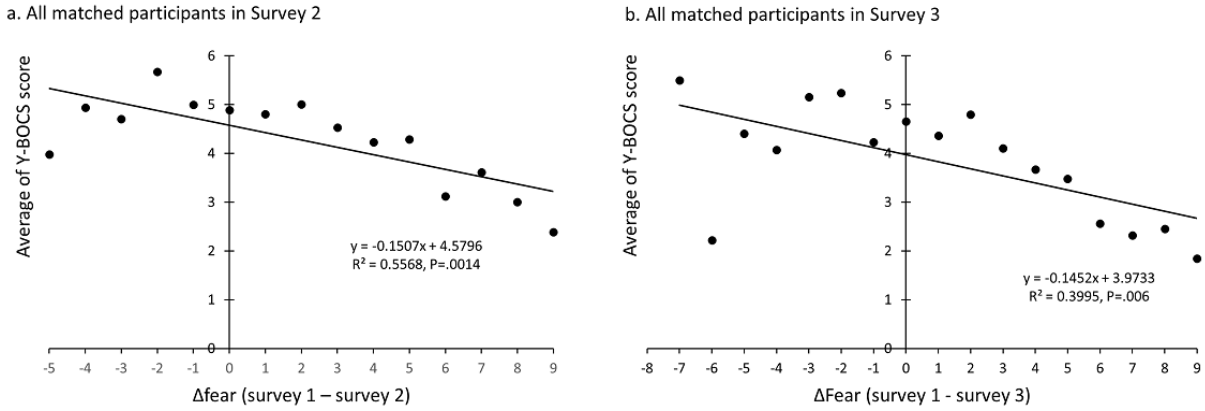


Table 4. Multiple linear regression analyses using the Yale-Brown Obsessive-Compulsive Scale score as the dependent variable.

Survey and variables	B	SE	Beta		*t*	*P* value
			Value	95% CI		
Survey 1						
(Constant)	–9.15	0.48	—[a]	–10.09 to –8.22	–19.22	<.001
SAS[b] standard score	0.29	0.01	0.40	0.28 to 0.30	52.08	<.001
Intensity of fear	0.54	0.02	0.19	0.50 to 0.58	24.67	<.001
Gender	0.50	0.09	0.04	0.32 to 0.68	5.44	<.001
Having sibling(s)	0.49	0.11	0.03	0.27 to 0.71	4.32	<.001
Major	0.27	0.06	0.03	0.15 to 0.40	4.29	<.001
Quarantine level	0.18	0.07	0.02	0.05 to 0.31	2.71	.008
Knowledge on COVID-19	–0.17	0.09	–0.01	–0.33 to 0.00	–1.96	.03
Survey 2						
(Constant)	–8.37	0.48	—	–9.32 to –7.42	–17.30	<.001
SAS standard score	0.28	0.01	0.45	0.27 to 0.29	46.35	<.001
Intensity of fear	0.18	0.02	0.08	0.13 to 0.22	8.10	<.001
Gender	0.67	0.10	0.06	0.47 to 0.87	6.51	<.001
Having sibling(s)	0.64	0.13	0.05	0.39 to 0.89	4.97	<.001
Major	0.21	0.07	0.03	0.07 to 0.35	2.97	.006
Knowledge on COVID-19	–0.22	0.09	–0.02	–0.40 to –0.05	–2.52	.01
Survey 3						
(Constant)	–9.82	0.36	—	–10.54 to –9.11	–27.02	<.001
SAS standard score	0.30	0.01	0.48	0.29 to 0.32	51.44	<.001
Intensity of fear	0.15	0.02	0.07	0.11 to 0.19	7.15	<.001
Gender	0.38	0.10	0.04	0.19 to 0.58	3.82	<.001
Having sibling(s)	0.73	0.12	0.05	0.49 to 0.98	5.90	<.001
Major	0.21	0.07	0.03	0.08 to 0.34	3.19	.006

[a]Not applicable.

[b]SAS: Zung Self-Rating Anxiety Scale.

Discussion

Principal Findings

This online prospective cohort study found that the prevalence of possible OCD (11.3%) in survey 1 at the early stage of the COVID-19 pandemic was significantly higher than that in survey 2 (middle stage, 3.6%), and survey 3 (late stage, 3.5%). The Y-BOCS score, anxiety level, quarantine level, and intensity of fear of COVID-19 were significantly lower at surveys 2 and 3 than at survey 1. Compared to those with a lower Y-BOCS score (<16), participants with possible OCD reported a greater intensity of fear and had a higher SAS standard score (P<.001). Intensity of fear was positively correlated to the rate of possible OCD and the average total scores for the Y-BOCS in each survey (P<.001 for all). Multiple regressions indicated that those with a higher intensity of fear, a higher anxiety level, of male gender, with sibling(s), and majoring in nonmedicine disciplines had a greater chance of having a higher Y-BOCS score across all surveys. In matched survey participants, the Y-BOCS score was negatively correlated to changes in the intensity of fear of COVID-19.

The prevalence of possible OCD in survey 1 was three folds higher than that in surveys 2 and 3, suggesting that possible OCD was induced in the early stage of the COVID-19 pandemic. In addition, the intensity of fear of COVID-19, anxiety level, and quarantine level were significantly higher in survey 1 compared to surveys 2 and 3. Changes in the level of fear were negatively correlated with Y-BOCS score in the follow-up surveys. In each survey, the fear score was strongly correlated to the Y-BOCS score and the rate of possible OCD. The multiple regression analysis showed that both the SAS standard score and intensity of fear significantly contributed to variations in Y-BOCS score. These findings suggest that the intensity of fear of COVID-19 played a role in OCD and that the interactions between fear, anxiety, and pandemic-induced quarantine may be risk factors for an increase in Y-BOCS score. Also, as expected, a high prevalence of possible OCD (11.3%) was observed in survey 1. The COVID-19 pandemic might invoke fear and individuals would manifest OCD-like symptoms when they overacted to this fear. This is an example of the effects of the environment (COVID-19 pandemic) × psychology (fear and/or anxiety) interaction on OCD. Regarding anxiety, a recent longitudinal study on the mental health of the general population during the COVID-19 pandemic that conducted two surveys (1 and 2) at similar intervals to our study did not find a significant reduction in anxiety score in survey 2 [24]. These inconsistent findings may be due to several factors. First, survey 2 in our study was conducted 2 weeks later than that of the longitudinal study. Second, quarantine restrictions had been relaxed in most parts of the country when survey 2 was conducted in this study. Third, we surveyed students in university while the anxiety study surveyed the general population. Fourth, the SAS, which was used in our study, is more sensitive than the Depression, Anxiety and Stress Scale (42 items) used in the other study [25]. Fifth, the participants in this study were taking online courses in surveys 2 and 3, which might have served as a distraction from the pandemic.

In this study, the prevalence of possible OCD decreased from 11.3% at baseline to 3.6% in 5 weeks and remained at 3.5% after 11 weeks. In all three surveys, the participants were at home; hence, the living environment did not vary significantly. At the time of survey 2, participants had been taking online courses at home for 2 weeks, and the pandemic was partially under control. At the time of survey 3, the COVID-19 pandemic was under control; the quarantine level was lowered further, and the participants continued taking online courses at home. The intervals between the three surveys were 5-6 weeks. Therefore, compared to survey 1, the changed environmental factors at surveys 2 and 3 mainly included the status of the pandemic, level of quarantine, and online courses, while the changed psychological factors at surveys 2 and 3 included the intensity of fear of COVID-19, the decreased anxiety level, and the provision of more knowledge on COVID-19. Statistical analyses indicated that the quarantine level and knowledge on COVID-19 were not correlated to Y-BOCS score in survey 3, while knowledge on COVID-19 was negatively correlated to Y-BOCS score and explained less than 1% of the variation seen in the scores. The decreases in intensity of fear and anxiety level may be related to reductions in quarantine level due to declines in new case reports and dissemination of more knowledge on COVID-19 and may also be related to the interaction between those factors and time. In addition, taking online courses could be considered as an intervention to reduce fear of COVID-19 and anxiety, since more time spent on coursework is less time spent on activities that may instigate fear of COVID-19 and COVID-19–related anxiety. Reduction in the intensity of fear and anxiety was correlated to Y-BOCS score, leading to lower rates of possible OCD. Taking online courses is similar to strategies used in OCD treatment (eg, cognitive-coping therapy) [10,11]. That, subsequently, might be related to the lower rate of possible OCD in a relatively short period of time.

Not all participants with fear of COVID-19 were categorized as possible OCD based on Y-BOCS score, although our findings indicate that a higher intensity of fear was related to a higher prevalence of possible OCD. Previously, we investigated the relationship between the fear of negative events and OCD on patients, and found that for most patients with OCD a fear of negative events contributed to their symptoms [10-12].

The findings in this study introduced a new perspective to understanding the relationship between fear of negative events and OCD in the general population. First, it is not unusual that most participants in this study reported a certain intensity of fear related to the COVID-19 pandemic. Second, the attitude, evaluation, and cognition of this fear may affect their response to it. Third, when fear is excessive and disproportionate to the situation, it could lead to the development of an anxiety disorder [26,27]. Those who took the fear seriously and overreacted to it were more likely to be categorized as possible OCD. Fourth, the environment × psychology interaction could be a risk factor for some people and a resilience factor for others due to value-system differences.

We noted that 3 out of 1519 cases of possible OCDs in survey 1, 3 out of 305 cases of possible OCDs in survey 2, and 3 out of 305 cases of possible OCDs in survey 3 reported that their intensity of fear was zero. However, they reported fear of bodily

waste/secretions, dirt or germs, infectious illnesses, and environmental contamination. In the matched samples, no participant who reported zero for intensity of fear had a Y-BOCS score ≥16.

In addition, the participants who reported having sibling(s) were more likely to be categorized as possible OCD than those who had no sibling(s). In China, an only child easily becomes the family's center of attention and gets care from not only parents but also grandparents, even throughout early adulthood. Only children have closer parent-child relationships, which is probably related to being dependent upon others and having relatively few familial responsibilities [28]. On the other hand, those with siblings generally take on more responsibilities, such as caring for siblings or assisting in family affairs. Participants with sibling(s) may be more responsible and thus overreact to COVID-19 during the pandemic and be involved in transient possible OCD.

Our findings demonstrated that the prevalence of possible OCD in surveys 2 and 3 was 3.6% and 3.5%, respectively. Additionally, the prevalence of possible OCD in male participants at all timepoints was higher (12.5%, 4.4%, and 4.5% for surveys 1, 2, and 3, respectively) compared to females (10.6%, 2.7%, and 2.9% for surveys 1, 2, and 3, respectively). The findings suggested that the male students in the present study seem to have a higher prevalence of OCD than the general population [5]. Previously, Torres et al [29] found that 3.8% of Brazilian medical students had a possible case of OCD based on the Obsessive-Compulsive Inventory–Revised. Using the OCD subsection of the Clinical Interview Schedule–Revised as a self-administered questionnaire, Jaisoorya et al [30] reported the point prevalence of OCD in Indian college students as 3.3% (males: 3.5%; females: 3.2%). Yoldascan et al [31] reported that, in Turkish university students, the prevalence of OCD is 4.2%, and OCD was significantly associated with the male gender. These findings, along with ours, suggest that the prevalence of OCD among university students is similar to or higher than the general population across different cultures.

Our survey results indicated that the prevalence of possible OCD was positively correlated to the participant's academic discipline. Basic medical students (ie, those not involved in clinical practice) had a higher prevalence of possible OCD than medical students, but a lower prevalence of possible OCD than nonmedical students. Additionally, students in their third and fourth years had a higher prevalence of possible OCD than first-, second-, and fifth-year students. The findings suggested that the knowledge that students obtained in their academic majors may play a role in the "onset" of possible OCD, possibly by affecting one's cognition and appraisal of fear of COVID-19. Cognitive-behavior therapy might be useful to treat possible OCD and anxiety symptoms [32].

Limitations

Although this study included a pragmatic design and a large sample size, there are several limitations that need to be addressed. First, individuals with possible OCD were defined only according to the Y-BOCS score and were not verified via face-to-face interview, which might be related to the higher prevalence of possible OCD in this study compared to the general population. Second, we did not collect any biological samples and therefore could not analyze the relationship between OCD and factors such as genetics and/or the expression of certain proteins. Third, all participants were university students aged 17 to 50 years. Therefore, caution is advised if using the findings to infer patterns in the general population.

Conclusions

Our findings indicate that an environment × psychology interaction might be involved in the onset of OCD and that a fear of negative events should be considered as a target of interventions for mental health and well-being in both stressful situations and in clinical practice.

Acknowledgments

We would like to thank all participants for taking part in the surveys. We would also like to thank statistician and professor Li-Hui Gui for sharing his knowledge and helping in the analysis of the data.

Financial support was provided by the National Natural Foundation of China (81671365). The funder played no role in the design, conduct of the study; the collection, management, analysis and interpretation of data; the preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

GJ, WW, and K-CY contributed equally to this work. X-ZH contributed to the design of the study. X-ZH, WW, GJ, and K-CY contributed to the questionnaire design and dissemination. GJ, K-CY, HL, L-JS, J-DM, C-YH, S-SZ, ZZ, TL, JC, S-CY, and WW were responsible for dissemination of the questionnaire, data collection, and data management. X-ZH, GJ, and K-CY performed the statistical analyses. GJ, K-CY, and X-ZH drafted the manuscript. All authors contributed to the interpretation of the data and offered critical revisions of the draft. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Lu H, Stratton CW, Tang Y. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol 2020 Apr;92(4):401-402 [FREE Full text] [doi: 10.1002/jmv.25678] [Medline: 31950516]
2. Yang Y, Li W, Zhang Q, Zhang L, Cheung T, Xiang Y. Mental health services for older adults in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e19. [doi: 10.1016/s2215-0366(20)30079-1]
3. Rogers JP, Chesney E, Oliver D, Pollak TA, McGuire P, Fusar-Poli P, et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. The Lancet Psychiatry 2020 Jul;7(7):611-627 [FREE Full text] [doi: 10.1016/S2215-0366(20)30203-0] [Medline: 32437679]
4. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):e [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
5. Holes R, Yudofsky S, Roberts L. Textbook of Psychiatry, 6th ed. Arlington, VA: The American Psychiatric Publishing; 2014.
6. Ustün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL. Global burden of depressive disorders in the year 2000. Br J Psychiatry 2004 May;184:386-392. [doi: 10.1192/bjp.184.5.386] [Medline: 15123501]
7. O'Neill ML, Nenzel ME, Caldwell W. Intrusive thoughts and psychopathy in a student and incarcerated sample. J Behav Ther Exp Psychiatry 2009 Mar;40(1):147-157. [doi: 10.1016/j.jbtep.2008.07.004] [Medline: 18762284]
8. Salkovskis PM. Understanding and treating obsessive-compulsive disorder. Behav Res Ther 1999 Jul;37 Suppl 1:S29-S52. [Medline: 10402695]
9. Clark DA. Focus on "cognition" in cognitive behavior therapy for OCD: is it really necessary? Cogn Behav Ther 2005;34(3):131-139. [doi: 10.1080/16506070510041194] [Medline: 16195052]
10. Hu X, Ma J, Huang P, Shan X, Zhang Z, Zhang J, et al. Highly efficacious cognitive-coping therapy for overt or covert compulsions. Psychiatry Res 2015 Oct 30;229(3):732-738. [doi: 10.1016/j.psychres.2015.08.010] [Medline: 26275705]
11. Hu X, Wen Y, Ma J, Han D, Li Y, Wang S. A promising randomized trial of a new therapy for obsessive-compulsive disorder. Brain Behav 2012 Jul;2(4):443-454 [FREE Full text] [doi: 10.1002/brb3.67] [Medline: 22950048]
12. Ma J, Wang C, Li H, Zhang X, Zhang Y, Hou Y, et al. Cognitive-coping therapy for obsessive-compulsive disorder: a randomized controlled trial. J Psychiatr Res 2013 Nov;47(11):1785-1790. [doi: 10.1016/j.jpsychires.2013.08.002] [Medline: 23988179]
13. Abramowitz JS, Taylor S, McKay D. Obsessive-compulsive disorder. The Lancet 2009 Aug 08;374(9688):491-499. [doi: 10.1016/S0140-6736(09)60240-3] [Medline: 19665647]
14. Hauser TU, Eldar E, Dolan RJ. Neural Mechanisms of Harm-Avoidance Learning: A Model for Obsessive-Compulsive Disorder? JAMA Psychiatry 2016 Nov 01;73(11):1196-1197. [doi: 10.1001/jamapsychiatry.2016.1938] [Medline: 27654061]
15. Coles ME, Heimberg RG, Frost RO, Steketee G. Not just right experiences and obsessive-compulsive features: experimental and self-monitoring perspectives. Behav Res Ther 2005 Feb;43(2):153-167. [doi: 10.1016/j.brat.2004.01.002] [Medline: 15629747]
16. Olatunji BO, Tart CD, Ciesielski BG, McGrath PB, Smits JAJ. Specificity of disgust vulnerability in the distinction and treatment of OCD. J Psychiatr Res 2011 Sep;45(9):1236-1242 [FREE Full text] [doi: 10.1016/j.jpsychires.2011.01.018] [Medline: 21353249]
17. Geller DA, McGuire JF, Orr SP, Pine DS, Britton JC, Small BJ, et al. Fear conditioning and extinction in pediatric obsessive-compulsive disorder. Ann Clin Psychiatry 2017 Feb;29(1):17-26 [FREE Full text] [Medline: 28207912]
18. Tran BX, Ha GH, Nguyen LH, Vu GT, Hoang MT, Le HT, et al. Studies of Novel Coronavirus Disease 19 (COVID-19) Pandemic: A Global Analysis of Literature. Int J Environ Res Public Health 2020 Jun 08;17(11):e [FREE Full text] [doi: 10.3390/ijerph17114095] [Medline: 32521776]
19. Goodman WK, Price LH, Rasmussen SA, Mazure C, Delgado P, Heninger GR, et al. The Yale-Brown Obsessive Compulsive Scale. II. Validity. Arch Gen Psychiatry 1989 Nov;46(11):1012-1016. [doi: 10.1001/archpsyc.1989.01810110054008] [Medline: 2510699]
20. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Arch Gen Psychiatry 1989 Nov;46(11):1006-1011. [doi: 10.1001/archpsyc.1989.01810110048007] [Medline: 2684084]
21. Federici A, Summerfeldt LJ, Harrington JL, McCabe RE, Purdon CL, Rowa K, et al. Consistency between self-report and clinician-administered versions of the Yale-Brown Obsessive-Compulsive Scale. J Anxiety Disord 2010 Oct;24(7):729-733. [doi: 10.1016/j.janxdis.2010.05.005] [Medline: 20561767]
22. Hauschildt M, Dar R, Schröder J, Moritz S. Congruence and discrepancy between self-rated and clinician-rated symptom severity on the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) before and after a low-intensity intervention. Psychiatry Res 2019 Mar;273:595-602. [doi: 10.1016/j.psychres.2019.01.092] [Medline: 30716599]
23. Zung WW. A Rating Instrument For Anxiety Disorders. Psychosomatics 1971 Nov;12(6):371-379. [doi: 10.1016/S0033-3182(71)71479-0] [Medline: 5172928]

24. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
25. Dunstan DA, Scott N, Todd AK. Screening for anxiety and depression: reassessing the utility of the Zung scales. BMC Psychiatry 2017 Sep 08;17(1):329 [FREE Full text] [doi: 10.1186/s12888-017-1489-6] [Medline: 28886698]
26. Diagnostic and statistical manual of mental disorders, 5th ed. Washington, DC: American Psychiatric Association; 2013.
27. Shackman AJ, Fox AS. Contributions of the Central Extended Amygdala to Fear and Anxiety. J Neurosci 2016 Aug 03;36(31):8050-8063 [FREE Full text] [doi: 10.1523/JNEUROSCI.0982-16.2016] [Medline: 27488625]
28. Li S, Chen R, Cao Y, Li J, Zuo D, Yan H. Sexual knowledge, attitudes and practices of female undergraduate students in Wuhan, China: the only-child versus students with siblings. PLoS One 2013;8(9):e73797 [FREE Full text] [doi: 10.1371/journal.pone.0073797] [Medline: 24023905]
29. Torres AR, Cruz BL, Vicentini HC, Lima MCP, Ramos-Cerqueira ATA. Obsessive-Compulsive Symptoms in Medical Students: Prevalence, Severity, and Correlates. Acad Psychiatry 2016 Feb;40(1):46-54. [doi: 10.1007/s40596-015-0357-2] [Medline: 26108391]
30. Jaisoorya TS, Janardhan Reddy YC, Nair BS, Rani A, Menon PG, Revamma M, et al. Prevalence and correlates of obsessive-compulsive disorder and subthreshold obsessive-compulsive disorder among college students in Kerala, India. Indian J Psychiatry 2017;59(1):56-62 [FREE Full text] [doi: 10.4103/0019-5545.204438] [Medline: 28529361]
31. Yoldascan E, Ozenli Y, Kutlu O, Topal K, Bozkurt AI. Prevalence of obsessive-compulsive disorder in Turkish university students and assessment of associated factors. BMC Psychiatry 2009 Jul 06;9:40 [FREE Full text] [doi: 10.1186/1471-244X-9-40] [Medline: 19580658]
32. Ho CS, Chee CY, Ho RC. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. Ann Acad Med Singap 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]

Abbreviations

ANOVA: analysis of variance
OCD: obsessive-compulsive disorder
OR: odds ratio
SAS: Zung Self-Rating Anxiety Scale
XXMU: Xinxiang Medical University
Y-BOCS: Yale-Brown Obsessive-Compulsive Scale

Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by L Zhang, R Ho; comments to author 21.08.20; revised version received 26.08.20; accepted 14.09.20; published 30.09.20.

Please cite as:
Ji G, Wei W, Yue KC, Li H, Shi LJ, Ma JD, He CY, Zhou SS, Zhao Z, Lou T, Cheng J, Yang SC, Hu XZ
Effects of the COVID-19 Pandemic on Obsessive-Compulsive Symptoms Among University Students: Prospective Cohort Survey Study
J Med Internet Res 2020;22(9):e21915
URL: http://www.jmir.org/2020/9/e21915/
doi:10.2196/21915
PMID:32931444

Original Paper

Comparison of a Mobile Health Electronic Visual Analog Scale App With a Traditional Paper Visual Analog Scale for Pain Evaluation: Cross-Sectional Observational Study

Alexandra Turnbull[1*], BSc, Hons; Dean Sculley[2*], BSc, Hons, MSc, PhD; Carles Escalona-Marfil[3,4*], BSc, Hons, MSc, PhD; Lluís Riu-Gispert[3*], BSc; Jorge Ruiz-Moreno[5*], BSc, MA; Xavier Gironès[3*], BSc, Hons, PhD; Andrea Coda[1,6*], BSc, Hons, PhD

[1]School of Health Sciences, Faculty of Health and Medicine, The University of Newcastle, Ourimbah, Australia

[2]School of Biomedical Sciences and Pharmacy, Faculty of Health and Medicine, The University of Newcastle, Ourimbah, Australia

[3]Facultat de Ciències de la Salut de Manresa, Universitat de Vic-Universitat Central de Catalunya, Manresa, Spain

[4]Department of Physical Therapy, Escola Universertària de la Salut i l'Esport (EUSES), University of Girona, Salt, Girona, Spain

[5]MIXESTAT SL, Barcelona, Spain

[6]Priority Research Centre Health Behaviour, Hunter Medical Research Institute, Newcastle, Australia

[*]all authors contributed equally

Corresponding Author:
Andrea Coda, BSc, Hons, PhD
School of Health Sciences
Faculty of Health and Medicine
The University of Newcastle
Health Precinct, BE154, PO Box 127
Ourimbah, 2258
Australia
Phone: 61 0243484507
Email: andrea.coda@newcastle.edu.au

Abstract

Background: Accurate quantification of pain in a clinical setting is vital. The use of an electronic pain scale enables data to be collected, analyzed, and utilized much faster compared with traditional paper-based scales. The advancement of smart technology in pediatric and adult pain evaluation may offer opportunities to introduce easy-to-use and reliable pain assessment methods within different clinical settings. If promptly introduced within different pediatric and adult pain clinic services, validated and easily accessible mobile health pain apps may lead to early pain detection, promoting improvement in patient's quality of life and leading to potentially less time off from school or work.

Objective: This cross-sectional observational study aimed to investigate the interchangeability of an electronic visual analog scale (eVAS) app with a traditional paper visual analog scale (pVAS) among Australian children, adolescents, and adults for pain evaluation.

Methods: Healthy participants (age range 10-75 years) were recruited from a sporting club and a secondary school in Melbourne (Australia). The data collection process involved application of pressure (8.5 kg/cm^2) from a Wagner Force Dial FDK 20 to the midpoint of the thumb. The pressure was applied twice with a 5-minute interval. At each pressure application, participants were asked to randomly record their pain perception using the "eVAS" accessible via the "Interactive Clinics" app and the traditional pVAS. Statistical analysis was conducted to determine intermethod and intramethod reliabilities.

Results: Overall, 109 healthy participants were recruited. Adults (mean age 42.43 years, SD 14.50 years) had excellent reliability, with an intraclass correlation coefficient (ICC) of 0.94 (95% CI 0.91-0.96). Children and adolescents (mean age 13.91 years, SD 2.89 years) had moderate-to-good intermethod and intramethod reliabilities, with an ICC of 0.80 (95% CI 0.70-0.87) and average ICC of 0.80 (95% CI 0.69-0.87), respectively.

Conclusions: The eVAS app appears to be interchangeable compared with the traditional pVAS among children, adolescents, and adults. This pain evaluation method may offer new opportunities to introduce user-friendly and validated pain assessment apps for patients, clinicians, and allied health professionals.

(J Med Internet Res 2020;22(9):e18284) doi:10.2196/18284

KEYWORDS

pain; mobile app; mHealth; digital health; electronic visual analog scale; visual analog scale; symptom; eHealth; reliability

Introduction

Pain is a complex and multifactorial phenomenon that can negatively impact a patient's health-related quality of life [1]. Pain outcome measures are commonly used to assess the severity of symptoms in children, adolescents, and adults [2]. Traditionally, symptom progression has been recorded using the visual analog scale (VAS), Wong Baker scale, numeric rating scale, verbal rating scale, and faces pain scale-revised [3-6]. These tools have been extensively validated as appropriate measures for assessing pain and are commonly used daily by allied health professionals (AHPs).

Evidence suggests there are limitations associated with the more traditional paper pain outcome measures that are still commonly used in various clinical settings [7]. Drawn face scales may result in incorrect recordings if a child experiences difficulty in distinguishing between the feeling of pain and the emotional state, and smiling faces could result in overestimation of pain intensity [4]. These limitations are mostly based on paper pain scales being cumbersome, occasionally complex to use, and at risk of possible practitioner error [8].

The continuous growth of mobile health (mHealth) offers unparalleled opportunities to address issues related to health systems and accessing accurate, reliable, and frequent health data [9,10]. The introduction of smart technology in pediatric and adult pain evaluation may offer opportunities to implement tailored pain assessments within different clinical settings. Recently, novel technologies have emerged that utilize smart devices to improve the existing traditional pain outcome measures [11-16]. Few of these novel technologies have been examined for reliability and validity among children, adolescents, and adults, with most employing small sample sizes. There is growing evidence to suggest that electronic pain outcome measures are interchangeable with existing traditional pain outcome measures, but more mHealth and eHealth research is needed to test the validity and reliability of the electronic VAS (eVAS) among children and adolescents [17]. By evaluating the validity of mHealth and eHealth interventions available to patients and clinicians, we can equip AHPs with a more effective tool to measure symptom progression. This study adheres to the standards of digital health interventions set by the World Health Organization (WHO) in 2016 [18] and the Australian Government Health Authorities' guidelines with regard to how to validate new digital mHealth systems for the benefit of patients, clinicians, and the community [19]. These include monitoring and evaluating linear stages of development from prototype through to pilot studies showing efficacy, demonstration of effectiveness, scaling up, and integration into the clinical environment [18]. The traditional paper VAS (pVAS) was used owing to its well-established validity across all age groups [20]. This cross-sectional observational study aimed to investigate the interchangeability between an eVAS app and a traditional pVAS among children, adolescents, and adults. This may validate the use of the eVAS in a clinical setting, enhancing the collection, analysis, and dissemination of data in line with the eHealth and mHealth pathways envisaged by health authorities worldwide.

Methods

Recruitment

Healthy participants were recruited through a convenient sample of participants in Melbourne (Victoria, Australia) from John Monash Science School (Clayton) and KBH Brumbies Hockey Club (Mont Albert North). English speaking children and adolescents (age range 10-18 years) and adults (age range 18-75 years) were eligible to participate in the study. A mean age of 9.8 years appears to be suitable to evaluate the concept of experienced pain [3]; therefore, this finding was utilized to inform the age group for this mHealth trial, despite the possibility to reliably use the traditional pVAS from the age of 7 years [20]. Participants were excluded if they were diagnosed with neurological disorders, were receiving medication that would alter pain perception or threshold, and had severe visual impairments that may prevent viewing the pVAS and eVAS.

Prior to consent approval, a participant information sheet was supplied to potential participants, and they were made aware of the procedure and the time commitments required to take part in the study. The participant information sheet was adapted according to age (adults, and children and adolescents) including more visual aids in the children's version. Ethical approval was obtained from the University of Newcastle Human Research Ethics Committee (Dev-005638). Approval was also sought from and granted by the Victorian Schools and the Department of Education (Victoria) (2018_00373). Participants' gender and age were recorded and their identities were completely anonymized, with a unique ID assigned to each participant.

Measuring Instruments

In order to apply a standardized pressure to the participant's thumb, a Wagner Force Dial FDK 20 with a 1 cm^2 circular rubber end was adopted for each data collection, and the same data collector (AT) completed all measurements. The setup of the collection included a simple table and chair. The participant sat on the chair with the thumb on the edge of the table and other fingers underneath. A vertical pressure (8.5 kg/cm^2) using the Force Dial was applied to the midpoint of the thumb for 3 seconds (Figure 1).

Figure 1. Wagner Force Dial pressure applied on the participant's thumb.

The chosen amount of pressure (kg/cm^2) applied to the participant's thumb was previously successfully used by Escalona-Marfil et al and provides light enough pressure to mimic symptoms of mild pain while not being too extensive as to generate stronger pain or skin damage [17,21]. After application of pressure, the participant was asked to randomly complete either the pVAS (Figure 2) or eVAS (Figure 3).

Figure 2. Standard paper visual analog scale that was printed on a white background A4-sized paper.

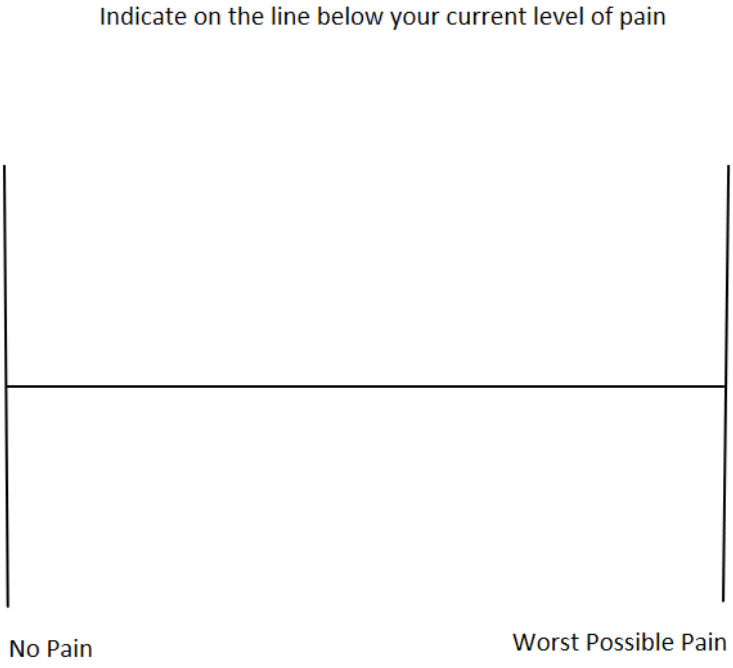
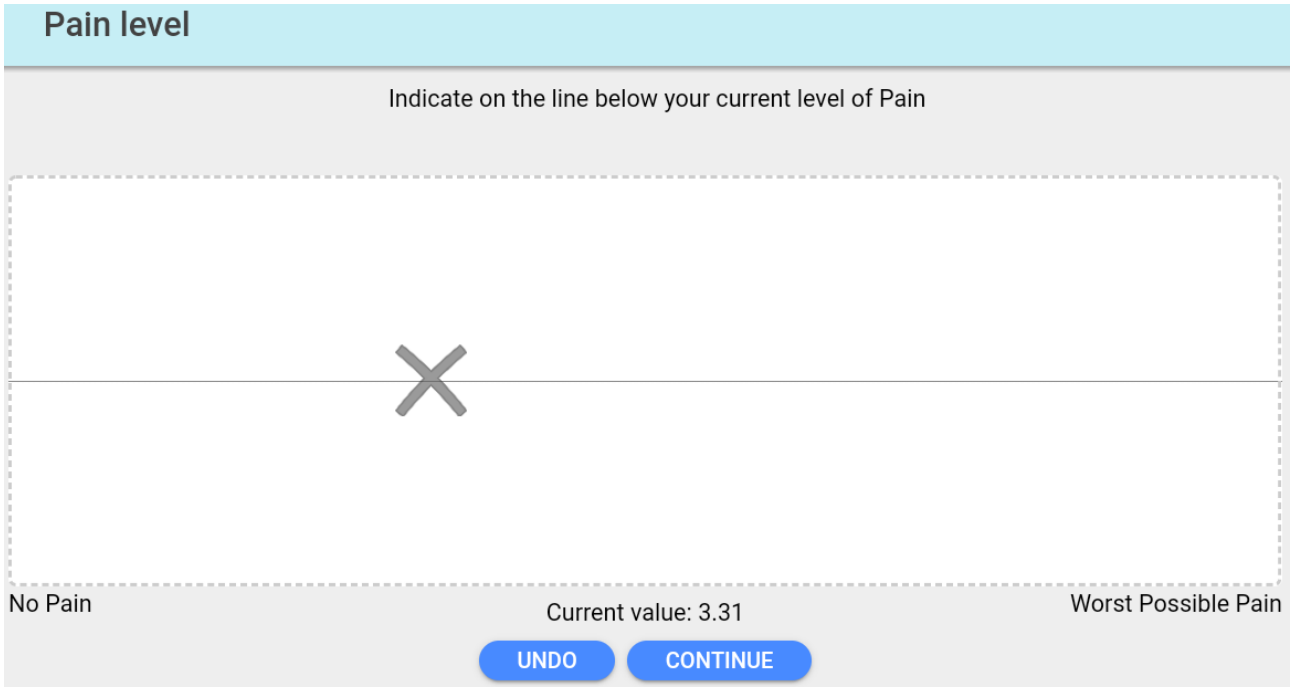


Figure 3. The electronic visual analog scale accessible via the Interactive Clinics app developed by BitGenoma Ltd, Spain.



The order of measurement using the pVAS and eVAS was block randomized (groups of 10). Sequential allocation was achieved using a freely available random number generator [22]. Allocation concealment for the eVAS and pVAS sequence was achieved by using sequentially numbered, opaque, and sealed envelopes. Both sequential generation and allocation concealment were conducted by an independent research team member (AC) who was not involved during the data collection process and did not have any prior or ongoing contact with enrolled participants.

With regard to the pVAS, participants were asked to draw a vertical line that corresponded with their symptom level, with the left side corresponding to “no pain” and the right side corresponding to “the worst pain imaginable” [21]. The traditional pVAS was a 12 × 7.5-cm white paper sheet with a 10-cm horizontal line drawn and two vertical 6-cm lines drawn either side. The eVAS was accessible via the “Interactive Clinics” app, and all recordings were conducted on a 7-inch (17.8 cm) Samsung Galaxy Tab 3 with Android operating system (v5.1.1), displaying a 13.5-cm straight horizontal line on a white background. When using the eVAS, the tablet was placed horizontally on the table at all times and each participant was asked to place one finger on the line on the screen [17]. In order to prevent bias, the data collector (AT) wiped the tablet

screen completely between measurements to prevent the participant from being able to see the position of the fingerprint previously placed on the tablet's screen.

After a period of 5 minutes, the sequence of data collection was reversed for each individual. Data collection only took approximately 10-15 minutes in total for each participant, and no follow-up was required. The use of standardized pressure application with 1-minute intervals was reliable for absolute pressure thresholds in multiple studies [23,24]. Additionally, simple pressure algometry is a repeatable measure of pain threshold [25,26]. A 5-minute interval was introduced to reduce possible risk of reacting pre-emptively to stimuli and to prevent temporal summation that could have impacted the quality of the assessment. For logistical reasons, the data collector (AT) was not blinded to this process. However, as soon as the values on the pVAS and eVAS were recorded, the tablet or the paper was immediately withheld from the participant in order to prevent any possible modifications to the data entered by the participant. The results gathered from the pVAS were extrapolated (by AT) using a standard plastic ruler, whereas the eVAS value was automatically calculated by the software.

Statistical Analysis

For a 5% two-sided *t* test with α=.05 and 80% power in an observational cross-sectional study with one intervention observation and a moderate effect size, it was estimated that a total of 100 subjects would be required [27,28]. The study was overpowered to an estimated 110 subjects (55 aged 10-18 years and 55 aged 18-75 years) to allow for a 9% dropout rate during the data collection period. Summary statistics for eVAS and pVAS results were calculated by splitting the measurement and method. Two approaches have been used to evaluate the agreement of the two methods (intermethod and intramethod reliabilities by means of intraclass correlation coefficients). STATA 15 (StataCorp LLC) was utilized for statistical analysis [29]. The independent statistician was blinded to the eVAS and pVAS allocation concealment and to the participant identity.

Intermethod and Intramethod Agreement Analysis

A mixed factorial model was employed to derive two intraclass correlation coefficients according to Shrout-Fleiss reliability fixed set [29,30]. One coefficient was a measure of intermethod reliability (ρ) estimated by the intraclass correlation coefficient (ICC). This coefficient was defined as the correlation between VAS values from different methods in the same subject and same replication. The other intraclass coefficient (γ) estimated by the average ICC (ICCa) was used as a measure of intramethod reliability. This was defined as the correlation between VAS values in the same method and same subject. A two-way balanced mixed analysis of variance model without interaction, a random subject effect, and a fixed method effect was fitted to estimate ICCs. The mean of squares for subjects, subject-method interactions, and errors from components of variance were also calculated [31]. Statistical inference of the ICCs was performed with CIs. In order to improve reliability coefficients, 95% CIs were calculated from the estimated sum of squares. For both intermethod reliability and intramethod reliability, the ICCs were higher than 0.8. In order to specify the precision of the estimated ICC, the length of the 95% CI was expressed as a function of the ICC value. Given that it was not possible to increase the number of methods to evaluate the VAS, the number of subjects was increased. In children and adolescents, with 94 ratings per method (47 subjects with two replicates per subject) and an anticipated ICC value of at least 0.8, an acceptable length for the 95% CI will be less than or equal to 0.2. In adults, with 124 ratings per method (62 subjects with two replicates per subject) and an anticipated ICC value of at least 0.8, an acceptable length for the 95% CI will be less than or equal to 0.1. Good agreement among methods was evaluated by plotting both methods against subjects and performing a Wilcoxon rank-sum test. According to Portney and Watkins, ICC values are classified as follows: ICC < 0.5, poor; $0.5 \leq$ ICC ≤ 0.75, moderate; $0.75 <$ ICC ≤ 0.9, good; and ICC > 0.9, excellent [30].

Results

Participant Characteristics

A total of 109 participants were included in the study. The study population consisted of 47 children and adolescents (mean age 13.9 years, SD 2.89 years; range 10-18 years; 16 female and 31 male participants) and 62 adults (mean age 42.44 years, SD 14.50 years; range 19-73 years; 37 female and 25 male participants) (Table 1 and Table 2, respectively). No participants were lost to the analysis.

Table 1. Summary statistics (visual analog scale measurements) for children and adolescents.

Variable	First measure		Second measure	
	eVAS[a]	pVAS[b]	eVAS	pVAS
Number	47	47	47	47
Mean value	1.692553	1.657447	1.774681	1.642979
SD value	0.977331	1.039177	1.03974	1.000408
P50[c]	1.45	1.60	1.55	1.50
Minimum value	0.05	0.00	0.00	0.00
Maximum value	3.65	4.00	4.42	4.00

[a]eVAS: electronic visual analog scale.

[b]pVAS: paper visual analog scale.

[c]P50: middle estimate.

Table 2. Summary statistics (visual analog scale measurements) for adults.

Variable	First measure		Second measure	
	eVAS[a]	pVAS[b]	eVAS	pVAS
Number	62	62	62	62
Mean value	1.738387	1.690323	1.819839	1.759677
SD value	1.550611	1.571723	1.748486	1.743415
P50[c]	1.245	1.100	1.265	1.300
Minimum value	0.05	0.00	0.00	0.00
Maximum value	7.91	8.10	8.26	8.50

[a]eVAS: electronic visual analog scale.

[b]pVAS: paper visual analog scale.

[c]P50: middle estimate.

Analysis of the eVAS and pVAS in Children, Adolescents, and Adults

Tables 1 and 2 show summary statistics for VAS measurements by measurement order and instrument (eVAS and pVAS) for the child and adolescent group and adult group, respectively. Differences between methods for median values ranged from 0.05 to 0.15 in children and adolescents and from 0.035 to 0.145 in adults. In Figures 4 and 5, the scatter plots for eVAS compared with pVAS are displayed for the child and adolescent group and the adult group, respectively, showing agreement between the two methods.

It is possible to observe in adults, the dispersion from values of VAS around 3, which is corrected when taking natural logarithms into account (Figure 6).

Figure 4. Scatter plot of the data in children and adolescents. Points are represented by subject number. eVAS: electronic visual analog scale; pVAS: paper visual analog scale.

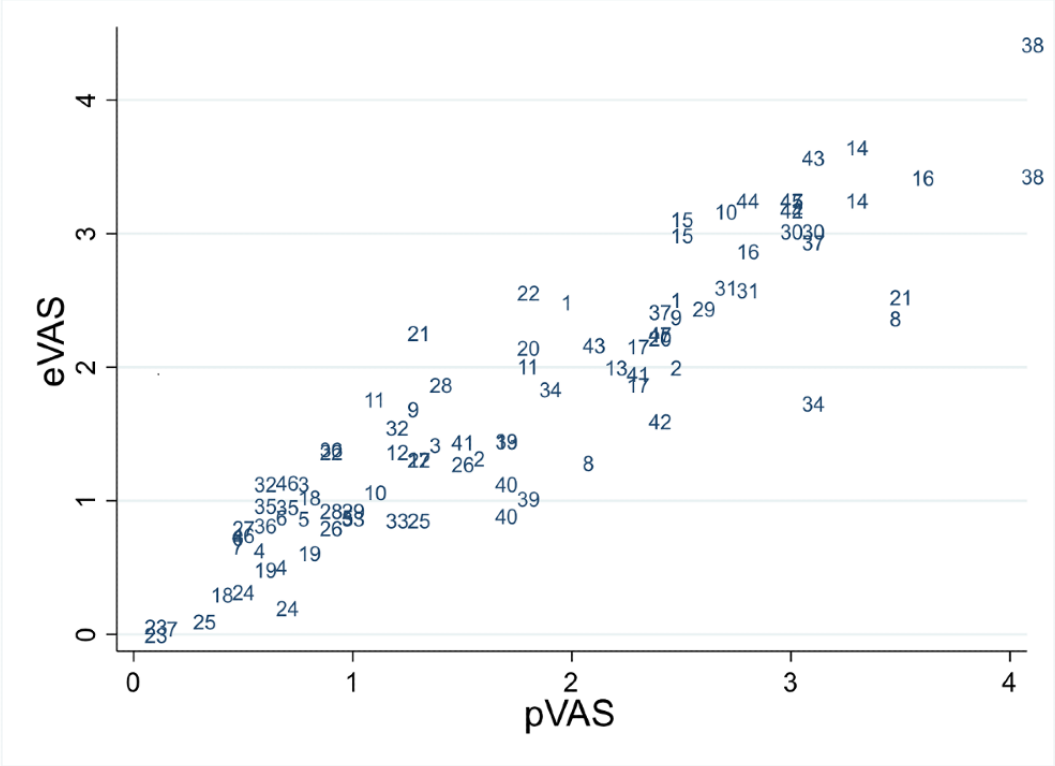


Figure 5. Scatter plot of the data in adults. Points are represented by subject number. eVAS: electronic visual analog scale; pVAS: paper visual analog scale.

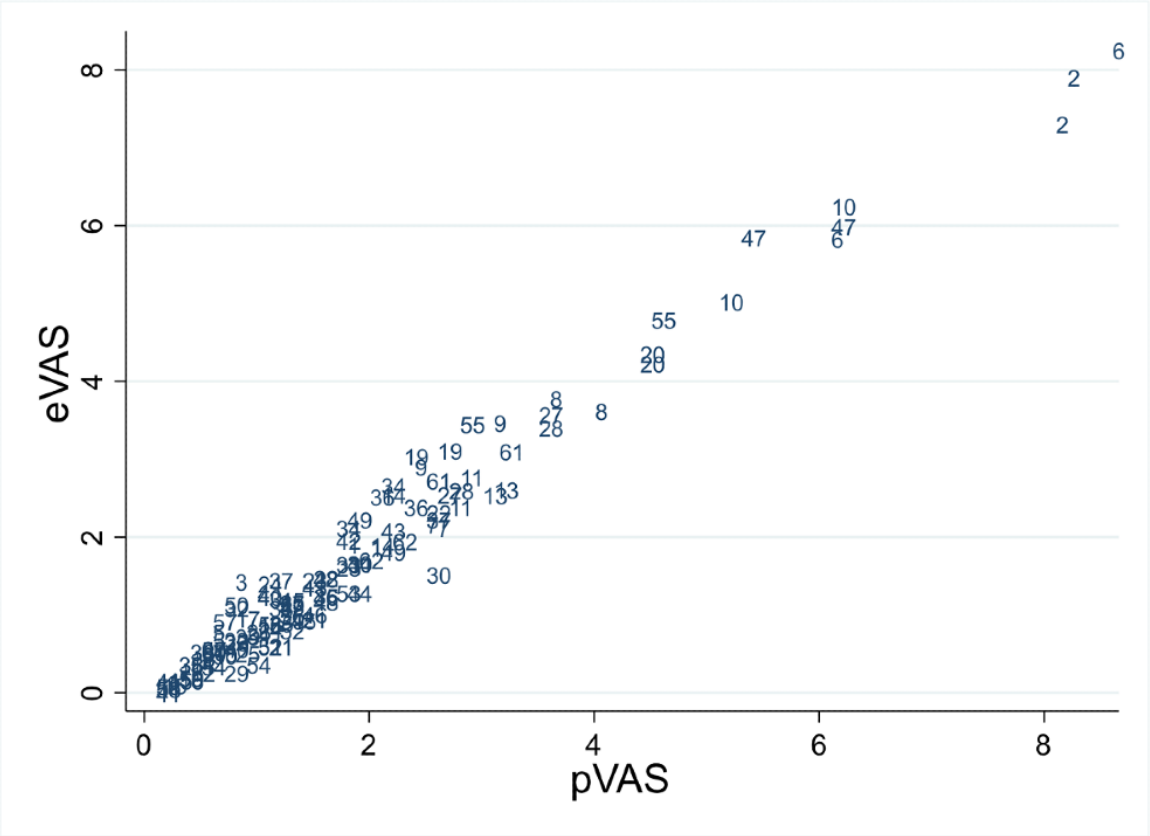
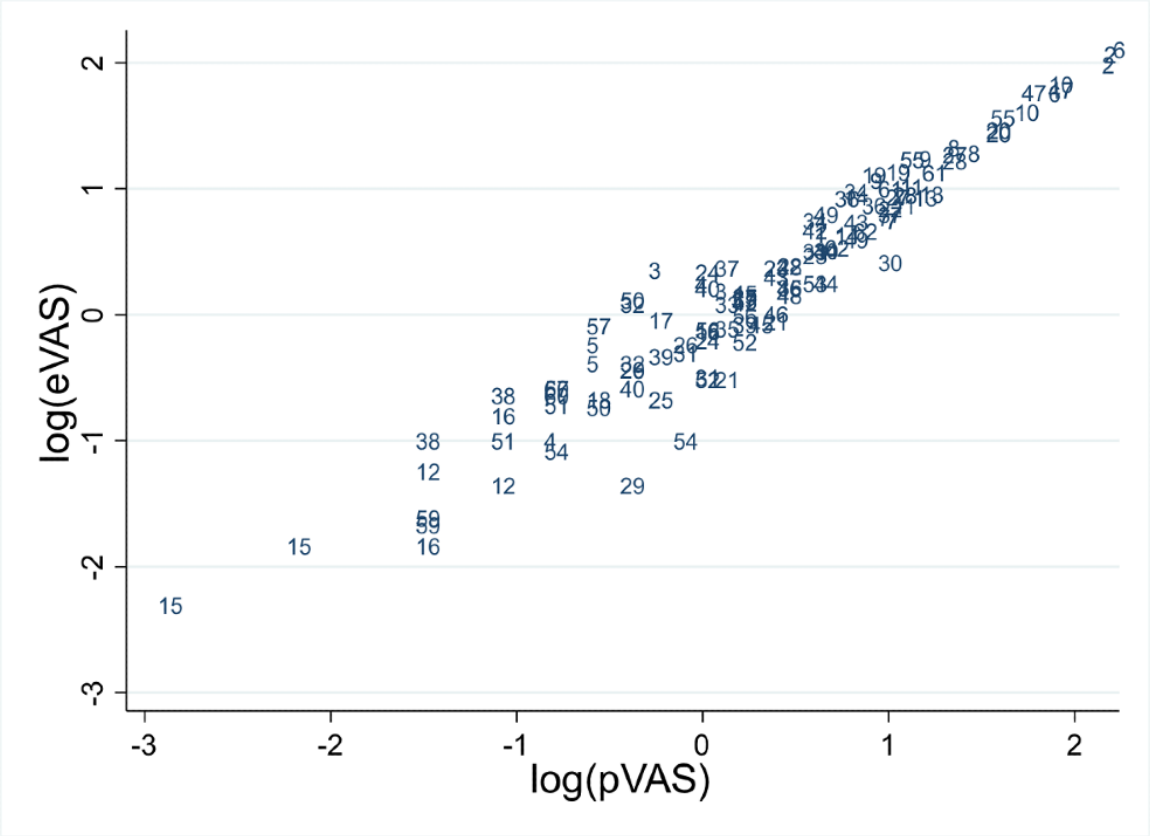


Figure 6. Scatter plot of the data (log) in adults. Points are represented by subject number. eVAS: electronic visual analog scale; pVAS: paper visual analog scale.

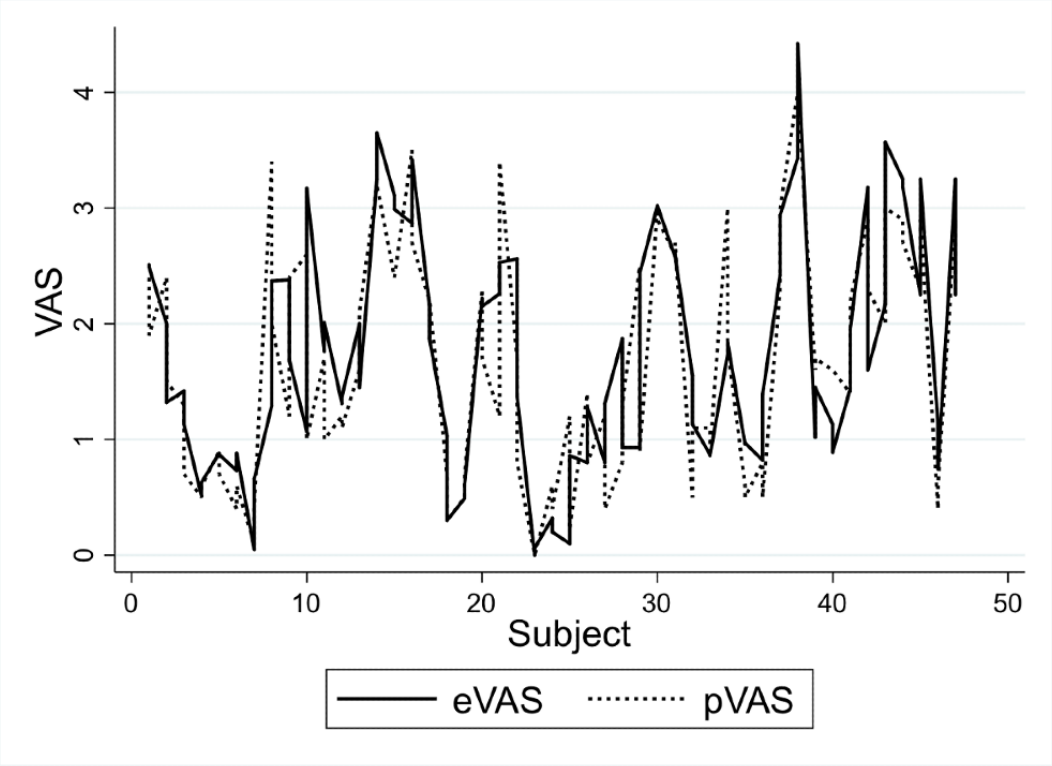


Figures 7 and 8 show the agreement between the methods for children and adolescents, and adults, respectively.

The two-sample Wilcoxon rank-sum test for comparing methods was not significant (children and adolescents, *P*=.48; adults, *P*=.73). The normality of residuals of the model for the child and adolescent group and the adult group showed a centered distribution (children and adolescents: Shapiro-Francia test, *P*=.05 and Shapiro-Wilks test, *P*=.06; adults: Shapiro-Francia test, *P*=.13 and Shapiro-Wilks test, *P*=.21).

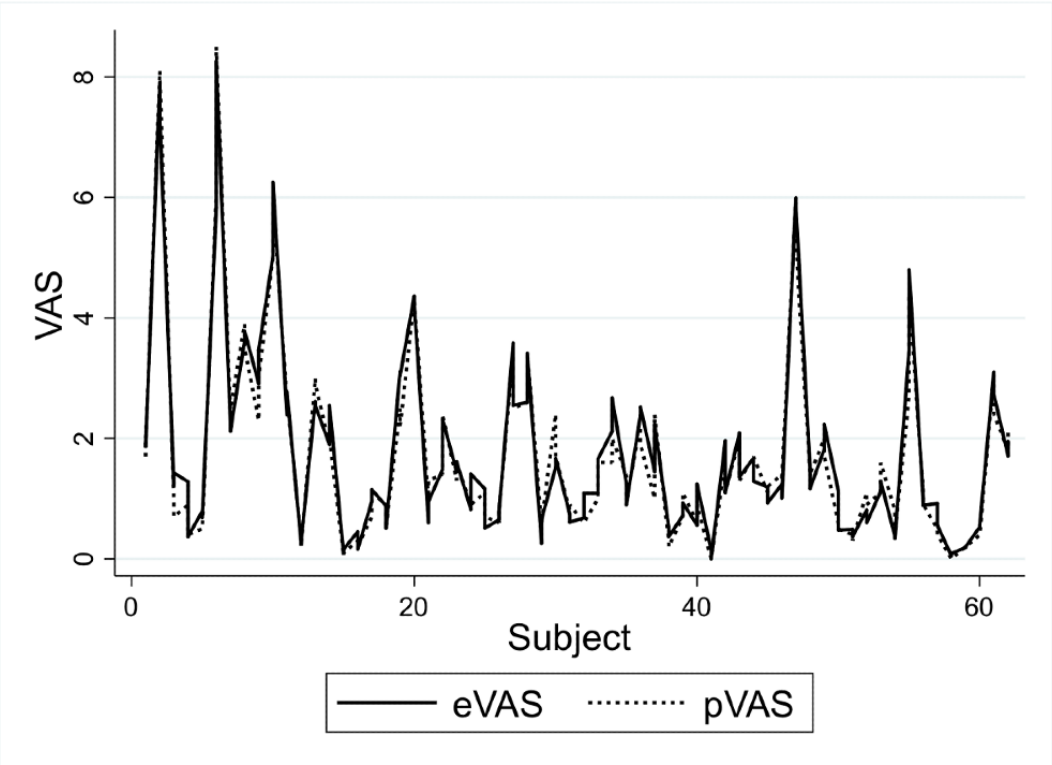
In children and adolescents, the intermethod reliability estimated by ICC reached the value of 0.80 (95% CI 0.70-0.87), indicating moderate-to-good reliability. The intramethod reliability estimated by ICCa reached the value of 0.80 (95% CI 0.69-0.87), indicating moderate-to-good reliability. For both coefficients, the length of the interval was less than 0.2.

Figure 7. Rating data for the two methods in children and adolescents. eVAS: electronic visual analog scale; pVAS: paper visual analog scale.



In adults, the intermethod reliability estimated by ICC reached the value of 0.94 (95% CI 0.91-0.96), indicating excellent reliability. The intramethod reliability estimated by ICCa reached the value of 0.94 (95% CI 0.91-0.96), indicating excellent reliability [32]. For both coefficients, the length of the interval was 0.1 or less.

Figure 8. Rating data for the two methods in adults. eVAS: electronic visual analog scale; pVAS: paper visual analog scale.



Discussion

Principal Findings

In the adult population, this study supported the interchangeability of the eVAS and pVAS owing to excellent intermethod and intramethod reliabilities as determined by an ICC value of 0.94 (95% CI 0.91-0.96) and ICCa value of 0.94 (95% CI 0.91-0.96), respectively. This supports previous findings by Bird et al who also reported excellent reliability in older adults (age range 65-85 years), using an apple iPad eVAS [11]. The eVAS also demonstrated excellent reliability, with regard to both individual ICC 0.90 (95% CI 0.82-0.95) and ICCa 0.97 (95% CI 0.95-1.0). Within a child and adolescent population, this study recognized the interchangeability of the eVAS and pVAS owing to moderate-to-good intermethod and intramethod reliabilities as determined by an ICC value of 0.80 (95% CI 0.70-0.87) and ICCa value of 0.80 (95% CI 0.69-0.87), respectively. Although this methodology has never been performed within this age range, Sánchez-Rodríguez et al reported that the mobile app "Painometer," which includes a 100-mm VAS scale, was concordant with its traditional counterparts from ages 12 to 19 years [13].

There are multiple methodological strengths of this study. Specifically, as previously recommended by Escalona-Marfil et al, this trial included block randomization during data collection for the eVAS and pVAS allocation sequence and also introduced a 5-minute interval between measures to reduce possible pain recall bias. This study suggested that fingerprints might remain visible on the screen during the eVAS recordings [17]. The current Australian study prevented traceable fingerprints by wiping clean the tablet's screen at the end of each recording. For the first time, this study reports the interchangeability of using eVAS compared with pVAS in pediatric participants from 10 years of age.

Occasionally, when using the traditional pVAS, patients may draw the line before the zero (0) or after the 100 point; therefore, these scores become invalid. Instead, the introduction of the eVAS app as part of regular pain evaluation prevents patients from recording their pain level outside the pVAS line, subsequently avoiding invalid scores.

For clinicians and researchers, especially those involved in community nursing or domiciliary visits, the time, cost, and space savings of data storage using the eVAS may be considerable when compared with the traditional pVAS, where manual transcription into clinical notes is required. The proposed eVAS app allows for automatic calculation of the VAS score, preventing possible human errors while using a ruler.

Patients may draw multiple lines on the pVAS or use inappropriate pens or thick highlighters. This can become confusing for clinicians to thoroughly interpret and record the intended results. A recent scoping review of systematic reviews highlighted that mHealth and hand-held electronic devices allow for accurate and complete medical documentation, providing instant access to reliable health data that may support clinical decision making [33]. Novel mHealth tools, such as the eVAS app, may make the work of health professionals even more efficient and increase reliability during their clinical assessments.

The eVAS app allows for the objective monitoring and recording of patients' pain levels. This mHealth tool might become advantageous for those patients living in geographically remote areas, where limited access to specialists is apparent. Patients and parents/caregivers may not always be required to visit the hospital, consequently saving the time and money required to travel long distances from rural areas. If promptly introduced within different pediatric and adult pain clinic services, the eVAS may support early pain detection, preventing incidences of unnecessary prolonged pain, with a consequent improvement in the patient's quality of life. This improved clinical management of pain may also lead to a reduction in absence from school or work.

Although not utilized within this study for data analysis, the eVAS app is capable of recording the time and day when the measures are taken. This important feature could be integrated within clinical settings and automatically reported within patient clinical records to highlight any diurnal variation in pain perception. Future trials could therefore also investigate the possible fluctuation of pain within a day. Notably, this may provide greater understanding of the complex nature of pain in response to environmental conditions or treatment plans [34].

This study further adds to the growing body of evidence that supports the use of digital technologies in health care. eHealth and mHealth have already been extensively used as tools for education, diagnosis, and management of pathologies such as diabetes [35,36], pediatric rheumatology [37], polycystic ovarian syndrome [38], and alcoholism [39]. At present, there are limited approaches available that combine evidence-based practice with health apps [40]. Portelli et al reported that most of the current apps available for pain management are rarely supported by an evidence base and may be misleading with their claims [41]. In addition, there are still limited regulations regarding data privacy for information collected from these apps [42]. Alarmingly, Blenner et al highlighted that many apps for diabetes management sold data to third parties without disclosure, even with a privacy policy stating that data were not going to be shared for commercial benefit [43]. This indicates the importance of further high-quality research into mHealth and eHealth regarding data privacy. More effort is also needed with regard to educating patients and practitioners in the use of apps that fully adhere to the guidelines clearly set by the WHO in evaluating digital health outcomes [18,44,45].

Limitations

There are some limitations that should be considered while interpreting the findings of this study. First, despite a recruitment effort, a balanced number of children and adolescents (n=47, 43%) and adults (n=62, 57%) was not obtained. This was due to logistical school issues in obtaining signed consent forms from parents during busy school terms. The relatively smaller pediatric sample size may have had an impact on the overall ICC values obtained from the child and adolescent population, in comparison with the adult population. Second, it should be noted that data were collected in a convenient sample of people from the community who were not exposed to high levels of

pressure pain with the Wagner Force Dial. Future studies may consider testing the eVAS with different intensities of pressure that would be deemed ethically acceptable. Third, a 1-minute interval is typically considered a suitable time gap to measure pain generated by pressure application [22,23]. During this trial, a 5-minute gap was adopted to record the pain generated by the pressure application (8.5 kg/cm^2) on the participant's thumb. The 5-minute gap was chosen to reduce any possible pain recall, especially among pediatric participants and to allow full recovery of sensory function of the thumb. There is possible anchoring bias on repetition of the two tests within a 5-minute interval. A 1-minute interval is deemed appropriate to assess a noxious stimulus without temporal summation [22,23,26]. Within the concept of pain measurement, recall bias may include a psychosocial aspect acknowledging that pain may be amplified or reduced after an extended period of time [46]. Additionally, pain perception within a population may fluctuate day to day [47]. However, for the purpose of this cross-sectional study and for determining interchangeability between the pVAS and eVAS, it is imperative that a single controlled stimulus is used and that measurements are undertaken within the same environment to eliminate confounding. Finally, although the child and adolescent population had slightly lower reliability relative to the adult population, this evidence supports the use of the eVAS in the pediatric population. A possible cause for the lower reliability relative to the adult population is the difference in scale length across platforms. The eVAS line width was 13.5 cm compared with 10 cm in the pVAS. Conceptual understanding of scales may differ between adult, and child and adolescent groups [48]. To encourage consistent conceptual understanding of the study in a large age range (10-75 years), the participants were made aware of the difference in sizes of the scales and asked to mark in a ratio. It is plausible that owing to the possible limited understanding of spatial ratio, there might be an impact on the results from younger participants [48].

In conclusion, the use of technology by children, adolescents, and adults is growing and is evident across multiple settings [49,50]. This study highlights the need for further investigation regarding the transferability of an eVAS pain app to different smartphone and tablet screen sizes that are already largely accessible within the community.

Clinical Implications

Monitoring and evaluating digital health interventions can be challenging, but have become requirements within the mHealth and eHealth fields [18]. This study specifically supports the adoption of these easy-to-use and validated pain assessment mHealth methods that have excellent reliability in adults and moderate-to-good reliability in children and adolescents. The emerging field of digital health presents an evolving cultural shift within health care settings. The growing use of digital mHealth has the potential to improve pain management. eHealth and mHealth have the ability to improve adherence to pain reporting [51,52], allow real-time data capture [53], and improve communication between practitioners and patients [54].

Conclusion

This study provides supporting evidence on the interchangeability of the eVAS and pVAS in child and adolescent, and adult populations. The introduction of similar validated eVAS pain apps may greatly increase the quality of reliable data accessible to clinicians, thereby improving the well-being of symptomatic patients. Most importantly, the use of mHealth in pain management may also facilitate timely clinical decisions, improve patients' self-management and overall awareness in the progression of their pain levels, and become an integrated approach consistent with the eHealth goals of the WHO and Australian Health Authorities [18,19]. Further research is needed on the use of these pain apps among symptomatic children, adolescents, and adults to ascertain the possible impacts of this new technology in these populations [55,56].

Acknowledgments

The authors would like to thank John Monash Science School (Clayton, Victoria) and KBH Brumbies Hockey Club (Mont Albert North, Victoria) for their generous support of this project. In particular, we would like to thank KBH Brumbies Hockey Club president Mike Hanlon and coach Daniel Brennan; the principal team of JMSS Peter Corkill, Sally Cheah-Johnson, and Andrew Chisholm; and JMSS teaching staff Adriana Abels and Peter Thompson for their enthusiasm and support.

Conflicts of Interest

None declared.

References

1. Rabbitts JA, Holley AL, Groenewald CB, Palermo TM. Association Between Widespread Pain Scores and Functional Impairment and Health-Related Quality of Life in Clinical Samples of Children. The Journal of Pain 2016 Jun;17(6):678-684. [doi: 10.1016/j.jpain.2016.02.005]
2. Hersh AO, Salimian PK, Weitzman ER. Using Patient-Reported Outcome Measures to Capture the Patient's Voice in Research and Care of Juvenile Idiopathic Arthritis. Rheumatic Disease Clinics of North America 2016 May;42(2):333-346. [doi: 10.1016/j.rdc.2016.01.004]
3. Shields BJ, Cohen DM, Harbeck-Weber C, Powers JD, Smith GA. Pediatric Pain Measurement Using a Visual Analogue Scale: A Comparison of Two Teaching Methods. Clin Pediatr (Phila) 2016 Jul 02;42(3):227-234. [doi: 10.1177/000992280304200306]

4. Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B. The Faces Pain Scale – Revised: toward a common metric in pediatric pain measurement. Pain 2001;93(2):173-183. [doi: 10.1016/s0304-3959(01)00314-1]
5. Tomlinson D, von Baeyer CL, Stinson JN, Sung L. A systematic review of faces scales for the self-report of pain intensity in children. Pediatrics 2010 Nov;126(5):e1168-e1198. [doi: 10.1542/peds.2010-1609] [Medline: 20921070]
6. Ruskin D, Lalloo C, Amaria K, Stinson JN, Kewley E, Campbell F, et al. Assessing Pain Intensity in Children with Chronic Pain: Convergent and Discriminant Validity of The 0 To 10 Numerical Rating Scale in Clinical Practice. Pain Research and Management 2014;19(3):141-148. [doi: 10.1155/2014/856513]
7. Zisk-Rony RY, Lev J, Haviv H. Nurses' Report of In-hospital Pediatric Pain Assessment: Examining Challenges and Perspectives. Pain Management Nursing 2015 Apr;16(2):112-120. [doi: 10.1016/j.pmn.2014.05.003]
8. Quinn BL, Sheldon LK, Cooley ME. Pediatric Pain Assessment by Drawn Faces Scales: A Review. Pain Management Nursing 2014 Dec;15(4):909-918. [doi: 10.1016/j.pmn.2014.03.004]
9. Davis TL, DiClemente R, Prietula M. Taking mHealth Forward: Examining the Core Characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
10. Ryu S. Book Review: mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
11. Bird M, Callisaya ML, Cannell J, Gibbons T, Smith ST, Ahuja KD. Accuracy, Validity, and Reliability of an Electronic Visual Analog Scale for Pain on a Touch Screen Tablet in Healthy Older Adults: A Clinical Trial. Interact J Med Res 2016 Jan 14;5(1):e3. [doi: 10.2196/ijmr.4910]
12. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with Pain™: User-Centred Design of a Web- and Mobile-Based Self-Management Program for Youth with Chronic Pain Based on Identified Health Care Needs. Pain Research and Management 2014;19(5):257-265. [doi: 10.1155/2014/935278]
13. Sánchez-Rodríguez E, de la Vega R, Castarlenas E, Roset R, Miró J. An App for the Assessment of Pain Intensity: Validity Properties and Agreement of Pain Reports When Used with Young People. Pain Med 2015 Oct 01;16(10):1982-1992. [doi: 10.1111/pme.12859]
14. Sun T, West N, Ansermino JM, Montgomery CJ, Myers D, Dunsmuir D, et al. A smartphone version of the Faces Pain Scale-Revised and the Color Analog Scale for postoperative pain assessment in children. Paediatr Anaesth 2015 Oct 28;25(12):1264-1273. [doi: 10.1111/pan.12790]
15. Stinson JN, Stevens BJ, Feldman BM, Streiner D, McGrath PJ, Dupuis A, et al. Construct validity of a multidimensional electronic pain diary for adolescents with arthritis. Pain 2008 Jun;136(3):281-292. [doi: 10.1016/j.pain.2007.07.002] [Medline: 17723279]
16. Luca NJ, Stinson JN, Feldman BM, Benseler SM, Beaton D, Campillo S, et al. Validation of the Standardized Universal Pain Evaluations for Rheumatology Providers for Children and Youth (SUPER-KIDZ). J Orthop Sports Phys Ther 2017 Oct;47(10):731-740. [doi: 10.2519/jospt.2017.7375]
17. Escalona-Marfil C, Coda A, Ruiz-Moreno J, Riu-Gispert L, Gironès X. Validation of an Electronic Visual Analog Scale mHealth Tool for Acute Pain Assessment: Prospective Cross-Sectional Study. J Med Internet Res 2020 Feb 12;22(2):e13468 [FREE Full text] [doi: 10.2196/13468] [Medline: 32049063]
18. World Health Organization. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. Geneva, Switzerland: World Health Organization; 2016.
19. eHealth Strategy for NSW Health: 2016-2026. NSW Government. URL: https://www.health.nsw.gov.au/eHealth/Pages/default.aspx [accessed 2020-09-02]
20. Shields BJ, Palermo TM, Powers JD, Grewe SD, Smith GA. Predictors of a child's ability to use a visual analogue scale. Child Care Health Dev 2003 Jul;29(4):281-290 [FREE Full text] [doi: 10.1046/j.1365-2214.2003.00343.x] [Medline: 12823333]
21. Breivik H, Borchgrevink P, Allen S, Rosseland L, Romundstad L, Breivik Hals E, et al. Assessment of pain. British Journal of Anaesthesia 2008 Jul;101(1):17-24. [doi: 10.1093/bja/aen103]
22. RANDOM.ORG. URL: https://www.random.org/sequences/ [accessed 2019-01-01]
23. Cole LJ, Farrell MJ, Gibson SJ, Egan GF. Age-related differences in pain sensitivity and regional brain activity evoked by noxious pressure. Neurobiology of Aging 2010 Mar;31(3):494-503. [doi: 10.1016/j.neurobiolaging.2008.04.012]
24. Balaguier R, Madeleine P, Vuillerme N. Intra-session absolute and relative reliability of pressure pain thresholds in the low back region of vine-workers: effect of the number of trials. BMC Musculoskelet Disord 2016 Aug 18;17(1). [doi: 10.1186/s12891-016-1212-7]
25. Buchanan HM, Midgley J. Evaluation of pain threshold using a simple pressure algometer. Clin Rheumatol 1987 Dec;6(4):510-517. [doi: 10.1007/bf02330587]
26. Frank L, McLaughlin P, Vaughan B. The repeatability of pressure algometry in asymptomatic individuals over consecutive days. International Journal of Osteopathic Medicine 2013 Sep;16(3):143-152 [FREE Full text] [doi: 10.1016/j.ijosm.2012.10.004]
27. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. Nephrology Dialysis Transplantation 2010 Jan 12;25(5):1388-1393. [doi: 10.1093/ndt/gfp732]

28. Gignac GE, Szodorai ET. Effect size guidelines for individual differences researchers. Personality and Individual Differences 2016 Nov;102:74-78. [doi: 10.1016/j.paid.2016.06.069]
29. Stata: Software for Statistics and Data Sciences. URL: https://www.stata.com/ [accessed 2019-05-02]
30. Portney L, Watkins M. Foundations of Clinical Research: Applications to Practice. London, England: Pearson; 2000.
31. Gwet K. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. California, United States: Advanced Analytics; 2014.
32. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
33. Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C. Evidence of effectiveness of health care professionals using handheld computers: a scoping review of systematic reviews. J Med Internet Res 2013;15(10):e212 [FREE Full text] [doi: 10.2196/jmir.2530] [Medline: 24165786]
34. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S240-S252. [doi: 10.1002/acr.20543] [Medline: 22588748]
35. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
36. Coda A, Sculley D, Santos D, Girones X, Acharya S. Exploring the Effectiveness of Smart Technologies in the Management of Type 2 Diabetes Mellitus. J Diabetes Sci Technol 2017 May 25;12(1):199-201. [doi: 10.1177/1932296817711198]
37. Coda A, Sculley D, Santos D, Girones X, Brosseau L, Smith DR, et al. Harnessing interactive technologies to improve health outcomes in juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2017 May 16;15(1):40 [FREE Full text] [doi: 10.1186/s12969-017-0168-y] [Medline: 28511689]
38. Boyle JA, Xu R, Gilbert E, Kuczynska-Burggraf M, Tan B, Teede H, et al. Ask PCOS: Identifying Need to Inform Evidence-Based App Development for Polycystic Ovary Syndrome. Semin Reprod Med 2018 Dec;36(1):59-65. [doi: 10.1055/s-0038-1667187] [Medline: 30189452]
39. Cohn AM, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
40. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a Pain App for That": Review of Patient-targeted Smartphone Applications for Pain Management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
41. Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. British Journal of Pain 2016 Jul 08;10(3):135-140. [doi: 10.1177/2049463716638700]
42. Rosenfeld L, Torous J, Vahia IV. Data Security and Privacy in Apps for Dementia: An Analysis of Existing Privacy Policies. Am J Geriatr Psychiatry 2017 Aug;25(8):873-877. [doi: 10.1016/j.jagp.2017.04.009] [Medline: 28645535]
43. Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C, Andrews LB. Privacy Policies of Android Diabetes Apps and Sharing of Health Information. JAMA 2016 Mar 08;315(10):1051-1052. [doi: 10.1001/jama.2015.19426] [Medline: 26954415]
44. Zaidan BB, Haiqi A, Zaidan AA, Abdulnabi M, Kiah ML, Muzamel H. A security framework for nationwide health information exchange based on telehealth strategy. J Med Syst 2015 May 3;39(5):51. [doi: 10.1007/s10916-015-0235-1] [Medline: 25732083]
45. World Health Organization. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. Geneva, Switzerland: World Health Organization; 2018.
46. Isselée H, De LA, Bogaerts K, Lysens R. Long-term fluctuations of pressure pain thresholds in healthy men, normally menstruating women and oral contraceptive users. Eur J Pain 2001;5(1):27-37. [doi: 10.1053/eujp.2000.0213] [Medline: 11394920]
47. Hutchings A, Calloway M, Choy E, Hooper M, Hunter DJ, Jordan JM, et al. The Longitudinal Examination of Arthritis Pain (LEAP) study: relationships between weekly fluctuations in patient-rated joint pain and other health outcomes. J Rheumatol 2007 Nov;34(11):2291-2300. [Medline: 17937461]
48. Sophian C. Perceptions of proportionality in young children: matching spatial ratios. Cognition 2000 May 15;75(2):145-170. [doi: 10.1016/s0010-0277(00)00062-7] [Medline: 10771276]
49. Edwards S, Henderson M, Gronn D, Scott A, Mirkhil M. Digital disconnect or digital difference? A socio-ecological perspective on young children's technology use in the home and the early childhood centre. Technology, Pedagogy and Education 2016 Apr 06;26(1):1-17. [doi: 10.1080/1475939x.2016.1152291]
50. Digital lives of older Australians. Australian Communications and Media Authority. 2020. URL: https://www.acma.gov.au/publications/2016-08/publication/digital-lives-older-australians [accessed 2020-06-09]

51. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 2003 Jul;104(1-2):343-351. [doi: 10.1016/s0304-3959(03)00040-x] [Medline: 12855344]
52. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004 Feb;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005] [Medline: 14736583]
53. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015 Dec;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
54. Jacob E, Pavlish C, Duran J, Stinson J, Lewis MA, Zeltzer L. Facilitating pediatric patient-provider communications using wireless technology in children and adolescents with sickle cell disease. J Pediatr Health Care 2013 Jul;27(4):284-292 [FREE Full text] [doi: 10.1016/j.pedhc.2012.02.004] [Medline: 22446036]
55. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A Pilot Comparison of a Smartphone App With or Without 2-Way Messaging Among Chronic Pain Patients: Who Benefits From a Pain App? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
56. Sánchez-Rodríguez E, Castarlenas E, de la Vega R, Roset R, Miró J. On the electronic measurement of pain intensity: Can we use different pain intensity scales interchangeably? J Health Psychol 2017 Nov 09;22(13):1658-1667. [doi: 10.1177/1359105316633284] [Medline: 26966092]

Abbreviations

AHP: allied health professional
eVAS: electronic visual analog scale
pVAS: paper visual analog scale
VAS: visual analog scale
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by M Lozano-Lozano, R Cruz Martínez; comments to author 28.03.20; revised version received 09.07.20; accepted 26.07.20; published 17.09.20.

Please cite as:
Turnbull A, Sculley D, Escalona-Marfil C, Riu-Gispert L, Ruiz-Moreno J, Gironès X, Coda A
Comparison of a Mobile Health Electronic Visual Analog Scale App With a Traditional Paper Visual Analog Scale for Pain Evaluation: Cross-Sectional Observational Study
J Med Internet Res 2020;22(9):e18284
URL: http://www.jmir.org/2020/9/e18284/
doi:10.2196/18284
PMID:32940621

Original Paper

Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study

Fabio Masina[1*], PhD; Valeria Orso[1*], PhD; Patrik Pluchino[1,2], PhD; Giulia Dainese[3], MS; Stefania Volpato[3], MS; Cristian Nelini[4], PhD; Daniela Mapelli[1,2], PhD; Anna Spagnolli[1,2], PhD; Luciano Gamberini[1,2], PhD

[1]Human Inspired Technologies Research Center, University of Padova, Padova, Italy

[2]Department of General Psychology, University of Padova, Padova, Italy

[3]L'Incontro Social Enterprise, Castelfranco Veneto, Italy

[4]Day Center for Severe Acquired Brain Injury, Opere Pie—Istituto Pubblico di Assistenza e Beneficienza, Pederobba, Italy

[*]these authors contributed equally

Corresponding Author:
Luciano Gamberini, PhD
Human Inspired Technologies Research Center
University of Padova
Via Luzzatti, 4
Padova
Italy
Phone: 39 049 827 5796
Email: luciano.gamberini@unipd.it

Abstract

Background: Voice assistants allow users to control appliances and functions of a smart home by simply uttering a few words. Such systems hold the potential to significantly help users with motor and cognitive disabilities who currently depend on their caregiver even for basic needs (eg, opening a door). The research on voice assistants is mainly dedicated to able-bodied users, and studies evaluating the accessibility of such systems are still sparse and fail to account for the participants' actual motor, linguistic, and cognitive abilities.

Objective: The aim of this work is to investigate whether cognitive and/or linguistic functions could predict user performance in operating an off-the-shelf voice assistant (Google Home).

Methods: A group of users with disabilities (n=16) was invited to a living laboratory and asked to interact with the system. Besides collecting data on their performance and experience with the system, their cognitive and linguistic skills were assessed using standardized inventories. The identification of predictors (cognitive and/or linguistic) capable of accounting for an efficient interaction with the voice assistant was investigated by performing multiple linear regression models. The best model was identified by adopting a selection strategy based on the Akaike information criterion (AIC).

Results: For users with disabilities, the effectiveness of interacting with a voice assistant is predicted by the Mini-Mental State Examination (MMSE) and the Robertson Dysarthria Profile (specifically, the ability to repeat sentences), as the best model shows (AIC=130.11).

Conclusions: Users with motor, linguistic, and cognitive impairments can effectively interact with voice assistants, given specific levels of residual cognitive and linguistic skills. More specifically, our paper advances practical indicators to predict the level of accessibility of speech-based interactive systems. Finally, accessibility design guidelines are introduced based on the performance results observed in users with disabilities.

(J Med Internet Res 2020;22(9):e18431) doi:10.2196/18431

KEYWORDS

voice assistants; accessibility; cognitive functions; disability; ambient assisted living

Introduction

Background

Voice-activated technologies are becoming pervasive in our everyday life [1,2]. In 2017, 46% of Americans reported using voice-activated technologies [3-5]. One of the most prominent application domains is the domestic environment, where voice assistants, a branch of voice-activated technologies, allow the user to control and interact with several home appliances in a natural way by uttering voice commands [6,7]. When integrated into a smart house, voice assistants allow the user to perform numerous everyday actions without the need to move and reach the actual object. More specifically, the user can operate all the devices that are connected, ranging from switching the lights on and off to opening and closing the doors and windows, for instance.

Research on voice assistants is focused mainly on the general population. Indeed, the studies investigating user experience and usability of voice assistants mainly involved able-bodied users [3,8-11], thereby neglecting a broad community of users with disabilities. However, people suffering from motor and cognitive impairments would significantly benefit from the possibility of controlling home appliances and personal devices remotely. Voice assistants hold the potential to enable individuals with disabilities to govern their houses without the need to constantly depend on caregivers [3,12].

One of the obvious barriers that some users with disabilities can encounter by interacting with voice assistants is related to speech impairments [13] that are a frequent secondary consequence of motor disorders [14]. Although most voice assistants exploit machine learning algorithms to adapt to the user and increase their speech recognition accuracy over time [15,16], these systems are still designed and developed for people with clear and intelligible speech. Thus, the difficulty of clearly utter sentences and speaking with adequate vocal intensity may represent a relevant accessibility challenge of voice assistants. The accessibility of voice assistants has not been thoroughly investigated yet. In this study, we explored how users with motor, linguistic, and cognitive disabilities interact with a commercial voice assistant in a natural situation. More specifically, the aim was to investigate the role of cognitive and linguistic functions to predict the performance of individuals affected by physical, linguistic, and cognitive difficulties in interacting with a voice assistant.

Voice Assistants for Users With Disabilities

Studies investigating the interaction between users with disabilities and voice assistants are still sparse. However, some evidence is starting to shed light in this field. Recently, Pradhan and colleagues [7] investigated the opinions of disabled users who regularly deploy voice assistants by examining their reviews. Most comments (about 86%) were positive, highlighting how the device has made it easier to accomplish specific tasks autonomously (eg, playing songs). The complaints were mainly focused on the lack of desired features, yet users pointed out that the main challenges they have in interacting with the voice assistant were due to the need to speak aloud and respect a precise timing for uttering the command. On the whole, these findings were confirmed by a following interview-based study with users with disability [7].

Ballati and colleagues [17] investigated to what extent people affected by speech impairments could be understood by three different voice assistants available off the shelf. More specifically, accuracy in speech processing was tested using sentences extracted from the TORGO database, which includes the recordings of 8 English speakers with dysarthria [18]. The sentences extracted were spoken to the voice assistants, and the accuracy across systems was compared. Each system processed the sentences one by one, while the experimenter scored the system accuracy with respect to the ability of the system to understand the sentence and consistency of the answer by the system. Results of this study revealed a general speech recognition accuracy of 50% to 60%, with all three systems having similar performance. These findings were partially confirmed with dysarthric Italian patients [19], where authors found different performance accuracy across the voice assistants.

While insightful, the studies reported above have limitations that might make it challenging to generalize the results. First, the actual speech abilities of the users were not assessed because they were either self-reported [7] or not reported at all [17,19]. This approach fails to provide clear indications for the design of voice assistants, as it does not highlight the users' needs. In addition, previous studies focused on speech abilities, neglecting cognitive skills. Cognitive skills were proven to affect the ability to operate a voice-controlled device by Weiner and colleagues [20]. In this study, the voice-controlled system showed a decrease in accuracy of speech recognition when the speakers suffered from Alzheimer disease or age-related cognitive decline. Furthermore, patients with Alzheimer disease experience difficulties interacting with a voice-controlled robot because of the timing imposed by the device [21].

Some of the previous research [17,19] did not even involve humans as participants, as they relied on prerecorded sentences. While this ensures high reliability in terms of assessing the robustness of the system, it also fails to account for the variability of individual performances and motivation behind the actual use of the device. Likewise, Pradhan and colleagues [7] found that about a third of the reviews analyzed were written by caregivers, who may have reported their viewpoint, misleading the perspective of the person they assisted. Finally, to the best of the writers' knowledge, the research available so far was conducted in laboratory settings, where the background noise is controlled, if any, and where there are no group interactions, as is likely to happen in a household.

Speech and Cognitive Factors Accounting for User Performance

Speech and cognitive skills play a significant role in the ability to effectively control voice assistants [17,19,22]. To properly convey a voice command, users must adequately control the speed and rhythm of speech. As reported in a previous study, speech disfluency can represent an accessibility barrier to voice assistants. For instance, long hesitations or pauses can be misinterpreted by the system as a sentence delimiter [23], causing an alteration of speech segmentation. Moreover, users must be able to correctly articulate words, especially

multisyllable words (eg, temperature) or specific words that may require more effort to be articulated [24]. A further aspect to properly interact with these devices is the voice intensity, which should be sufficiently loud to make voice assistants detect and segment the sounds [7].

Along with these speech skills, cognitive abilities are required to utter a command. The user must remember specific keywords and specific sequences of words to operate the system. These abilities involve memory functions, specifically long-term memory and working memory, both crucial when interacting with voice interfaces [25]. In addition, the user must respect a specific timing to provide the commands, a capacity that counts on executive functions, namely a set of functions needed to plan and control actions [26]. Not least, to properly use a voice assistant, the user must also monitor the feedback of the system (which sometimes consists of simple lights) and correctly interpret it. Such skills rely on underlying attention processes.

Methods

Study Design

This study was meant to assess the accessibility of a commercial voice assistant. In particular, we investigated whether specific cognitive and/or linguistic skills were related to the effectiveness of the interaction. To this end, the study consisted of two phases. In phase 1, participants were involved in group sessions, in which they were invited to interact with the voice assistant by performing several realistic tasks in a living laboratory (eg, switching on the light). Each group session involved 4 participants. This choice was motivated by our desire to build a friendly and informal setting that could facilitate interaction and prevent the feeling of being in a testing situation. Group sessions were video recorded to allow offline analysis of participant performances. In phase 2, participants received an evaluation of their neuropsychological and linguistic functions. The two phases of the study took place in different settings and on separate days and required different experimental materials. The study was approved by the Ethics Committee of the Human Inspired Technologies Research Center, University of Padova, Italy (reference number 2019_39).

Participants

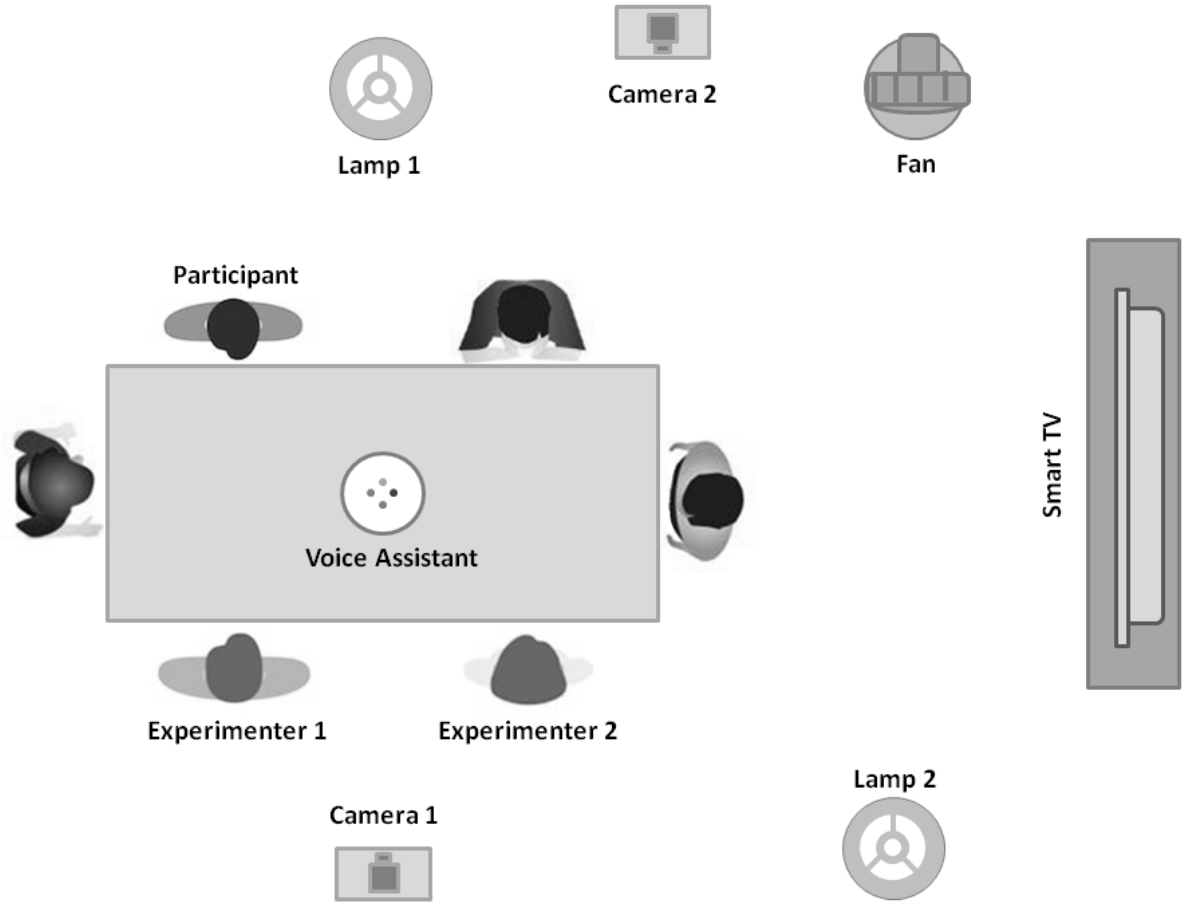
A total of 16 participants (9 males, 7 females) took part in the study. The mean age of the sample was 38.3 (SD 8.6) years (range 22 to 51 years). On average, they had 11.8 (SD 2.7) years of education (range 8 to 18 years). To partake in the study, participants had to meet the following inclusion criteria: (1) suffering from ascertained motor impairments and related language difficulties and (2) needing daily assistance from at least one caregiver. The sample was characterized by 6 participants affected by congenital disorders, 2 participants with neurodegenerative disorders, 4 participants affected by traumatic brain injury, and 4 participants with nontraumatic brain injury (ie, tumor). The heterogeneity of the sample well represents the population that can be found in daycare centers. Participants were indeed recruited from a daycare center for people with disabilities, with which the research team collaborates. Before enrollment, all invited participants received an explanation of the activity. Upon agreement, they were provided written informed consent (if necessary, the individual's legal guardian was informed about the scope and unfolding of the activity and gave the informed consent for the person they assisted to partake in the study). In any case, informed consent was given prior to their enrollment. Participants received no compensation for taking part in the study.

Phase 1: Interaction With the Voice Assistant

Setting

The first phase took place in a living laboratory. The room was furnished to resemble a living room with a large table in the middle. The voice assistant was placed at the center of the table, around which participants and experimenters were sitting (Figure 1). The laboratory was equipped with several devices that were connected to the voice assistant and could be controlled by prompting voice commands. All of the voice-controlled devices were placed so that users could easily see them. The room was also equipped with two camcorders to video record the sessions. One camera was placed above the table and enabled the observation of users' interactions with the voice assistant. The other camera served to record the outcomes of the interaction (Figure 1).

Figure 1. Representation of the experimental setting.



Equipment

For this study, a commercial voice assistant was deployed. More specifically, we chose to use Google Home (Google LLC), given its growing popularity. Two lamps and a floor fan were connected to smart plugs, which were in turn connected to the voice assistant, thereby enabling control of the switch on/off and light color change (for the lamps only). A 50-inch television was connected to Chromecast (Google LLC), which was in turn connected to Google Home. By doing this, it was possible to operate the TV using voice commands. For the video recordings, two video cameras were installed, one was a C920 Pro HD (Logitech) and the other one was a Handycam HDR-XR155E (Sony Europe BV).

Tasks

Participants were invited to individually prompt some commands to the voice assistant, as indicated by the experimenter. The tasks comprised turning on/off the fan and the lights, changing the color of the light, interacting with the TV (activating YouTube, Spotify, and Netflix), and making specific requests to the voice assistant (eg, “set an alarm for 1 pm”). The full list of commands that participants were asked to speak can be seen in Textbox 1.

Textbox 1. The list of voice commands that participants were asked to speak during the first phase of the study.

Fans

- Turning on/off

Lamps

- Turning on/off
- Changing colors
- Changing light intensity

TV (YouTube)

- Selecting videos
- Increasing/decreasing volume

TV (Netflix)

- Selecting movies
- Pausing movies
- Playing movies

TV (Spotify)

- Selecting songs
- Increasing/decreasing volume

Voice assistant

- Asking for the latest news
- Asking for the weather forecast
- Setting an alarm

Procedure

Participants were first welcomed in the living laboratory and invited to make themselves comfortable. They were reminded about the aim and the unfolding of the activity. In addition, they were shown the camcorders and after they all proved to be aware of them, the video recording started. At this point, the experimenter showed how the voice assistant worked by prompting some example commands and properly explained the correct sequence of words to convey the command. Next, participants were allowed to familiarize themselves with the voice assistant until they felt confident. When they considered themselves ready, the experimental session started. The experimenter asked each participant to individually perform the selected tasks (Textbox 1). The tasks were not proposed in a strict order across participants. To keep the session lively and prevent boredom and fatigue, the tasks were alternated across participants. Should a participant fail to accomplish a requested task (eg, the voice assistant did not respond in the expected manner), the experimenter gently encouraged them to try again. A fixed number of attempts was not set a priori to prevent participants from feeling frustrated as a consequence of repeated failed attempts. Participants were allowed to try until they felt comfortable.

Once the task list was completed by all participants, the experimenter asked them their impressions about the voice assistant in a semistructured group interview. The questions regarded an overall evaluation of the pleasantness of the voice assistant (from 1 to 10), in which rooms it would be more helpful, if they would like to have it in their own houses, and which additional functions they would like to control. Phase 1 took about 2.5 hours.

Phase 2: Neuropsychological and Linguistic Assessment

Data Collection

All of the participants involved in phase 1 received an individual examination by a trained neuropsychologist and a speech therapist, who were both blind to the outcomes of the users' performances with the voice assistant. Several assessment tools were selected and adopted. More specifically, the neuropsychological functions were assessed with the Addenbrooke's Cognitive Examination–Revised (ACE-R) [27] and the Frontal Assessment Battery (FAB) [28]. The linguistic assessment was conducted by collecting several measures, namely participant vocal intensity, and other speech production indices gathered using the standardized Italian version of the Robertson Dysarthria Profile [29]. The evaluation sessions took place in a quiet room at the daycare center where participants were recruited and lasted about 1.5 hours for the neuropsychological evaluation and 2 hours for the linguistic evaluation.

Neuropsychological and Linguistic Tests

The ACE-R [27] is a screening test originally proposed as an extension of the Mini-Mental State Examination (MMSE) [30]. The ACE-R allows the evaluation of 5 cognitive domains, attention/orientation, memory, verbal/category fluency, language, and visuospatial ability, in addition to providing the MMSE score. Attention/orientation is assessed by asking the participant about the date, season, and current location where the evaluation is taking place, as well as repeating 3 single words and doing serial subtractions. Memory consists of items that evaluate episodic and semantic memory. Verbal and category fluency require the ability to list in 1 minute as many words as possible complying with a verbal criterion and a category criterion. Language includes several subtasks, requiring speech comprehension, naming figures, repeating words and sentences, reading regular and irregular words, and writing. Finally, visuospatial ability consists of copying and drawing specific pictures.

With respect to the MMSE, it represents a general index of cognitive functioning ranging from 0 to 30. A score below 24 may indicate the presence of cognitive impairment [30].

Frontal Assessment Battery

The FAB [28] is a brief inventory for the evaluation of executive functions. It is composed by 6 subscales exploring domains: conceptualization (similarities test), mental flexibility (verbal fluency test), motor programming (Luria motor sequences), sensitivity to interference (conflicting instructions), inhibitory control (go/no-go test), and environmental autonomy (prehension behavior). Each domain consists of 3 items and is scored from 0 (unable to complete the requests) to 3 (fully able to fulfill the requests). The maximum overall score for the FAB is 18.

Vocal Intensity

Vocal intensity reflects the loudness of the voice. Physically, it represents the magnitude of the oscillations of the vocal folds, and it is measured in decibels (dB). In this study, vocal intensity was collected by using the PRAAT software [31], a tool for speech analysis. Participants were invited to repeat aloud a specific sentence (ie, “Turn off the light, turn on the TV” in their native language) for 5 minutes at a distance of 1.5 meters from the recording device.

Speech Production

An expert speech and language therapist assessed participant speech production. The protocol adopted for the evaluation was extracted from the Robertson Dysarthria Profile [29]. This test is divided into 8 subscales (ie, intelligibility, respiration, phonation, facial muscles, diadochokinesis, oral reflexes, prosody, articulation), each including several items. The therapist assigns a score on a 4-point scale (1 = severe, 2 = moderate, 3 = mild, 4 = normal) for each item of the test. In this study, the subscales considered were prosody and articulation. More specifically, for prosody (2 items) the items assessed the speed and rhythm of speech production. With regard to articulation (5 items), the items considered the ability to articulate single letters (consonants and vowels) and clustered letters (groups of consonants and multisyllable words), as well as the capacity to repeat sentences.

Data Analysis

The data analysis comprised analysis of the video recordings to assess the extent to which users were capable to effectively interacting with the voice assistant. The outcomes of the analysis were summarized into a performance index. The index was then associated with the neuropsychological and linguistic measures collected in the second phase of this study. Since the main purpose of this study was the identification of predictors (cognitive and/or linguistic) capable of accounting for an effective interaction with the voice assistant, multiple linear regression models were run.

Video Analysis

The two video streams recorded during the sessions were synchronized into a single video file using a video editing software. The resulting video was then imported into a dedicated software for the analysis (The Observer XT 12, Noldus Information Technology Inc). The analysis was conducted in two passes. During the first pass, two of the authors watched the videos and selected the events of interest: the experimenter’s requests, participants’ actions, and voice assistant’s responses. The two researchers then agreed on the events to code, defining the objective triggers detailing the beginning and the end of each. A trained coder was in charge of rating the videos.

For each participant, the number of attempts they made for each task request and the resulting outcome were coded. More specifically, the beginning of an attempt was coded when the experimenter prompted the participant to try to accomplish a given task. The attempt ended with either the actual activation of the intended function (successful outcome) or with a failure to observe the expected outcome (unsuccessful outcome). In particular, unsuccessful outcomes were further categorized based on the type of error made by the participants. Four categories of errors were identified:

- Timing errors included all of the unsuccessful outcomes caused by the participant not respecting the timing imposed by the system (eg, the participant uttered the waking command “Hey Google” and did not wait for the system to reply before prompting the full command)
- Phrasing errors comprised all the failed attempts that followed an incorrect sequence of words to prompt the command (eg, the participant saying “Hey Google...put the red the lamp” instead of “Hey Google...make the lamp red”)
- Comprehension errors referred to all mistakes participants made because they could not understand the experimenter’s request (eg, changing the color of the lamp instead of turning it off)
- Pronunciation errors included all of the failures that followed a wrong articulation of one or more words within the sentence (eg, participants struggling to pronounce words that were not in their native language, such as Netflix)

Participants’ attempts could also be coded as self-corrections (with successful or unsuccessful outcome) when the participant realized autonomously that the command was wrong and tried to amend it.

To understand whether participants were able to prompt commands to the voice assistant, an overall performance index was computed expressing the percentage of successful attempts and the total number of attempts. Importantly, self-corrections with successful outcomes were considered successful attempts whereas self-corrections with unsuccessful outcomes were considered unsuccessful attempts.

Neuropsychological and Linguistic Assessment

Regarding the neuropsychological measures, not all participants were able to complete all of the subscales of the ACE-R. More specifically, several participants could not fully complete some items of the ACE-R (eg, drawing a clock) because of their physical impairments (eg, dystonia). However, since all participants could complete at least the items of the MMSE, only the MMSE score was considered in the multiple linear regression models, in addition to the FAB score. With regard to the linguistic assessments, all the collected measures were considered in the regression models.

Multiple Linear Regression Models

Data were statistically analyzed using RStudio software version 1.2 (RStudio PBC). To investigate which predictors of the performance index (participant performances during the use of the voice assistant) are best, multiple linear regression models were adopted. In order to make accurate predictions, we considered, among several models, the one that best described the data. The best model was identified by adopting a selection strategy based on the Akaike information criterion (AIC). The AIC value provides an estimation of the quality of a model given several other candidate models. The AIC considers both the complexity of a model and its goodness of fit. According to the AIC, given a set of models, the one characterized by the lowest AIC is the best [32].

The neuropsychological and linguistic predictors entered in the models were the MMSE score, FAB score, vocal intensity (dB), and scores obtained from the 2 items of the prosody subscale and 5 items of the articulation subscale of the Robertson Dysarthria Profile. More specifically, the linear regression models were performed entering the predictors grouped into four clusters: (1) neuropsychological cluster (ie, MMSE and FAB), (2) vocal intensity cluster (ie, dB), (3) prosody cluster (ie, speed and rhythm), and (4) articulation cluster (ie, initial consonants, vowels, groups of consonants, multisyllable words, and repetition of sentences). The latter two clusters consisted of the items in the Robertson Dysarthria Profile. Since the forced entry method was adopted, the order in which predictors were entered in the model did not affect the results.

Results

Video Analysis

The performance index extracted from the video analysis shows that participant accuracy was on average 58.5% (SD 18.6%). The most frequent type of errors made by participants were phrasing errors (75/182, 41.2%). Participants mainly had problems uttering long commands, especially when they were required to respect a specific syntax. It should be noted that uttering the right sequences of words was not problematic to the same extent for all participants, as one participant never made this type of error, while one made it 21 times.

Timing errors were the second most frequent type of error (74/182, 40.7%), and they can be clustered into anticipatory timing errors and delayed timing errors. More specifically, as for the anticipatory timing errors, participants tended not to wait for the system to reply to the waking command before prompting the actual command. For one participant, respecting the timing seemed particularly difficult, as they made this type of error 30 times. To a lesser extent, with regard to the delayed timing errors, participants waited too long after the system had replied to the waking command. In many cases, the actual command overlapped to the system prompting the error message “Sorry, I don’t know how to help you.”

Less frequent were the comprehension errors (19/182, 10.4%) and pronunciation errors (14/182, 7.7%). Regarding the former, participants mainly tended to misunderstand the most complex commands (eg, playing a video on YouTube). Regarding the latter, users had some difficulties with English words, like Netflix. Nevertheless, the system could successfully respond even when they had strong dialectal stress.

Overall, all participants enjoyed the interaction with the voice assistant. Indeed, the general evaluation of the system was extremely positive, with a mean score of 9.4 (SD 1.2). As for the rooms in which participants would like to install the voice assistant, 8 of them suggested the bedroom and 4 the kitchen. On the whole, all participants would like to have a voice assistant at their own house. Finally, with regard to the functions that participants would have liked to implement in their own house, they mentioned playing music (n=5) and controlling the home automation (n=5), such as opening/closing windows and doors.

Interestingly, during the interaction with the voice assistant, several participants provided their spontaneous opinions highlighting the benefits and drawbacks of the system. For instance, P3 stated: “Since my shoulder hurts, it is useful because it is easier when I have to open doors.” However, P3 claimed as well: “sometimes it does not understand me and I am afraid to crash the Google program.” Another participant mentioned some difficulties as well, especially concerning the general utility of having a voice assistant at home. P9 stated: “I cannot think as before [the accident], it is not so easy to have such a device at home, it might not be useful.”

Neuropsychological and Linguistic Assessment

Table 1 shows the raw scores from participants in the neuropsychological and linguistic assessments made in the second phase of the study. With regard to the neuropsychological scores, participants showed a mean MMSE score of 26.1 and a mean FAB score of 12.6. Concerning the linguistic assessment, participants had a mean vocal intensity of 61.6 dB. Finally, the mean scores of the Robertson Dysarthria Profile indicated speed of speech of 2.7, and rhythm of speech of 2.6. Overall, these scores indicated mild to moderate prosody difficulties. Finally, the mean scores of items measuring articulation abilities showed mild issues regarding the pronunciation of initial consonants (mean 3.3), vowels (mean

3.3), groups of consonants (mean 3.2), multisyllable words (mean 3.3), and the repetition of sentences (mean 3.1).

Table 1. Summary of participant scores from the neuropsychological and linguistic assessments.

Measure	Mean score (SD)
Mini-Mental State Examination	26.1 (2.9)
Frontal Assessment Battery	12.6 (3.8)
Vocal intensity (dB)	61.6 (4.2)
Prosody	
Speed of speech production	2.7 (0.7)
Rhythm of speech production	2.6 (0.7)
Articulation	
Initial consonants	3.3 (0.6)
Vowels	3.3 (0.5)
Groups of consonants	3.2 (0.7)
Multisyllable words	3.3 (0.6)
Repetition of sentences	3.1 (0.6)

Multiple Linear Regression Models

In order to identify the best model to predict participant accuracy (assessed as the performance index), several multiple linear regression models were considered. Multimedia Appendix 1 shows all estimated models with their respective AIC scores. Comparing the AICs in all the models, model *ad* ($F_{6,9}$=4.91, *P*=.02, R^2=.77), which included the neuropsychological and articulation clusters, was the best one (AIC 130.69; Multimedia Appendix 1).

When checking for the coefficients of this model, 2 predictors were found to explain a significant amount of the variance of accuracy. The predictors that significantly accounted for accuracy were the MMSE (β=6.16, t_9=3.88, *P*=.004) and repetition of sentences (β=31.14, t_9=2.71, *P*=.02). Of importance, among the nonsignificant predictors, 3 (ie, initial consonant, group of consonants, and multisyllable words) had a variance inflation factor (VIF) >10 (tolerance statistics: 1/VIF<0.1), showing multicollinearity [33,34]. As a consequence, a new model was performed, removing all the nonsignificant and collinear predictors by entering only the MMSE and repetition of sentences. The results confirmed the previous model, namely that the MMSE and repetition of sentences were significant predictors of accuracy: MMSE (β=3.70, t_{13}=3.26, *P*=.006) and repetition of sentences (β=22.06, t_{13}=4.16, *P*=.001). The AIC value of this final model was 130.11, showing that it was the best model compared with the previous ones (Multimedia Appendix 1).

To test the assumptions of the linear regression model, diagnostic statistics were performed. The model met the assumption of independence (Durbin-Watson 2.29, *P*=.68). The Q-Q plot of standardized residuals suggested that the residuals were normally distributed. Tolerance statistics (1/VIF) indicated that multicollinearity was not a concern (MMSE tolerance .92; repetition of sentences tolerance .92).

The standardized values were .57 (MMSE) and .73 (repetition of sentences). The first value suggests that as the MMSE increases by 1 standard deviation (2.89 points), the performance index increases by 1 standard deviation as well (10.6%). This prediction is true only if the repetition of sentences is constant. The second standardized value predicts that every time the repetition of sentences improves by 1 standard deviation (0.6 points), the performance index increases of 1 standard deviation (13.6%). This interpretation is true only if the MMSE is fixed.

Discussion

Principal Findings

This work aimed to investigate whether cognitive and/or linguistic functions could predict the user's performance in operating an off-the-shelf voice assistant. To this end, a group of users suffering from motor and cognitive difficulties was invited to a living laboratory. The lab was purposefully equipped with a voice assistant connected to several smart devices (ie, TV, lamps, floor fan), and participants were asked to perform specific tasks following the experimenter's instructions. In order to assess user performances, interactions with the voice assistant were video recorded. Cognitive and linguistic functions were assessed with standardized inventories and subsequently related to the user performances with the voice assistant.

The performance index was found to be predicted by the overall cognitive abilities, as assessed by the score on the MMSE and by the ability to repeat sentences. In other words, a minimum level of residual cognitive functioning (ie, MMSE score above the cutoff [≥24]) is recommended to effectively operate a voice assistant. Among the linguistic skills, the ability to repeat sentences was necessary. These findings contribute to provide specific indications of the level of inclusion of commercial voice assistants.

More generally, the average accuracy was around 60%, extending previous findings that were limited to synthesized

utterances [17]. Different than the previous studies, we arranged a living lab and involved actual users with disabilities in a realistic group situation. This approach allowed us to identify and categorize the most prominent types of errors emerging during the interaction. The most frequent mistakes were phrasing errors (41.2%), highlighting the difficulties of participants to respect the syntax of the voice command, especially when the command was long. The second most frequent error consisted of the difficulty in respecting the timing imposed by the device (40.7%), as already reported by Pradhan and colleagues [7]. Specifically, participants uttered the command too quickly or too slowly, showing a tendency to ignore the feedback of the voice assistant. This was probably due to the lack of saliency of the feedback provided by Google Home after the waking command [35], which consists only of dim lights moving on top of the device. Additionally, two other types of errors emerged, relating to difficulty of comprehension of the request (10.4%) and pronunciation issues (7.7%). The latter was limited to English words. These findings suggest significant implications for the design of universal voice assistants. First, more salient feedback should be included to make it easier for users with disabilities to interact with the system. Additionally, the timing should be adjustable to better respond to the actual abilities of the user and adapt to their proficiency in using it over time. Finally, to increase the likelihood of users remembering how to operate the voice assistant, commands should include familiar words.

These results are particularly relevant because they provide new implications for the design of voice assistants using an inclusive design perspective that also considers users with special needs. On the other hand, these findings can provide an indication to caregivers, both family members and health care professionals, for choosing assistant technologies that are suitable for the people they assist. More specifically, the ability to interact and use voice assistants does not depend exclusively on linguistic skills, as it could seem. In fact, aspects related to cognitive functions, in particular the global level of cognitive functioning, seem to play a crucial role. Hence, linguistic and cognitive abilities predict performance with voice assistants. Users with severe cognitive impairment (MMSE score <18) [36] may still be able to use these systems effectively (performance index >50%) if their level of linguistic skills is normal (Robertson Dysarthria Profile = 4) [29], which somehow compensates for the cognitive difficulties. Similarly, a user with severe difficulty in articulating sentences may successfully use voice assistants if they have a normal level of cognitive functioning such that they can invoke compensative strategies. Taken together, these findings may serve as promising indicators to foresee the degree of accessibility of voice assistants. Importantly, the predictors employed in this study are extracted from standardized inventories that are highly widespread and administered in many clinical environments.

Finally, despite the mistakes, participants positively received the system and enjoyed their experience, consistent with the findings of Pradhan and colleagues [7]. Users found the system useful and reported that they would like to have it at their own houses. In addition, they suggested that such a system would be helpful in compensating for their difficulties with movements (eg, opening doors). The positive user opinions about the system revealed the general acceptance of voice assistants, highlighting the importance of using these mainstream systems in the field of assistive technologies in order to help users with disabilities regain some independence and increase their quality of life.

This study suggests that with specific and targeted adjustments a commercial voice assistant can be turned into an assistive technology that can effectively complement the individual's skills. Indeed, voice assistants could offer tremendous benefits. First of all, these systems are widespread and inexpensive compared with assistive technologies, which are often harder to find and costly. Furthermore, assistive technologies can be stigmatizing. The fear of feeling exposed and feelings of autonomy and dignity loss are significant barriers to the adoption of assistive technology [37]. On the other hand, the popularity of voice assistants, as well as their appealing design, may make them really inclusive technology, being helpful to individuals with or without disabilities.

Limitations

We acknowledge that this study has some limitations. First, the sample size was limited to 16 participants. Therefore, further studies should extend our findings with larger and even more heterogenous samples. In addition, we have explored a likely use scenario, where users interact with the voice assistant in a group situation, as happens in shared living environments. Nevertheless, future experiments should also investigate a use scenario in which the user operates the system individually to examine more closely the interaction between the individual and the voice assistant.

Conclusions

In this work we report on a group experiment involving users with motor, linguistic, and cognitive difficulties that was meant to predict participant performances based on their level of cognitive and linguistic skills. Previous studies did not involve actual users or consider their capabilities. For the first time, we conducted an experiment in a living lab with individuals with disabilities and provide a detailed report of their performances and difficulties. More importantly, participant performances showed they could be predicted by their residual level of cognitive and linguistic capabilities. In addition, these results contribute to the field of assistive technology by describing the different types of errors made by users and providing design implications.

The enthusiastic reaction of participants highlights the potential of voice assistants to provide or return some autonomy in basic activities, like turning the light on/off when they are lying in bed. Further research effort should be devoted to fine-tuning voice assistants to better serve users' needs and evaluating in the field to what extent the systems are actually helpful. To conclude, by polishing the existing widespread voice assistants, there will be the concrete opportunity to increase the quality of life of people with disabilities by providing them with truly inclusive technology.

Acknowledgments

This paper was supported by the project “Sistema Domotico IoT Integrato ad elevata Sicurezza Informatica per Smart Building” (POR FESR 2014-2020).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Linear regression models considered in the analyses with their respective values of R2, adjusted R2, and Akaike information criterion.
[DOCX File , 15 KB - jmir_v22i9e18431_app1.docx]

References

1. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
2. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
3. Pyae A, Joelsson TN. Investigating the usability and user experiences of voice user interface. 2018 Presented at: Proc 20th Int Conf Human-Computer Interact with Mob Devices Serv Adjun—MobileHCI 18; 2018; Barcelona. [doi: 10.1145/3236112.3236130]
4. Sayago S, Neves B, Cowan B. Voice assistants and older people: some open issues. Proceedings of the 1st International Conference on Conversational User Interfaces 2019;7:1-3. [doi: 10.1145/3342775.3342803]
5. Olmstead K. Nearly half of Americans use digital voice assistants, mostly on their smartphones. Washington: Pew Internet and American Life Project; 2017. URL: https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/ [accessed 2020-08-31]
6. Noda K. Google Home: smart speaker as environmental control unit. Disabil Rehabil Assist Technol 2018 Oct;13(7):674-675. [doi: 10.1080/17483107.2017.1369589] [Medline: 28830267]
7. Pradhan A, Mehta K, Findlater L. Accessibility came by accident: use of voice-controlled intelligent personal assistants by people with disabilities. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems 2018:1-13. [doi: 10.1145/3173574.3174033]
8. Corbett E, Weber A. What can i say? Addressing user experience challenges of a mobile voice user interface for accessibility. Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services 2016:72-82. [doi: 10.1145/2935334.2935386]
9. Lopatovska I, Oropeza H. User interactions with “Alexa” in public academic space. Proc. Assoc. Info. Sci. Tech 2019 Feb;55(1):309-318. [doi: 10.1002/pra2.2018.14505501034]
10. Lopatovska I, Rink K, Knight I, Raines K, Cosenza K, Williams H, et al. Talk to me: exploring user interactions with the Amazon Alexa. J Librariansh Inf Sci 2018 Mar 07;51(4):984-997. [doi: 10.1177/0961000618759414]
11. López G, Quesada L, Guerrero L. Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces BT. In: Nunes I, editor. Advances in Human Factors and System Interactions. Cham: Springer; 2018.
12. Branham S, Roy A. Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users. 2019 Presented at: ASSETS st Int ACM SIGACCESS Conf Comput Access; 2019; Pittsburgh. [doi: 10.1145/3308561.3353797]
13. De Russis L, Corno F. On the impact of dysarthric speech on contemporary ASR cloud platforms. J Reliable Intell Environ 2019 Jul 6;5(3):163-172. [doi: 10.1007/s40860-019-00085-y]
14. Duffy J. Motor speech disorders e-book: substrates, differential diagnosis, and management. Philadelphia: Elsevier Health Sciences; 2013.
15. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015 Jul 16;349(6245):261-266. [doi: 10.1126/science.aaa8685]
16. Hoy MB. Alexa, Siri, Cortana, and more: an introduction to voice assistants. Med Ref Serv Q 2018 Jan 12;37(1):81-88. [doi: 10.1080/02763869.2018.1404391]
17. Ballati F, Corno F, De Russis L. Hey Siri, do you understand me: virtual assistants and dysarthria. Intell Environ 2018:1 [FREE Full text]
18. Rudzicz F, Namasivayam AK, Wolff T. The TORGO database of acoustic and articulatory speech from speakers with dysarthria. Lang Resour Eval 2011 Mar 26;46(4):523-541. [doi: 10.1007/s10579-011-9145-0]

19. Ballati F, Corno F, De Russis L. Assessing virtual assistant capabilities with Italian dysarthric speech. 2018 Presented at: ASSETS - Proc 20th Int ACM SIGACCESS Conf Comput Access; 2018; Galway. [doi: 10.1145/3234695.3236354]
20. Weiner J, Engelbart M, Schultz T. Manual and automatic transcriptions in dementia detection from speech. 2017 Presented at: Proc Annu Conf Int Speech Commun Assoc INTERSPEECH; 2017; Stockholm. [doi: 10.21437/interspeech.2017-112]
21. Rudzicz F, Wang R, Begum M, Mihailidis A. Speech interaction with personal assistive robots supporting aging at home for individuals with Alzheimer's disease. ACM Trans Access Comput 2015 Jul 20;7(2):1-22. [doi: 10.1145/2744206]
22. Malapaschas A. Accessibility and acceptability of voice assistants for people with acquired brain injury. SIGACCESS Access Comput 2020 Mar 03:1. [doi: 10.1145/3386402.3386406]
23. Kobayashi M, Kosugi A, Takagi H, Nemoto M, Nemoto K, Arai T, et al. Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2019:1. [doi: 10.1007/978-3-030-29390-1_4]
24. Balasuriya S, Sitbon L, Bayor A, Hoogstrate M, Brereton M. Use of voice activated interfaces by people with intellectual disability. 2018 Presented at: ACM Int Conf Proceeding Ser; 2018; Chengdu. [doi: 10.1145/3292147.3292161]
25. Wolters M, Georgila K, Moore JD, Logie RH, MacPherson SE, Watson M. Reducing working memory load in spoken dialogue systems. Interact Comput 2009 Aug;21(4):276-287. [doi: 10.1016/j.intcom.2009.05.009]
26. Anderson P. Towards a developmental model of executive function. Exec Funct Front Lobes A Lifesp Perspect 2011:1. [doi: 10.4324/9780203837863]
27. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. Int J Geriat Psychiatry 2006 Nov;21(11):1078-1085. [doi: 10.1002/gps.1610]
28. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: A frontal assessment battery at bedside. Neurology 2000 Dec 12;55(11):1621-1626. [doi: 10.1212/wnl.55.11.1621]
29. Fussi F, Cantagallo A, Bertozzini L. Profilo di valutazione della disartria: adattamento italiano del test di Robertson, raccolta di dati normativi e linee di trattamento. Turin: Omega; 1999.
30. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state:" A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6]
31. Boersma P, van Heuven V. Speak and unspeak with Praat. Glot Int 2001;5:341-347.
32. Akaike H. Information theory and the maximum likelihood principle. 1973 Presented at: 2nd Int Symp Inf Theory; 1973; Tsahkadsor. [doi: 10.1007/978-1-4612-1694-0_15]
33. Miles J. Tolerance and variance inflation factor. In: Wiley StatsRef: Statistics Reference Online, 2014. Hoboken: John Wiley & Sons; 2014.
34. Myers R. Classical and Modern Regression With Applications, 2nd Edition. Boston: Duxbury; 1990.
35. Kowalski J, Skorupska K, Kope W, Jaskulska A, Abramczuk K, Biele C, et al. Older adults and voice interaction: a pilot study with google home. 2019 Presented at: Conf Hum Factors Comput Syst—Proc; 2019; Glasgow. [doi: 10.1145/3290607.3312973]
36. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc 1992 Oct;40(9):922-935. [doi: 10.1111/j.1532-5415.1992.tb01992.x] [Medline: 1512391]
37. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: a systematic review. Int J Med Inform 2016 Dec;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]

Abbreviations

ACE-R: Addenbrooke's Cognitive Examination–Revised
AIC: Akaike information criterion
dB: decibel
FAB: Frontal Assessment Battery
MMSE: Mini-Mental State Examination
VIF: variance inflation factor

Edited by G Eysenbach; submitted 26.02.20; peer-reviewed by W Wilkowska, C Simons; comments to author 10.04.20; revised version received 05.06.20; accepted 26.07.20; published 25.09.20.

Please cite as:
Masina F, Orso V, Pluchino P, Dainese G, Volpato S, Nelini C, Mapelli D, Spagnolli A, Gamberini L
Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study
J Med Internet Res 2020;22(9):e18431
URL: http://www.jmir.org/2020/9/e18431/
doi:10.2196/18431
PMID:32975525

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